Ecological Informatics 32 (2016) 134–144

Contents lists available at ScienceDirect

Ecological Informatics

journal homepage: www.elsevier.com/locate/ecolinf

Adaptive frequency scaled wavelet packet decomposition for frog call classification

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

Electrical Engineering and Computer Science School, Queensland University of Technology, Brisbane, Australia

ARTICLE INFO

Article history:
Received 8 September 2015
Received in revised form 26 January 2016
Accepted 27 January 2016
Available online 4 February 2016

Keywords:
Frog call classification
Spectral peak track
Adaptive frequency scaled wavelet packet decomposition
k-means clustering
k-nearest neighbour
Support vector machine

ABSTRACT

Environmental changes have put great pressure on biological systems leading to the rapid decline of biodiversity. To monitor this change and protect biodiversity, animal vocalizations have been widely explored by the aid of deploying acoustic sensors in the field. Consequently, large volumes of acoustic data are collected. However, traditional manual methods that require ecologists to physically visit sites to collect biodiversity data are both costly and time consuming. Therefore it is essential to develop new semi-automated and automated methods to identify species in automated audio recordings. In this study, a novel feature extraction method based on wavelet packet decomposition is proposed for frog call classification. After syllable segmentation, the advertisement call of each frog syllable is represented by a spectral peak track, from which track duration, dominant frequency and oscillation rate are calculated. Then, a k-means clustering algorithm is applied to the dominant frequency, and the centroids of clustering results are used to generate the frequency scale for wavelet packet decomposition (WPD). Next, a new feature set named *adaptive frequency scaled wavelet packet decomposition sub-band cepstral coefficients* is extracted by performing WPD on the windowed frog calls. Furthermore, the statistics of all feature vectors over each windowed signal are calculated for producing the final feature set. Finally, two well-known classifiers, a k-nearest neighbour classifier and a support vector machine classifier, are used for classification. In our experiments, we use two different datasets from Queensland, Australia (18 frog species from commercial recordings and field recordings of 8 frog species from James Cook University recordings). The weighted classification accuracy with our proposed method is 99.5% and 97.4% for 18 frog species and 8 frog species respectively, which outperforms all other comparable methods.

1. Introduction

During the past decades, a rapid decline in frog biodiversity has been noted worldwide. There are many reasons for this decline, including habitat destruction (Clauzel et al., 2015), invasive species (Shine, 2014), and climate change (Garcia et al., 2014). Researchers investigate frogs to retain their biodiversity and develop effective protection strategies. Due to the development of acoustic sensor techniques, many sensors have been widely deployed for monitoring biodiversity, which produces large volumes of acoustic data (Wimmer et al., 2013). Compared with the traditional manual methods that require ecologists to physically visit sites for collecting biodiversity data, acoustic sensors can help collect audio data over larger spatio-temporal scales (Wimmer et al., 2010; Gage and Axel, 2014) Since several gigabytes of compressed data can be generated by an acoustic sensor per day, enabling automating species identification in acoustic data sets has become important (Zhang et al., 2013).

In recent years, acoustic data has been studied for the recognition and classification of animal calls by many researchers. Almost all the recognition and classification methods consist of four parts: pre-processing, syllable segmentation, feature extraction, and recognition or classification.

Frog call classification has been addressed in several papers. Huang et al. (2009) extracted spectral centroid, signal bandwidth, and threshold crossing rate from each segmented frog syllable. Then, two classifiers, k-nearest neighbour (k-NN) classifier and support vector machine (SVM), were used for classification. However, signal bandwidth and threshold crossing rate are very sensitive to the background noise, which results in low classification accuracy in noisy environments. Han et al. (2011) introduced spectral centroid, Shannon entropy and Rènyi entropy to classify frog calls with a k-NN classifier. Chen et al. (2012) first calculated syllable length for pre-classification of frog calls based on segmented frog syllables. Then, a multi-stage average spectrum was calculated for automatic recognition based on template matching. However, extracting features based on the Fourier transform has a tradeoff between time and frequency resolution, which restricts the discriminability of the features. Bedoya et al. (2014) proposed an automatic recognition system for frog calls based on the Mel-frequency cepstral coefficients (MFCCs) and a fuzzy classifier. However, MFCCs

* Corresponding author.
E-mail address: xiej8734@gmail.com (J. Xie).

http://dx.doi.org/10.1016/j.ecoinf.2016.01.007
1574-9541/

are designed for the human auditory system, and might be not suitable for the classification of frogs (Sahidullah and Saha, 2012). Meanwhile, MFCCs are not suitable for dealing with recordings with a low signal to noise ratio (SNR). In those previous studies (Huang et al., 2009; Han et al., 2011; Chen et al., 2012; Bedoya et al., 2014) most features used are either based on Fourier transform or transplanted from speech, speaker and music fields. To further improve the recognition and classification performance, it is necessary to develop more accurate species identification methods.

Wavelet analysis has been widely employed for acoustic data, because it can preserve both frequency and temporal information (Ren et al., 2008). Yen and Fu (2002) introduced wavelet packet transform (WPT) for individual frog identification. After applying WPT to the frog calls, energy of all the node coefficient were calculated as features. Then, Fisher's criterion (Yen and Lin, 2000) was used for dimension reduction. Finally, the feature vector after dimension reduction was fed into a neural network classifier for identification. Colonna et al. (2012) proposed to use discrete wavelet transform (DWT) for frog call classification. Based on the node coefficients of DWT, energy, power, zero-crossing rate and pitch of each node coefficients were calculated. However, applying WPT and DWT without any modifications cannot provide a good frequency domain resolution for classifying frog calls.

