Procedia Computer Science

Volume 80, 2016, Pages 1–12

ICCS 2016. The International Conference on Computational Science

Detecting frog calling activity based on acoustic event detection and multi-label learning

Jie Xie, Towsey Michael, Jinglan Zhang, and Paul Roe

Queensland University of Technology, Brisbane, Australia
j3.xie@hdr.qut.edu.au
{m.towsey, jinglan.zhang, p.roe}@qut.edu.au

Abstract

Frog population has been declining the past decade for habitat loss, invasive species, climate change, and so forth. Therefore, it is becoming ever more important to monitor the frog population. Recent advances in acoustic sensors make it possible to collect frog vocalizations over large spatio-temporal scale. Through the detection of frog calling activity with collected acoustic data, frog population can be predicted. In this paper we propose a novel method for detecting frog calling activity using acoustic event detection and multi-label learning. Here, frog calling activity consists of frog abundance and frog species richness, which denotes number of individual frog calls and number of frog species respectively. To be specific, each segmented recording is first transformed to a spectrogram. Then, acoustic event detection is used to calculate frog abundance. Meanwhile, those recordings without frog calls are filtered out. For frog species richness, three acoustic features, linear predictive coefficients, Mel-frequency Cepstral coefficients and wavelet-based features are calculated. Then, multi-label learning is used to predict frog species richness. Lastly, statistical analysis is used to reflect the relationship between frog calling activity (frog abundance and frog species richness) and weather variables. Experiment results show that our proposed method can accurately detect frog calling activity and reflect its relationship with weather variables.

Keywords: Frog abundance, frog species richness, multi-label learning, acoustic event detection, multiple regression analysis

1 Introduction

Over the past decade, a dramatic decline of frog population has been noticed worldwide [9]. Reasons for this decline can be summarized as habitat loss, invasive species, climate change. On one hand, frog population is rapidly declining, and on the other frogs are greatly important for the environment. First, frogs are an integral part of the food web, and the decline of their population can result in negative impacts through a whole-ecosystem. Second, frogs are

Selection and peer-review under responsibility of the Scientific Programme Committee of ICCS 2016

important indicator species for environmental health. Third, frogs are very useful in medical research that benefit humans[1]. Therefore, it is becoming necessary to protect frogs.

To monitor the change of frog population and optimize the protection policy, a growing number of researchers have shown interest in studying frogs [7, 3, 16]. Compared with counting frogs by visual observation, hearing frog vocalizations is much easier. Consequently, frogs' vocalizations are often used to study. Currently, there are two approaches for acosutic data collection. The traditional field survey method, which requires ecologists to physically visit sites for collecting bioacoustic data, is both time-consuming and costly [13]. Comparatively, recent advances in acoustic sensor techniques have greatly extended the spatio-temporal scale for collecting frog vocalizations. By the aid of acoustic sensors, it is possible to record frog vocalizations continuously and store them permanently. However, this technique provides us several gigabytes of compressed data per acoustic sensor every day. Develop automatic methods for exploring these quantities of acoustic data is becoming necessary.

Unfortunately, most previous methods focus on the recognition and classification of short-term recordings (segmented frog syllables) [7, 3, 16]. Few studies have explored long-term acoustic monitoring of frogs. Canavero *et al* [4] studied the relationship between calling activity of anuran assemblage and seasonal changes of weather factors such as temperature and rainfall. However, the activity of anuran species was simply quantified as rank abundance estimations of calling mate, which cannot accurately reflect frog species richness of each segmented recording. Ospina *et al* [10, 1] introduced a method named Region of Interest to identify the presence/absence of each species. Then, the detection model was built based on Hidden Markov Model with five frog parameters: maximum and minimum frequency, bandwidth, duration and slice between notes. The disadvantage of this method is that it can detecte the presence/absence of each frog species, but not frog abundance[2]. Akementins *et al* [2] explored the influence of abiotic cues on calling activity. Like [4], frog calling activity was quantified according to the numerical classification scheme. However, this method cannot accurately recognize frog species of each segmented recording. Xie *et al* [14] introduced acoustic indices for frog calling activity detection. Four indices, spectral peak track index, harmonic structure index, oscillation structure index, and Shannon entropy index, were selectively combined for detecting frog calling activity with a Gaussian mixture model. Then, the correlation between the frog calling activity and climate information was studied. However, the correlation was not strong because of the limitation of recording duration.

In this paper, we proposed a novel method for detecting frog calling activity. Here, frog calling activity, which consists of frog abundance and frog species richness, is detected based on acoustic event detection and multi-label learning. Frog abundance and frog species richness denote the number of individual frog calls and the number of different frog species of each segmented recording, respectively. Specifically, we first sample 10 seconds from every 10-minute recordings. Then, short-time Fourier transform (STFT) is used to obtain a spectrogram for each 10-second recording. Next, acoustic event detection is applied to the spectrogram image for frog abundance detection, which is also used to recognize those recordings without frog calls. Finally, multi-label learning is used to calculate frog species richness with three acoustic features: linear predictive coefficients, Mel-frequency Cepstral coefficients and wavelet-based features. After detecting frog abundance and frog species richness, statistical analysis is used to find the relationship between frog calling activity (frog abundance and frog species richness) and weather variables (temperature and rainfall). Experiment results show that our proposed method can accurately monitor frog calling activity and reflect its relationship with weather

[1] http://www.savethefrogs.com/why-frogs/

[2] Frog abundance denotes the number of individual frog calls

variables.

2 Materials and methods

The architecture of our calling activity detection system is shown in Figure 1. The system consists of three parts: frog abundance detection, frog species richness detection, and correlation analysis.

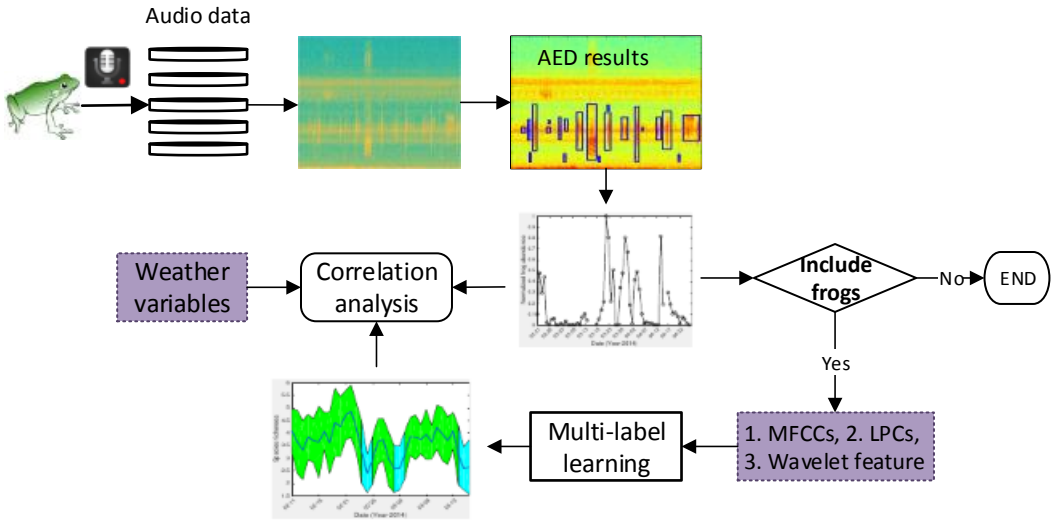


Figure 1: Flowchart of frog call classification system

2.1 Acquisition of frog call recordings

All recordings selected for this study were obtained from three sites in Queensland, Australia: *Kiyomi dam*, *Stony creek dam* and *BG creek dam*, using a battery-powered acoustic sensor (stored in a weather proof metal box) with an external microphone. The recordings were stored on 16GB SD cards in 64 kbps MP3 mono format. All recordings were collected from February, 2014 to April, 2014, because it is the breeding season in Queensland when male frogs make calls to attract female for reproducing. All recordings started around sunset, finished around sunrise every day and have 12 hour duration. We sampled 10-second recordings every 10 minutes for those continuous recordings. There are 4170, 4908, and 1544 10-second recordings for *Kiyomi dam*, *Stony creek dam* and *BG creek dam* respectively, because of data loss. A representative sample of 342 10-second recordings was selected to train and evaluate the proposed method. The ground truth of those 342 10-second recordings is generated by a frog expert who manually tags each recording with frog species.

We first manually inspected spectrograms of ten randomly selected call examples for each frog species. Two parameters, dominant frequency and syllable duration, were then measured and averaged, as listed in Table 1, which are used as prior information for subsequent analysis.

Table 1: Dominant frequency (F_0) and syllable duration (T_s) of eight frog species averaged for ten randomly selected syllables.