In this study, the WPD is applied to the frog calls with an adaptive frequency scale for feature extraction. Frog species that are genetically similar often share close advertisement calls (Gingras and Fitch, 2013). Therefore, the dominant frequency which is directly calculated from the trace of advertisement call is an important feature for differentiating frog species. We use dominant frequency to produce the frequency scale for WPD, which is different from using minimum and maximum frequency to generate the frequency scale for WPD in Ren et al. (2008). Specifically, continuous acoustic data are first segmented into syllables using Härmä's method (Harma, 2003). Then, spectral peak tracks are extracted from each syllable where possible. Three features are extracted from each track: track duration, dominant frequency and oscillation rate. Next, a k-means clustering algorithm is applied to the dominant frequency, and the centroids of clustering results are used to generate the frequency scale for WPD. After applying the adaptive frequency scaled WPD to the frog calls, a new feature set named *adaptive frequency scaled wavelet packet decomposition sub-band cepstral coefficients* (AWSCCs) is extracted. Finally, two classifiers, a k-NN classifier and a SVM classifier, are employed for the classification with the proposed feature set.

2. Methods

The architecture of the proposed classification method consists of five modules: syllable segmentation, syllable feature extraction, adaptive frequency scale generation, WPD feature extraction and classification (see Fig. 1). Each module is described in the following sections.

2.1. Sound recording and pre-processing

Two datasets obtained from a commercial recording (Stewart, 1999) and James Cook University (JCU) were selected for this study. Recordings, which were collected from the CD, are two-channel, sampled at 44.10 kHz and saved in MP3 format. All recordings were obtained with a directional microphone and have a high signal to noise ratio (SNR). Each recording includes one frog species, and has a duration ranging from twenty-one to fifty-four seconds. The calls of eighteen frog species recorded in Queensland, Australia were used to develop the detailed methodology. To reduce the subsequent computational burden, all recordings were re-sampled at 16 kHz per second, mixed to mono, and saved in WAV format.

The JCU recordings were obtained from Kiyomi dam (S 19°22′16.0′,E146°27′31.3″) BG creek dam (S19°27′1.23″,E146°24′5.65″) and Stony creek dam (S 19°24′07.0″,E146°2551.3) in Townsville, using Song Meter (SM2) (Xie, 2016). The recordings were stored on 16 GB SD cards in 64 kbps MP3 mono format and have a low

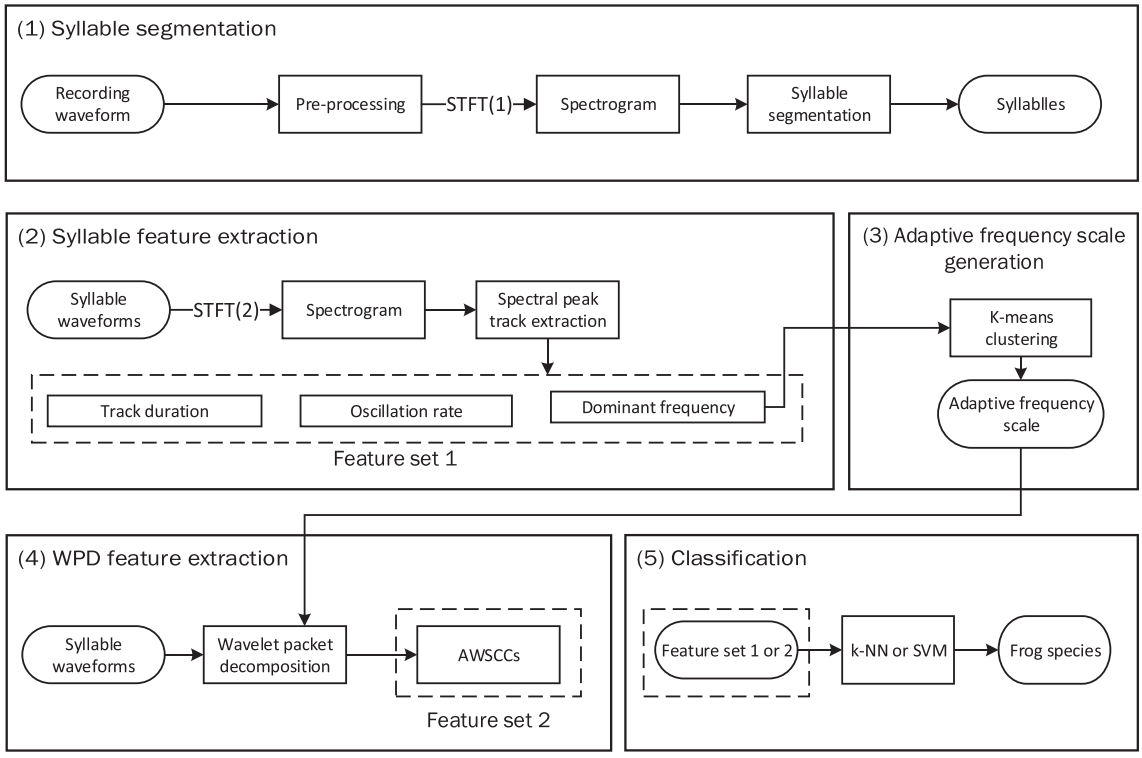


Fig. 1. Block diagram of the frog call classification system. The line of dashes indicates the extracted feature set. AWSCCs is the abbreviation of *adaptive wavelet packet decomposition sub-band cepstral coefficients*. STFT is short-time Fourier transform. For STFT(1), the window function, size and overlap are Kaiser window, 512 samples and 25%. For STFT(2), the window function, size and overlap are Hamming window, 128 samples and 90%. In this diagram, two feature sets are extracted, the description of other feature sets is shown in Fig. 6.

Table 1
Parameters of 18 frog species averaged of three randomly selected syllable samples in the commercial recording. These selected samples make the *reference data set*.