Frog species	Code	Dominant frequency (Hz)	Syllable duration(ms)
Canetoad	CAD	560	NA
Cyclorana novaehollandiae	CNE	610	400
Limnodynastes terraereginae	LTE	610	100
Litoria fallax	LFX	4000	280
Litoria nasuta	LNA	2800	160
Litoria rothii	LRI	1800	500
Litoria rubella	LRA	2300	580
Uperolela mimula	UMA	2400	120

2.2 Frog abundance monitoring

Frog abundance is monitored through the detection of acoustic events in a spectrogram image. Here, the spectrogram was generated by applying short-time Fourier transform (STFT) to each 10-second recording. Acoustic event detection, which consists of multiple image processing steps, are modified from our previous study [15] and summarized as follows.

Step 1: Wiener filter

To de-noise and smooth the spectrogram, a 2-dimensional Wiener filter is applied to the spectrogram image over a 5×5 time-frequency grid, where the filter size is selected after the consideration of trade-off between removing the background graininess and blurring acoustic events.

$$\hat{S_{tf}} = \mu + \frac{(\sigma^2 - \nu^2)}{\sigma^2}(S_{tf} - \nu) \tag{1}$$

where μ and σ^2 are local mean and variance, respectively. ν^2 is the noise variance estimated by averaging all local variances.

Step 2: Spectral subtraction

After Wiener filtering, the graininess has been removed. However, some noises such as wind, insect, motor engine that cover the whole recording are still remained. Here, a modified spectral subtraction is employed for dealing with those noises. Description of this algorithm can be found in our previous study [16].

Step 3: Adaptive thresholding

Following noise reduction, the next step is to convert the noise reduced spectrogram $\hat{S_{tf}^{'}}$ into the binary spectrogram S_{tf}^{b} for events detection. Here, an adaptive thresholding method named *Otsu thresholding* [11] is employed to find an optimal threshold.

$$\phi_b^2(k) = w_1(k)w_2(k)[\mu_1(k) - \mu_2(k)]^2 \tag{2}$$

where $w_1(k) = \sum_0^k p(j)$ is calculated from the histogram as k, $p(j) = n(j)/N$ are the values of the normalized gray level histogram, $n(j)$ is the number of values in level j, N is the total number of values over the whole spectrogram image, $\mu_1(k) = [\sum_0^k p(j)x(j)]/w_1$, $x(j)$ is the value at the center of the jth histogram bin. Then, the threshold, T_0, is calculated as

$$T_0 = (\phi_{b1}^2(k) + \phi_{b2}^2(k))/2 \tag{3}$$

Step 4: Events filtering using dominant frequency and event area

To further remove those events that are not belong to frog species shown in Table 1, dominant frequency (F_0) and area (number of pixels) within the event boundary (Ar) are used for filtering. First, large acoustic events, whose area is larger than A_{large}, are separated into small events, because the area of frog calls to be classified in Table 1 is empirically smaller than A_{large}. Then,

dominant frequency is used to filter the events. First, the averaged frequency is calculated by averaging the peak frequency within each acoustic event. Then, the event, whose averaged frequency are not within allowed fluctuation in both sides of dominant frequency, are discarded. Finally, small acoustic events, whose area is smaller than A_{small}, are filtered. Those events, whose average frequency are between 300 Hz and 800 Hz, are not filtered using A_{small}, because the area of LTE (averaged frequency is between 300 Hz and 800 Hz) is smaller than A_{small}. Figure 2 shows the acoustic event detection results.

To calculate frog abundance of each segmented recording, the frog abundance is calculated as follows.

$$F_{abun} = \sum_{n=1}^{N} A_{i,j}(n)^2 \tag{4}$$

Here, $A_{i,j}$ represents the decibel value of location (i, j) within each acoustic event n in the spectrogram.

2.3 Wavelet-based feature extraction for species richness analysis

Frog species richness is calculated by tagging each segmented recording. Since many segmented recordings consist of multiple frog species, one direct solution is to assign each recording with a set of labels (frog species) for explicitly expressing its semantics [17]. Therefore, multi-label learning is adopt to tag each segmented recording.

Extracting discriminating features, which maximize between-group (inter-specie) dissimilarity and minimize within-group (intra-specie) dissimilarity, is very important for achieving high classification performance [7, 3]. In this study, feature extraction is performed based on wavelet packet decomposition using a modified version of the method introduced in [16] and summarized below.

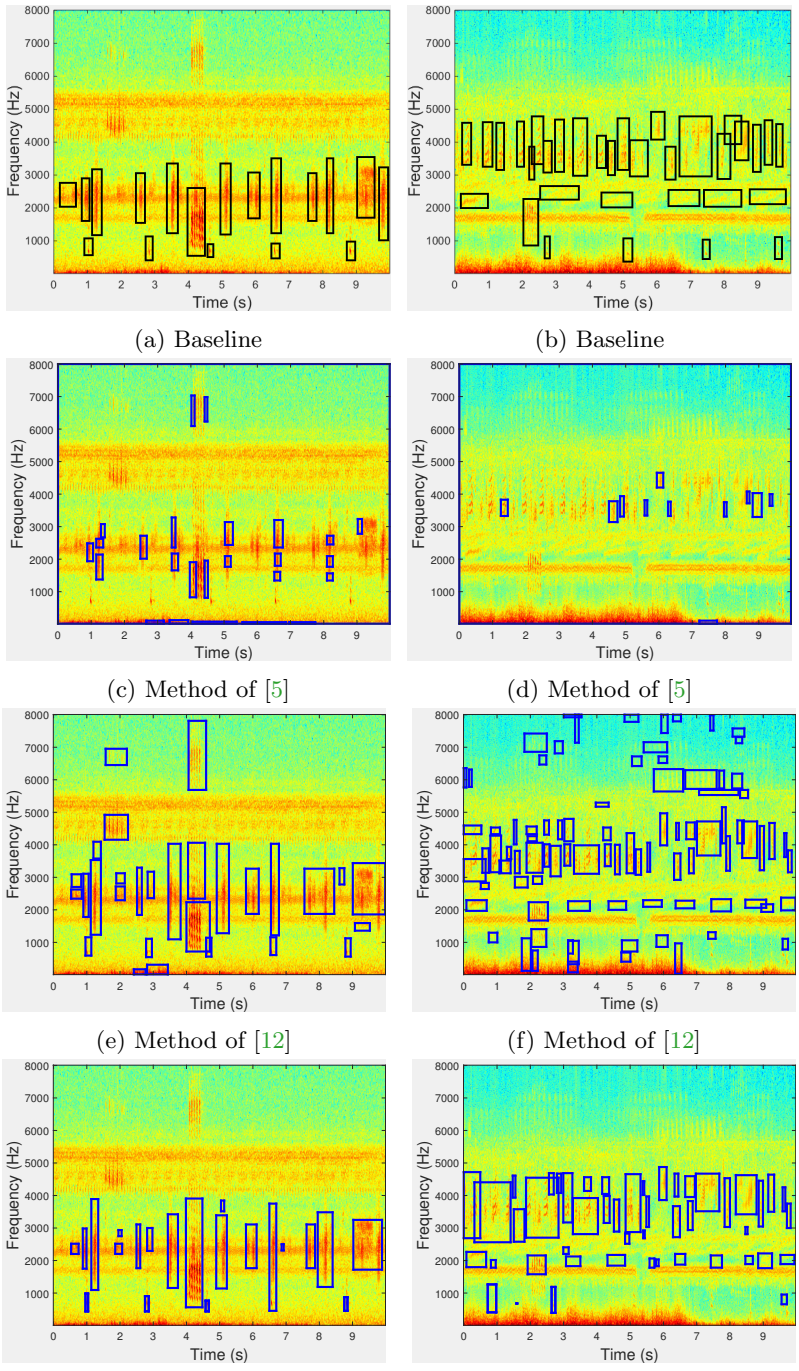
For feature extraction, constructing a suitable frequency scale for wavelet packet (WP) tree based on the dominant frequency of each frog species is the first step, because different frog species tend to have different dominant frequencies [6]. In [16], k-means clustering was first applied to the extracted dominant frequencies of training data. Then, the frequency scale was built by sorting clustering centroids to construct the WP tree. In this study, the prior information (F_0) obtained from Table 1 is directly used to construct the WP tree. We iteratively detect each WP tree sub-band node until the frequency range of each node includes more than one F_0. Then, the WP tree of that particular sub-band node will be further split until each sub-band node has only one dominant frequency value or none. After constructing the frequency scale, adaptive frequency scaled wavelet packet decomposition is applied to each segmented recording for feature extraction.