No.	Scientific name	Abbreviation	Syllable duration (millisecond)	Peak frequency (Hz)	Oscillation rate (cycle/s)
1	*Assa darlingtoni*	ADI	80	3200	160
2	*Crinia parinsignifera*	CPA	250	4300	350
3	*Litoria caerulea*	LCA	500	500	50
4	*Litoria chloris*	LCS	800	1700	220
5	*Litoria fallax*	LFX	430	4700	70
6	*Litoria gracilenta*	LGA	1400	2700	100
7	*Litoria latopalmata*	LLA	30	1400	2100
8	*Litoria nasuta*	LNA	100	2800	160
9	*Litoria revelata*	LRA	160	4100	70
10	*Litoria rubella*	LUA	500	2900	60
11	*Litoria verreauxii*verreauxii	LVV	270	3100	125
12	*Mixophyes fasciolatus*	MFS	200	1200	140
13	*Mixophyes fleayi*	MFI	50	1000	140
14	*Philoria kundagungan*	PKN	170	430	95
15	*Pseudophryne coriacea*	PCA	300	2400	80
16	*Pseudophryne raveni*	PRI	370	2500	45
17	*Rheobatrachus silus*	RSS	510	1500	60
18	*Uperoleia laevigata*	ULA	450	2400	150

SNR compared with the commercial recording. All the JCU recordings started around sunset, finished around sunrise every day and have 12 h duration.

2.2. Spectrogram analysis based on validation set

In this study, three syllables for each frog species are set aside and used as a *reference data set*. For the commercial recording, three parameters including syllable duration, dominant frequency, and oscillation rate, are manually calculated for those three syllables of each species and averaged, as listed in Table 1. The reference data set is excluded from the data used in the testing stage.

For the JCU recordings,[1] the corresponding parameters are described in Table 2. Compared with recordings from the commercial recording, peak frequency shows a smaller variation than syllable duration and oscillation rate.

2.3. Syllable segmentation

For frog calls, an elementary acoustic unit for classification is the syllable, which is a continuous vocalization emitted from an individual (Huang et al., 2009). Each commercial recording consists of the continuous multiple calls of one frog species. Therefore, it is necessary to segment each call into individual syllables. This syllable segmentation process is applied to the spectrogram, which is generated by applying short-time Fourier transform (STFT) to each recording. For STFT, the window function is the Hamming window with the size and overlap of 512 samples and 25%, respectively.

To further improve the segmentation result, those syllables whose averaged energy is less than 15% of the maximum energy are removed (Gingras and Fitch, 2013). The distribution of syllable numbers after segmentation for all frog species is shown in Fig. 2.

For the JCU recordings, bandpass filtering is applied to each recording before using Härmä's method. A bandpass filter is first used to filter specific frog species, because different frog species tend to call simultaneously. The filtering is

$$S'(t,f) = \begin{cases} S(t,f) & Fl_{ower} \le f \le F_{upper} \\ 0 & \end{cases}$$

[1] https://www.ecosounds.org/

Table 2
Parameters of 8 frog species obtained by averaging three randomly selected syllable samples from recordings of James Cook University. NA indicates there is no oscillation structure in the spectrogram for the background noise and frog chorus. Since syllable duration of *Rhinella marina* (commom name: Canetoad) is very different from each other, we manually set the duration of Canetoad using the maximum duration of other frog species, which is 500 ms.

No.	Scientific name	Abbreviation	Syllable duration (ms)	Peak frequency (Hz)	Oscillation rate (cycles/s)
1	*Rhinella marina*	CTD	500	680	12
2	*Cyclorana novaehollandiae*	CNE	350	600	NA
3	*Limnodynastes terraereginae*	LTE	80	630	NA
4	Litoria fallax	LFX	120	4100	50
5	Litoria nasuta	LNA	100	2700	NA
6	*Litoria rothii*	LRI	350	1150	15
7	Litoria rubella	LUA	500	2400	NA
8	Uperolela mimula	UMA	120	2400	40

Here, $S'(t,f)$ is the filtered spectrogram, the F_{lower} and F_{upper} are lower and upper cutoff frequency and calculated as

$$\begin{aligned} F_{upper} &= F_{peak} + \beta \\ F_{lower} &= F_{peark} - \beta \end{aligned} \quad (1)$$

where F_{peak} is the peak frequency (Table 2), β is a threshold for determining the frequency bandwidth and set at 300 Hz based on the *reference data set*.

After bandpass filtering, noise reduction is essential for improving the segmentation for the low signal to noise ratio in JCU recordings. Here, we use the method of Towsey et al. (2012) for noise reduction. Finally, we use Härmä's method to detect individual syllables (Fig. 3).

For *Canetoad*, the durations of different calls are very different, therefore, we manually selected 30 syllables whose combined duration is 500 ms.

For the JCU recordings, eight frog species were used for experiment. After syllable segmentation of continuous recordings, for each frog species, we randomly selected 30 syllables from segmentation results for subsequent analysis.

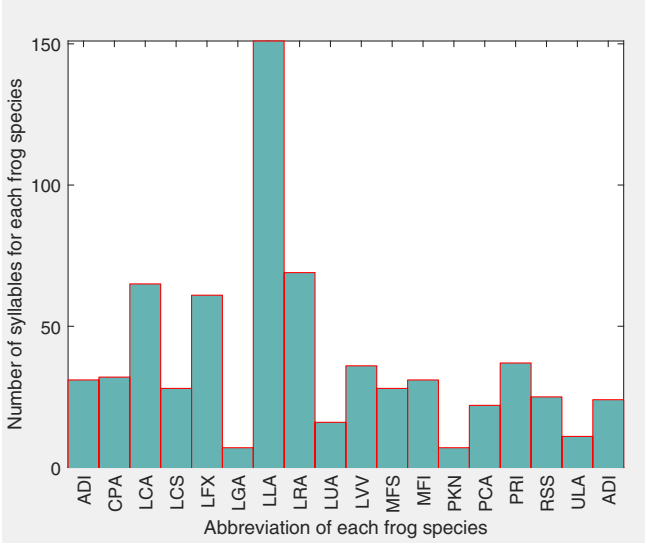


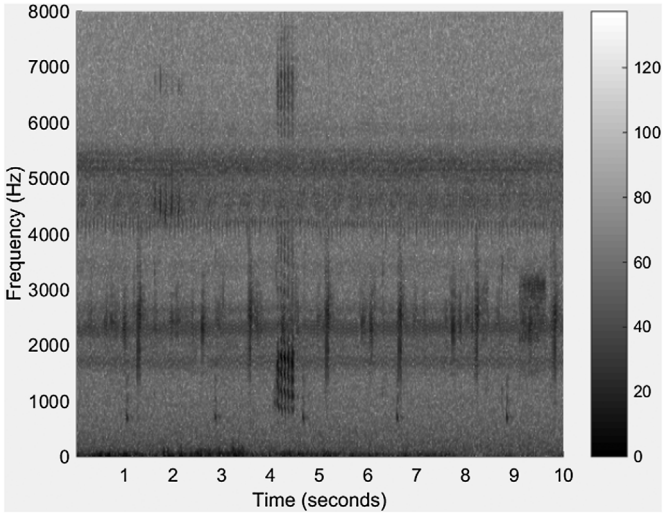
Fig. 2. Distribution of syllable number for all frog species. The x-axis is the abbreviation of each frog species, and the corresponding scientific name can be found in Table 1.

2.4. Spectral peak track extraction

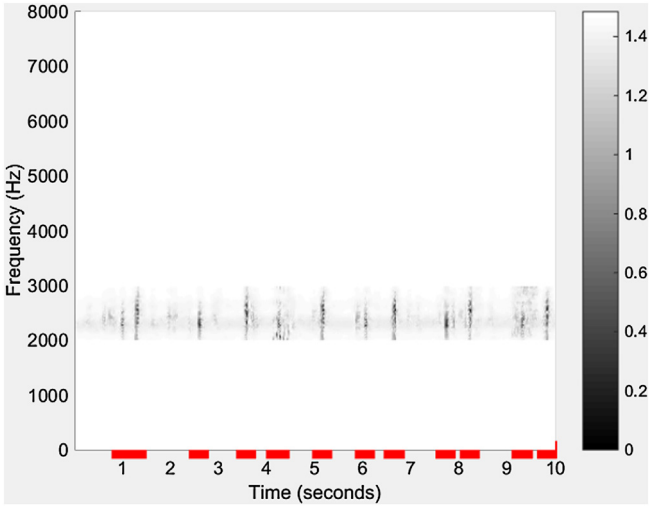
Spectral peak tracks (SPT) (also called frequency tracks) have been explored for studying birds (Jancovic and Kokuer, 2015; Heller and Pinezich, 2008) and whales (Roch et al., 2011). In this study, the spectral peak track is used to represent the trace of a frog advertisement call, because frogs which are genetically related share more similar advertisement calls than distantly related frogs (Gingras and Fitch, 2013). The reasons for using SPT are (1) to isolate the desired frog calls from the background noise; (2) to extract corresponding SPT features. Here, the SPT is extracted using a modified version of the method introduced in Xie et al. (2015) as follows.

For the SPT extraction algorithm, seven parameters need to be set (Table 3). The process for determining those parameters is explained in Section 3.

Before applying the SPT extraction algorithm, each syllable is transformed to a spectrogram with the following parameter settings (Hamming window, frame size is 128 samples, and window overlap is 90%). For the generated spectrogram, the maximum intensity (real peak) is selected from each frame with a minimum required intensity, *I*. Then, the time and frequency domain intervals between two successive peaks are calculated. If the time and frequency intervals are smaller than T_c and f_c respectively, one initial track (SPT_1) will be generated.



(a) Spectrogram.

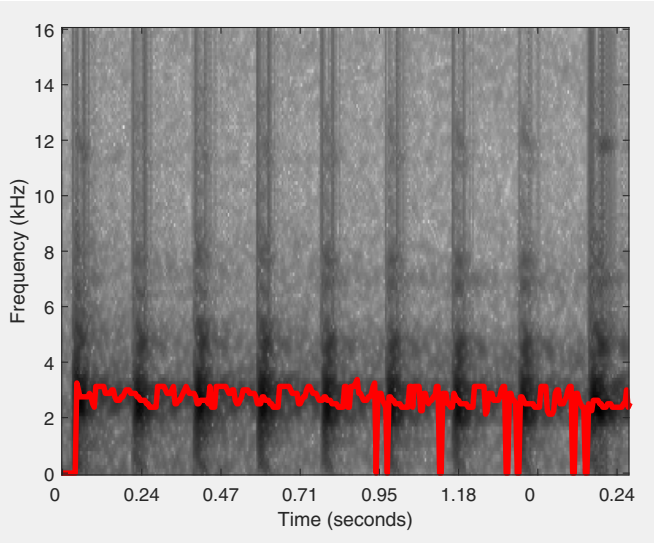


(b) Segmentation results with marked red lines.

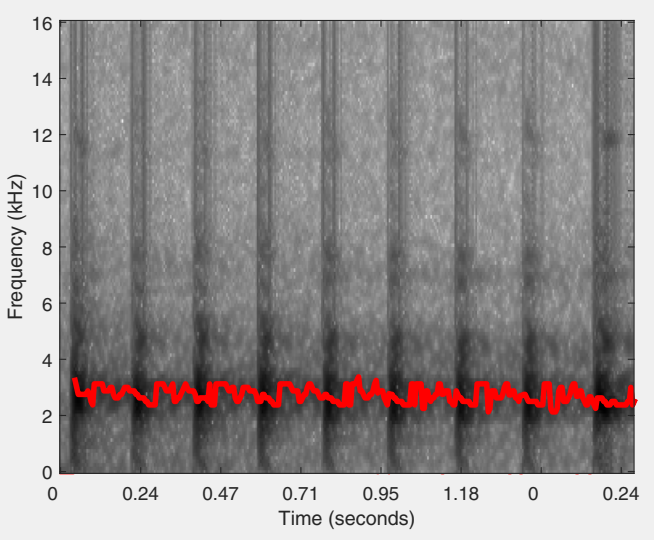
Fig. 3. Segmentation results based on bandpass filtering for *Uperolela mimula*, noise reduction and Härmä's method. The red line in (b) indicates the start and stop location of each segmented syllable. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Table 3
Parameters used for spectral peak extraction.