For each 10-second recording, it is represented as $y(n)$, $n = 1, ..., N$, where N is the length of each recording. Based on the $y(n)$, detailed description for WP-based feature extraction is list as follows:

Step 1: Add a Hamming window to the signal y(n) and perform wavelet packet decomposition spaced in adaptive frequency scale as described in [16].

$$WP(i, j) = \sum_{i=1}^{M} y(n)w(n)\psi_{(a,b)}(n) \tag{5}$$

where $w(L)$ is the Hamming window function, $WP(i, j)$ is the wavelet coefficients of the decomposition, i is the sub-band index, j is the index of wavelet coefficients, $\psi_{(a,b)}(n)$ is the



(a) Baseline (b) Baseline

(c) Method of [5] (d) Method of [5]

(e) Method of [12] (f) Method of [12]

(g) Proposed method (h) Proposed method

Figure 2: Acoustic event detection for frog abundance monitoring using different methods. For each row, different methods are applied to the same recordings. The baseline of the detection results is show in the first row; detected frog calls are draw using a blue rectangle.

wavelet base function, and we use 'db4' experimentally. Here, a and b are the scale and shift parameters, respectively.

Step 2: Calculate the total energy of each sub-band.

$$WP_i = \sum_{j=1}^{M_i} [WP(i,j)]^2 \tag{6}$$

where $i = 1, 2, ..., T$, and T is the total number of sub-band, and $j = 1, 2, ..., M_i$, M_i is the total number of wavelet coefficients.

Step 3: Normalize the energy of each sub-band.

$$SE_i = \frac{WP_i}{M_i} \tag{7}$$

where $i = 1, 2, ..., T$.

Step 4: Perform discrete cosine transform on the logarithm sub-band energy for dimension reduction and obtain the WP-based feature.

$$WP_{base}(d) = \sum_{i=1}^{T} logSE_i cos(\frac{d(i-0.5)}{T}\pi) \tag{8}$$

where $d = 1, 2, ..., d^{'}$, $1 \leq d^{'} \leq T$, here $d^{'}$ is the dimension of WP-based feature, and set as 12.

Different from [16], the recording is first segmented into frames using a Hamming window. Then, all frames are divided into three equal parts, and WP-feature within each part is averaged, respectively, because different frog species within similar frequency band may exist in one 10-second recording, segmenting each recording into small parts might be able to keep the information of different frog species in the same frequency band. Besides WP-based feature, two other acoustic features, linear predictive coefficients (LPCs) and Mel-frequency Cepstral coefficients (MFCCs), are also calculated for the comparison.

2.4 Multi-label classification for species richness analysis

Since many segmented recordings consist of calls from multiple frog species, frog call classification can be framed as a multi-label classification problem. However, previous studies have not adopted multi-label learning to classify frog calls. Therefore, it is worth to investigate different multi-label learning algorithms for the classification of multiple vocalizing frog species. In this study, four multi-label learning algorithms, whose base classifier is C4.5 decision tree, are employed: binary relevance (BR), classifier chains (CC), random k-labEL Pruned Sets (RAKEL and RAKEL1) [17]. The default parameter settings of those four multi-label learning algorithms are used. The trained classifier, which achieves the best classification performance, is then used to tag rest recordings. After tagging each 10-second recording, frog species richness is lastly calculated as follows.

$$F_{rich} = \frac{\sum_{k=1}^{K} f_{rich}(k)}{K} \tag{9}$$

where $f_{rich}(k)$ is the number of frog species of each tagged 10-second recording, K is the number of 10-second recording for each day.

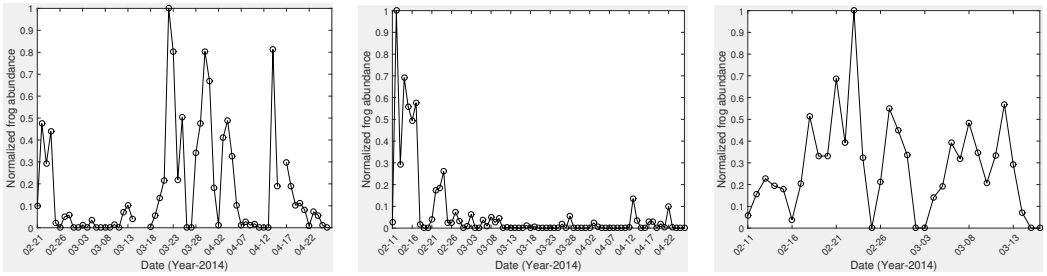
3 Experiment setup

Each 10-second recording is divided into frames of 512 samples and 50% frame overlap for STFT. A_{large} and A_{small}, which are used for area filtering in acoustic event detection, are empirically set at 3000 pixels and 300 pixels, respectively. Allowed fluctuations in both sides of dominant frequency are 300 Hz for dominant frequency filtering. For WP-based feature, window size and overlap are 512 samples and 50%, the window function is a Hamming window. All algorithms were programmed in Matlab 2014b except multi-label learning, which was implemented in Meka 1.7.7[4].