Parameter	Description
I (dB)	Minimum intensity threshold for peak selection
T_c (s)	Maximum time domain interval for peak connection
T_s (s)	Minimum time interval for stopping growing tracks
f_c (Hz)	Maximum frequency domain interval for peak connection
d_{min} (s)	Minimum track duration
d_{max} (s)	Maximum track duration
β (0~1)	Minimum density value



(a) selected peaks below the intensity threshold *I* and are set to zero.



(b) spectral peak track with predicted peaks using linear regression.

Fig. 4. Spectral peak track extraction results for *Neobatrachus sudelli*. By filling the gaps within the track, the dominant frequency can be more accurately calculated.

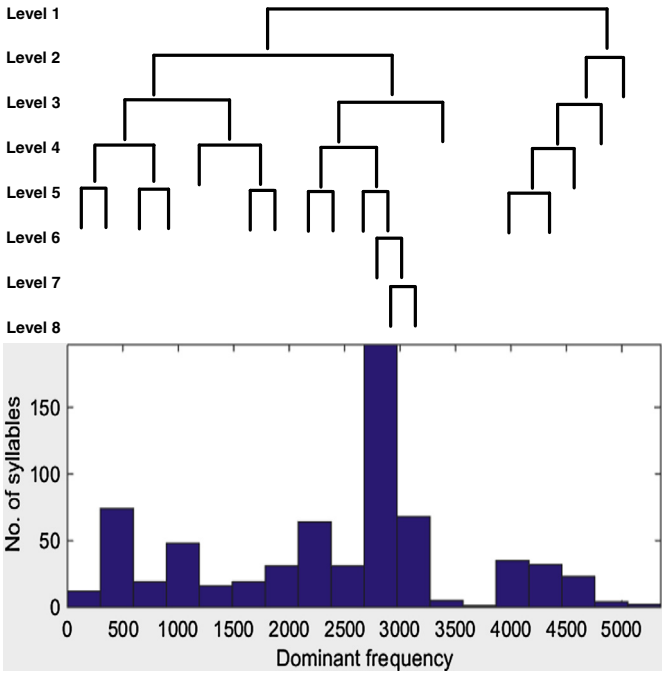


Fig. 5. Adaptive wavelet packet tree for classifying twenty frog species. The upper image is the wavelet packet tree; the lower image is the histogram of dominant frequency for twenty frog species.

After that, linear regression is applied to the generated track for calculating the position of the next predicted peak. Based on peaks $p_1(t_1,f_1)$ and $p_2(t_2,f_2)$ within the initial track (SPT_1), a and b in Eq. (2) can be solved.

$$f = at + b \tag{2}$$

Based on a and b, the predicted peak p_n of the following frame t_n can be calculated. Next, the time and frequency domain intervals between predicted peak (p_n) and the real peak of the successive frame are recalculated. If the time and frequency intervals are smaller than T_c and f_c respectively, the real peak will be added to the initial track. After each peak is added to the initial track, linear regression is repeated to recalculate the next predicted peak using at most the last 10 included peaks. This iterative process continues until T_s is no longer satisfied. When no more peaks will be added to one track, the next step is to compare the duration and density of the track with d_{min}, d_{max}, and β. If all conditions are satisfied, then the track will be saved to the track list. The SPT results for *Neobatrachus sudelli* are shown in Fig. 4. During the process of track extraction, time domain gaps are generated where the intensity threshold I is not reached. These gaps can be filled by predicting the correct frequency bin using linear regression, as illustrated in Fig. 4.

Table 4
Parameter setting for calculating spectral peak track.

Parameter	Commercial recordings	JCU recordings
I (dB)	3	3
T_c (s)	0.005	0.1
T_s (s)	0.05	0.2
f_c (Hz)	800	800
d_{min} (s)	0.01	0.05
d_{max} (s)	2	2
β (0 ~ 1)	0.8	0.6

2.5. Syllable SPT features

After SPT extraction, each SPT is expressed in the following format: (1) track start time t_s; (2) track stop time t_e; (3) frequency bin index for each of the peaks within the track f_t ($t_s \le t \le t_e$). Then, syllable features including track duration, dominant frequency, and oscillation rate are calculated based on the SPT.

(a) Track duration (second): Track duration (D_t) is directly obtained from the bounds of the track.

$$D_t = (t_e - t_s) * r_x \tag{3}$$

where r_x is the time domain resolution in unit second per frame.

(b) Dominant frequency (Hz): Dominant frequency ($\overline{f}$) is calculated by averaging the frequency of all peaks within one track

$$\overline{f} = \sum_{t=t_s}^{t_e} f_t / (t_e - t_s + 1) * r_y \tag{4}$$

where r_y is the frequency domain resolution with unit frequency per bin, f_t is the frequency bin index of peak t.

(c) Oscillation rate (Hz): Oscillation rate (O_r) represents the number of pulses per second. The algorithm for extracting oscillation rate

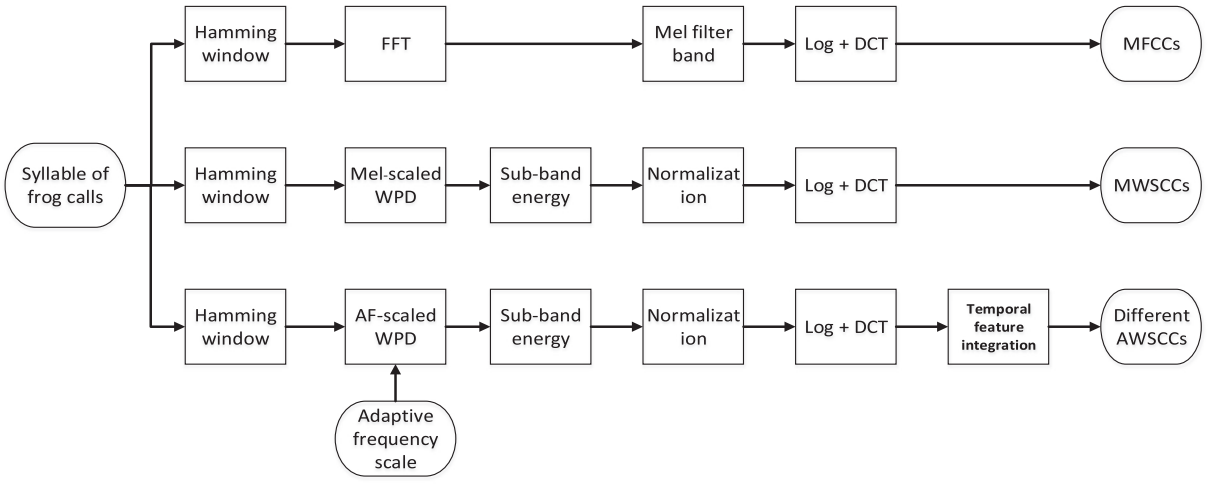


Fig. 6. Description of three feature extraction methods including MFCCs, MWSCCs, and different AWSCCs.

Table 5
Weighted classification accuracy (mean and standard deviation) comparison for five feature sets with two classifiers.

Feature set	Classification accuracy (%)	
	k-NN	SVM
SFs	82.2 ± 11.2	84.2 ± 10.5
MFCCs	90.8 ± 8.6	92.8 ± 11.0
MWSCCs	95.0 ± 7.7	97. ± 5.7
Averaged AWSCCs	98.8 ± 4.2	99.0 ± 3.6
Delta-AWSCCs	99.2 ± 2.1	99.6 ± 1.8

is introduced and summarized as follows. First, the frequency domain boundary is defined based on the dominant frequency, and the power within the boundary is calculated. Then, the power vector is normalized, and the first and last 20% part of the vector is discarded, because of the uncertainty in the start and end of the syllables. Next, the autocorrelation with the length of the vector is calculated. Furthermore, a discrete cosine transform (DCT) is applied to the vector after subtracting the mean, and the position of the highest frequency (P_f) is achieved. Finally, the oscillation rate is defined as

$$O_r = \frac{P_f}{L_{dct}} * r_x * \gamma \tag{5}$$

where P_f is the position of the highest frequency values of the DCT result, L_{dct} is the length for applying DCT to the power vector, and is experientially set as 0.2 s in this study.

2.6. Wavelet packet decomposition

Wavelet packet decomposition (WPD) is a powerful tool for the analysis of non-stationary signals, which includes multiple bases and different basis (Selin et al., 2007). With WPD, an original acoustic signal can be split into two frequency bands such as lower and higher frequency band. Then, both lower and higher frequency bands can be further

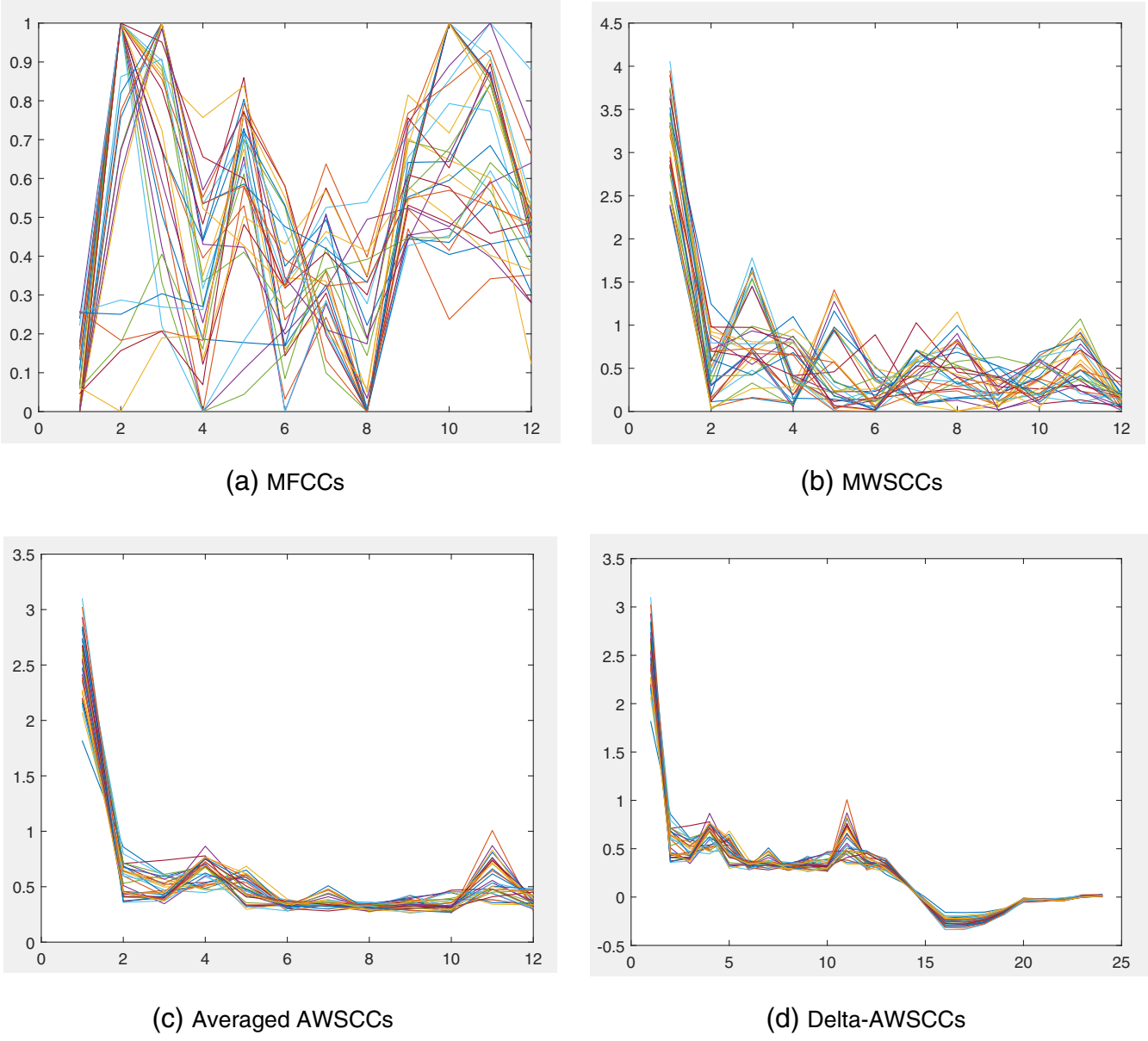


Fig. 7. The feature vectors for 31 syllables of the single species, *Assa darlingtoni*. x-axis is feature index and y-axis is the feature value. Note that the feature vectors for averaged AWSCCs (c) and delta-AWSCCs (d) are more highly correlated than for the other two methods (a) and (b).

continuously decomposed into two sub-bands, which produce a complete wavelet packet tree (Farooq and Datta, 2001). Due to its ability for analyzing a non-stationary signal, WPD has been used to analyze acoustic signals (Selin et al., 2007; Ren et al., 2008). Here, WPD is used to obtain features for frog call classification.