4 Experiment results

4.1 Frog abundance detection

Figure 3 shows the frog abundance result of three selected sites through the whole frog breeding season. It can be found that the frog abundance of the same site changes a great deal over time. In the *Kiyomi dam*, frog abundance is relatively high from February 21 to February 25. However, frog abundance is quite low in two period, which are February 26 to March 11 and April 07 to April 12. The highest abundance of this site is achieved on March 22. However, the highest abundance for *Stony creek dam* and *BG creek dam* is obtained in February, which shows that frog abundance of different sites often varies a lot for different environments. Recordings of 47 days of all three sites have no frog calls. In the subsequent analysis, only those recordings that consist of frog calls are used for frog species richness analysis.



(a) *Kiyomi dam* (b) *Stony creek dam* (c) *BG creek dam*

Figure 3: Frog abundance detection of different sites: *Kiyomi dam*, *Stony creek dam* and *BG creek dam*. For *Kiyomi dam*, three days do not record any acoustic data and then there is no value in those particular days. All the frog abundance value is normalized to [0 1].

4.2 Frog species richness analysis

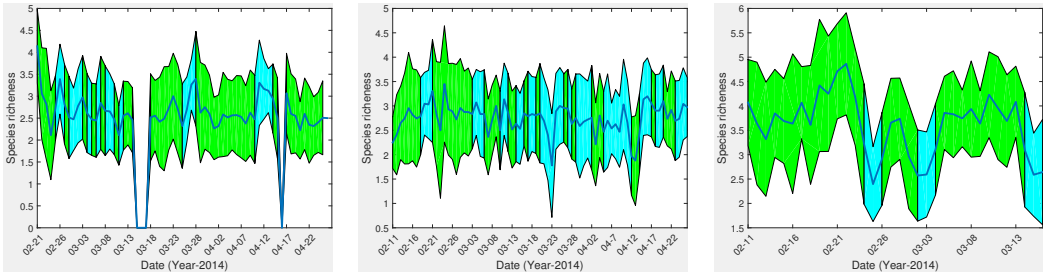
We apply different multi-label learning algorithms on 342 selected recordings to compare different feature sets. Then, six evaluation rules are used to compare the performance with the combination of four feature sets and four multi-label algorithms: Hamming loss, Rank loss, Average precision, One error, Exampled based F1, and Micro F1 [8, 17]. Experiment results are shown in Table 2.

[4] http://meka.sourceforge.net/

Table 2: Classification results based on four feature sets and four multi-label learning algorithms. Here the methods for multi-label algorithms are in accordance to the name in the Meka software. The base classifier of all methods is decision tree. For a metric, the best value is in bold. Here, ↓ indicates the smaller the better, while ↑ indicates the bigger the better.