2.7. WPD based on an adaptive frequency scale

To obtain robust features for frog call classification, the frequency scale used for WPD is crucial. In prior work (Litvin and Cohen, 2011; Biswas et al., 2014; Zhang and Li, 2015), different frequency scales have already been proposed for WPD. Bark-scaled WPD was proposed by Litvin and Cohen to separate blind source from a single channel audio source (Litvin and Cohen, 2011). (Biswas et al., 2014) used features based on ERB-scaled (Equivalent rectangular bandwidth) WPD for Hindi consonant recognition. (Zhang and Li, 2015) developed a method based on Mel-scaled WPD for bird sound detection with the SVMs classifier. However, most frequency scales used for WPD are developed for studying speech rather than frogs. Therefore, finding a suitable frequency scale for frogs to perform the WPD is important for obtaining features with strong discriminatory power. In this study, we propose an adaptive frequency scale for WPD for frog calls based on the dominant frequency of frog species to be classified. Specifically, the k-means clustering algorithm is used to cluster the dominant frequency of all syllables. Then, the centroids of the clustering result are used to generate the frequency scale. Here, the value of k for the k-means clustering algorithm is the same as the number of frog species to be classified, the distance function used is *city block* (Melter, 1987).

Based on the obtained frequency scale, an adaptive frequency scaled WPD method is proposed, which is described in Algorithm 1. The wavelet packet tree used for classifying 18 frog species is shown in Fig. 5.

Algorithm 1: Adaptive frequency scale for WPD.

Data: $c_i(i = 1, 2, ..., K)$, f_s, where K is the number of frog species to be classified, c_i is the centroid of the clustering results, $f_s = sr/2$ where sr is the sample rate of the audio recordings, which is 16 kHz here.
Result: Adaptive wavelet packet tree
begin
  Step 1: Sort the centroid $c_i(i = 1, 2, ..., K)$, and calculate the difference between the consecutive vectors of c, sort the difference and save it as $d_j(j = 1, 2, ..., K - 1)$
  Step 2: Calculate the decomposition level L based on the following rule
  $f_s/min(d) \leq 2^{L-1}$
  where L is the minimum integer that satisfies that equation.
  Step 3: Perform the wavelet packet decomposition
  for $l = 1 : L$ **do**
    1. Calculate the frequency resolution of level l
    for $i = 1 : K$ **do**
      1: Put the c_i into the right frequency band
      2: Count the number of c_i in each band (n)
    if $n \geq 2$ **then**
      perform further decomposition to that particular node
    else
      stop decomposition

2.8. Feature extraction based on adaptive frequency scaled WPD

In previous studies (Bedoya et al., 2014; Xie et al., 2015), Mel-frequency cepstral coefficients (MFCCs) have been used for studying bioacoustic data, and used as the baseline for feature comparison in this study. Besides MFCCs, another feature set called Mel-scaled wavelet packet decomposition sub-band cepstral coefficients (MWSCCs) is also included in the comparison experiment (Zhang and Li, 2015), because it shows better performance than MFCCs for bird detection in a complex environment. In this study, we propose a novel feature set named *adaptive frequency scale wavelet packet decomposition sub-band cepstral coefficients* (AWSCCs) for frog call classification. The extraction procedure of AWSCCs is similar to MWSCCs. However, the frequency scale used for our AWSCCs is based on an adaptive frequency scale rather than Mel-scale for MWSCCs. Meanwhile, after performing DCT, temporal feature integration is used for calculating the statistics of feature vectors which generates different AWSCCs (see Fig. 6).

After syllable segmentation, the signal of one syllable is represented as $y(n), n = 1,...,N$, where N is the length of one syllable of frog calls. Based on the $y(n)$, steps for AWSCCs extraction are described as follows:

1) Add Hamming window to the signal $y(n)$.

$$x(n) = w(n)y(n) \tag{6}$$

where $w(L)$ is the Hamming window function and defined as $w(n) = 0.54 - 0.46\cos(\frac{2n\pi}{L-1})$, L is the length of Hamming window and set as 128 samples here.

2) Perform wavelet packet decomposition spaced in adaptive frequency scale as described in Section 2.7.

$$WP(i, j) = \sum_{i=1}^{M} x(n)\psi_{(a,b)}(n) \tag{7}$$

where $WP(i,j)$ 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 wavelet base function, and we use 'Db 4' experimentally. Here, a and b are the scale and shift parameters, respectively. 'Db 4' represents the Daubechies wavelet transform which has four scaling and wavelet function coefficients.

3) Calculate the total energy of each sub-band.

$$WP_i = \sum_{j=1}^{M_i} [WP(i, j)]^2 \tag{8}$$

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.

4) Normalize the energy of each sub-band.

$$SE_i = \frac{WP_i}{M_i} \tag{9}$$

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

Table 6
Classification accuracy of five features for the classification of twenty-four frog species using the SVM classifier. Here, Avg AWSCCs means the averaged AWSCCs.