Features	Method	Hamming loss ↓	Rank loss ↓	Average precision ↑	One error ↓	Exampled based F1 ↑	Micro F1 ↑
MFCCs+LPCs	BR	0.155 ± 0.015	0.171 ± 0.037	**0.446 ± 0.061**	0.246 ± 0.063	0.699 ± 0.03	0.749 ± 0.024
	CC	0.147 ± 0.018	0.147 ± 0.02	0.35 ± 0.016	0.199 ± 0.042	0.722 ± 0.035	0.756 ± 0.029
	RAkEL	0.167 ± 0.038	0.122 ± 0.026	0.333 ± 0.017	0.194 ± 0.063	0.721 ± 0.044	0.752 ± 0.041
	RAkEL1	0.134 ± 0.012	0.099 ± 0.025	0.342 ± 0.023	**0.147 ± 0.056**	0.74 ± 0.044	0.783 ± 0.022
Multi-stage MFCCs + LPCs	BR	0.155 ± 0.016	0.169 ± 0.035	0.445 ± 0.062	0.249 ± 0.064	0.7 ± 0.03	0.75 ± 0.024
	CC	0.147 ± 0.018	0.147 ± 0.021	0.35 ± 0.016	0.199 ± 0.042	0.722 ± 0.034	0.756 ± 0.028
	RAkEL	0.166 ± 0.035	0.124 ± 0.027	0.334 ± 0.018	0.194 ± 0.069	0.724 ± 0.048	0.754 ± 0.04
	RAkEL1	0.134 ± 0.013	0.101 ± 0.026	0.342 ± 0.02	0.15 ± 0.063	0.737 ± 0.05	0.783 ± 0.023
WP-based feature + LPCs	BR	0.148 ± 0.025	0.139 ± 0.033	0.356 ± 0.065	0.254 ± 0.063	0.708 ± 0.046	0.762 ± 0.036
	CC	0.168 ± 0.031	0.168 ± 0.045	0.341 ± 0.027	0.272 ± 0.061	0.684 ± 0.054	0.723 ± 0.048
	RAkEL	0.155 ± 0.023	0.103 ± 0.022	0.324 ± 0.018	0.178 ± 0.031	0.729 ± 0.032	0.763 ± 0.030
	RAkEL1	0.14 ± 0.027	0.094 ± 0.018	0.333 ± 0.028	0.193 ± 0.063	0.727 ± 0.053	0.773 ± 0.042
Multi-stage WP-based feature + LPCs	BR	0.153 ± 0.014	0.147 ± 0.022	0.364 ± 0.056	0.266 ± 0.037	0.689 ± 0.035	0.75 ± 0.025
	CC	0.142 ± 0.029	0.146 ± 0.023	0.345 ± 0.019	0.254 ± 0.094	0.714 ± 0042	0.764 ± 0.045
	RAkEL	0.154 ± 0.022	0.11 ± 0.012	0.33 ± 0.027	0.196 ± 0.062	0.739 ± 0.022	0.768 ± 0.025
	RAkEL1	**0.131 ± 0.012**	**0.09 ± 0.014**	0.33 ± 0.026	0.173 ± 0.03	**0.743 ± 0.026**	**0.787 ± 0.018**

The combination of multi-stage WP-based feature+LCPs and the RAkEL1 method achieves the best performance. Therefore, this combination is used for the testing data. Figure 4 shows the frog species richness of the three selected sites. For all the three sites, the variation of species richness is not high, which shows that species richness of the same area is relatively stable. However, frog species richness of *BG creek dam* has a smaller variation over the time than *Kiyomi dam* and *Stony creek dam*. The comparison of the species richness for the three sites is shown in Figure 5. In contrast to other sites, the species richness in *BG creek dam* is the highest. This might be that *BG creek dam* is closer to a river and farther away from the human community.



(a) *Kiyomi dam* (b) *Stony creek dam* (c) *BG creek dam*

Figure 4: Frog species richness distribution of three selected sites. Here green bar represents the species variation, blue bar means there is no frog calls, zero value denotes the data loss of those particular days.

4.3 Statistical analysis

Multiple regression analysis is used to explore frog calling activity (frog abundance and frog species richness) along weather variables (mean temperature and rainfall)[5]. Frog calling activity is found to be highly correlated with mean temperature ($F=5.18$, $P<0.05$ for abundance, and $F=10.7$, $P<0.01$ for species richness). To calculate the correlation between rainfall and

[5]http://www.bom.gov.au/?ref=hdr

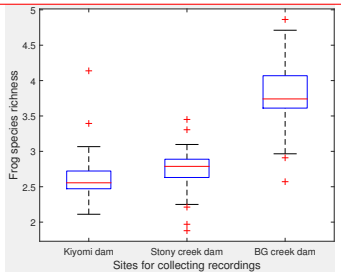


Figure 5: Averaged frog species richness of different sites.

frog calling activity, we first set the rainfall vaule as the dummy variable. Then, the correlation between frog calling activity and rainfall is also studied with multiple regression analysis ($F=4.63$, $P<0.05$ for abundance, and $F=4.64$, $P<0.05$ for species richness). The statistical analysis results indicate that frogs tend to make calls in the warm and humidity environment, which is in accordance to previous studies [2, 4].

5 Conclusion and future work

Acoustic sensors are more widely used to monitor frog calling activity than the traditional field survey method. However, the use of acoustic sensors generates large volumes of audio data, which makes it necessary to develop automated methods. This paper proposes a novel method for detecting frog calling activity based on acoustic event detection and multi-label learning. Specifically, acoustic event detection is the first step to calculate frog abundance. Meanwhile, each 10-second recording is analyzed to decide whether it has frog calls or not. For those recordings with frog calls, multi-label learning is further used for calculating frog species richness with multi-stage WP-based features and LPCs. Finally, statistic analysis is utilized to reflect the relationship between frog calling activity (frog abundance and frog species richness) and weather variables (mean temperature and rainfall). Experiment results show that our proposed method can accurately detect frog calling activity and reflect its relationship with weather variables. Future work will focus on a wider frog call database, including a larger number of frog species, and frog calls collected over a longer period.

6 Acknowledgement

The authors would like to thank Mingying Zhu for the help of statistic analysis, the support of James Cook University, China Scholarship Council and Wet Tropics Management Authority.

References

[1] T Mitchell Aide, Carlos Corrada-Bravo, Marconi Campos-Cerqueira, Carlos Milan, Giovany Vega, and Rafael Alvarez. Real-time bioacoustics monitoring and automated species identification. *PeerJ*, 1:e103, 2013.

[2] MS Akmentins, LC Pereyra, EA Sanabria, and M Vaira. Patterns of daily and seasonal

calling activity of a direct-developing frog of the subtropical andean forests of argentina. *Bioacoustics*, 24(2):89–99, 2015.

[3] Carol Bedoya, Claudia Isaza, Juan M Daza, and José D López. Automatic recognition of anuran species based on syllable identification. *Ecological Informatics*, 24:200–209, 2014.

[4] Andrés Canavero, Matías Arim, Daniel E Naya, Arley Camargo, Inés Da Rosa, and Raúl Maneyro. Calling activity patterns in an anuran assemblage: the role of seasonal trends and weather determinants. *North-Western Journal of Zoology*, 4(1):29–41, 2008.

[5] Gábor Fodor. The ninth annual mlsp competition: first place. In *Machine Learning for Signal Processing (MLSP), 2013 IEEE International Workshop on*, pages 1–2. IEEE, 2013.

[6] Bruno Gingras and William Tecumseh Fitch. A three-parameter model for classifying anurans into four genera based on advertisement calls. *The Journal of the Acoustical Society of America*, 133(1):547–559, 2013.

[7] Chenn-Jung Huang, Yi-Ju Yang, Dian-Xiu Yang, and You-Jia Chen. Frog classification using machine learning techniques. *Expert Systems with Applications*, 36(2):3737–3743, 2009.

[8] Gjorgji Madjarov, Dragi Kocev, Dejan Gjorgjevikj, and Sao Deroski. An extensive experimental comparison of methods for multi-label learning. *Pattern Recognition*, 45(9):3084 – 3104, 2012.

[9] Malcolm L McCallum. Amphibian decline or extinction? current declines dwarf background extinction rate. *Journal of Herpetology*, 41(3):483–491, 2007.

[10] Oscar E Ospina, Luis J Villanueva-Rivera, Carlos J Corrada-Bravo, and T Mitchell Aide. Variable response of anuran calling activity to daily precipitation and temperature: implications for climate change. *Ecosphere*, 4(4):art47, 2013.

[11] Nobuyuki Otsu. A threshold selection method from gray-level histograms. *Automatica*, 11(285-296):23–27, 1975.

[12] Michael W. Towsey and Birgit Planitz. Technical report : acoustic analysis of the natural environment. 2011, April 2011.

[13] Jie Xie, Michael Towsey, Anthony Truskinger, Philip Eichinski, Jinglan Zhang, and Paul Roe. Acoustic classification of australian anurans using syllable features. In *2015 IEEE Tenth International Conference on Intelligent Sensors, Sensor Networks and Information Processing (IEEE ISSNIP 2015)*, Singapore, Singapore, April 2015.

[14] Jie Xie, Michael Towsey, Kiyomi Yasumiba, Jinglan Zhang, and Paul Roe. Detection of anuran calling activity in long field recordings for bio-acoustic monitoring. In *2015 IEEE Tenth International Conference on Intelligent Sensors, Sensor Networks and Information Processing*, Singapore, April 2015.

[15] Jie Xie, Michael Towsey, Jinglan Zhang, and Paul Roe. Image processing and classification procedure for the analysis of australian frog vocalisations. In *Proceedings of the 2Nd International Workshop on Environmental Multimedia Retrieval*, EMR '15, pages 15–20, New York, NY, USA, 2015. ACM.

[16] Jie Xie, Michael Towsey, Jinglan Zhang, and Paul Roe. Adaptive frequency scaled wavelet packet decomposition for frog call classification. *Ecological Informatics*, 32:134 – 144, 2016.

[17] Min-Ling Zhang and Zhi-Hua Zhou. A review on multi-label learning algorithms. *Knowledge and Data Engineering, IEEE Transactions on*, 26(8):1819–1837, Aug 2014.