Code	Classification accuracy (%)				
	SFs	MFCCs	MelCCs	Avg AWSCCs	Delta-AWSCCs
ADI	76.7 ± 15.3	80.0 ± 22.1	83.3 ± 16.7	100.0 ± 0.0	100.0 ± 0.0
CPA	86.7 ± 16.3	100.0 ± 0.0	93.3 ± 13.3	100.0 ± 0.0	100.0 ± 0.0
LCA	93.3 ± 15.3	100.0 ± 0.0	100.0 ± 0.0	100.0 ± 0.0	100.0 ± 0.0
LCS	70.0 ± 23.3	63.3 ± 27.7	96.7 ± 10.0	93.3 ± 13.3	96.7 ± 10.0
LFX	91.7 ± 8.3	93.3 ± 8.2	93.3 ± 8.2	100.0 ± 0.0	100.0 ± 0.0
LGA	30.0 ± 45.8	100.0 ± 0.0	100.0 ± 0.0	100.0 ± 0.0	100.0 ± 0.0
LLA	92.7 ± 8.1	98.7 ± 2.7	98.0 ± 4.3	100.0 ± 0.0	100.0 ± 0.0
LNA	78.6 ± 14.6	94.3 ± 9.5	95.7 ± 9.1	100.0 ± 0.0	100.0 ± 0.0
LRA	40.0 ± 30.0	10.0 ± 20.0	100.0 ± 0.0	90.0 ± 20.0	98.2 ± 6.5
LUA	60.0 ± 20.0	100.0 ± 0.0	86.7 ± 22.1	100.0 ± 0.0	100.0 ± 0.0
LVV	100.0 ± 0.0	96.7 ± 10.0	80.0 ± 22.1	93.3 ± 13.3	100.0 ± 0.0
MFS	90.0 ± 15.3	76.7 ± 21.3	90.0 ± 15.3	100.0 ± 0.0	100.0 ± 0.0
MFI	90.0 ± 30.0	100.0 ± 0.0	100.0 ± 0.0	100.0 ± 0.0	100.0 ± 0.0
PKN	90.0 ± 20.0	100.0 ± 0.0	100.0 ± 0.0	100.0 ± 0.0	100.0 ± 0.0
PCA	72.5 ± 20.8	77.5 ± 20.8	95.0 ± 10.0	92.5 ± 11.5	100.0 ± 0.0
PRI	45.0 ± 35.0	80.0 ± 33.2	100.0 ± 0.0	100.0 ± 0.0	100.0 ± 0.0
RSS	50.0 ± 50.0	100.0 ± 0.0	100.0 ± 0.0	100.0 ± 0.0	100.0 ± 0.0
ULA	93.3 ± 13.3	100.0 ± 0.0	100.0 ± 0.0	100.0 ± 0.0	100.0 ± 0.0

Table 7
Paired statistical analysis of the results in Table 4. For the classification accuracy of each frog species, the paired Student t-test was conducted (Tanton, 2005).

Pairs	Test results
Delta-AWSCCs - Avg AWSCCs	$t = 1.95$ (not significant)
Delta-AWSCCs - MWSCCs	$t = 3.41$ (significant at $p < 0.01$, $df = 17$)
Delta-AWSCCs - MFCCs	$t = 2.91$ (significant at $p < 0.01$, $df = 17$)
Delta-AWSCCs - SFs	$t = 5.52$ (significant at $p < 0.001$, $df = 17$)

5) Perform DCT on the logarithm sub-band energy for dimension reduction and obtain the feature AWSCCs.

$$AWSCCs(d) = \sum_{i=1}^{T} logSE_i \cos\left(\frac{d(i-0.5)}{T}\pi\right) \tag{10}$$

where $d = 1, 2, \ldots, d'$, $1 \le d' \le T$, here d' is the dimension of AWSCCs, and set as 12 here. To keep the feature dimension consistency, the dimensions for MFCCs and MWSCCs are also set as 12 in this study, and the detailed steps for extraction can be found in (Bedoya et al., 2014) and (Zhang and Li, 2015).

6) Temporal feature integration
Here, the statistics of all feature vectors over each windowed signal are calculated, which include sum, average, standard deviation, and skewness. With randomly selected five instances for each frog species, the classification accuracy of averaged AWSCCs is higher than other statistics of AWSCCs. Therefore, only averaged AWSCCs are used in the subsequent experiment. To capture the dynamic information of the frog calls, the delta-AWSCCs are also calculated based on the averaged AWSCCs.

2.9. Classification

In this study, the k-nearest neighbour (k-NN) and support vector machine (SVM) classification algorithm are used for frog call classification. The input parameters for each classifier are syllable features (SFs), MFCCs, MWSCCs, and different AWSCCs, and the output is the frog species.

2.9.1. k-nearest neighbours

For the k-NN classifier, an object is classified to the class of the majority of its k-nearest neighbours (Huang et al., 2009). Specifically, frog feature vectors are stored with species labels in the training phase. For the test phase, the distances between an input frog feature vector and all stored vectors are calculated. Then, k closest vectors are used for selecting the most frequent vector as the label. For example, the Euclidean distance between an input feature vector $f_{i,c}$ and one stored feature vector $f_{i,c}$ is calculated as

$$d(i,j) = \sqrt{\sum_{c=1}^{n} \left(f_{i,c} - f_{j,c}\right)^2} \tag{11}$$

where i and j are indices of the feature vector, n means the dimension of the feature vector. Next, k nearest neighbours of the feature vector i are selected based on the Euclidean distance for selecting the most frequent vector as the label. If the following equation is satisfied

$$\frac{1}{k_1}\sum_{j \in s_1} d(i, j(s_1)) \le \frac{1}{k_2}\sum_{j \in s_2} s(i, j(s_2)) \tag{12}$$

where $k = k_1 + k_2$, k_1 is the number of frog species s_1, k_2 is the number of frog species s_2. Here, the input feature vector i will be classified as frog species s_2.

2.9.2. Support vector machines

Due to the high accuracy and superior generalization properties, support vector machines have been widely used for classifying animal sounds (Huang et al., 2009; Acevedo et al., 2009). In this study, the feature set obtained is first selected as training data. Then, the pairs $(v_l^n, L_l^n), l = 1, 2, \ldots, C_l$ are constructed using the selected training data, where C_l is the number of frog instance in the training data, v_l^n is the feature vector obtained from the *l-th* frog instance in the training data,

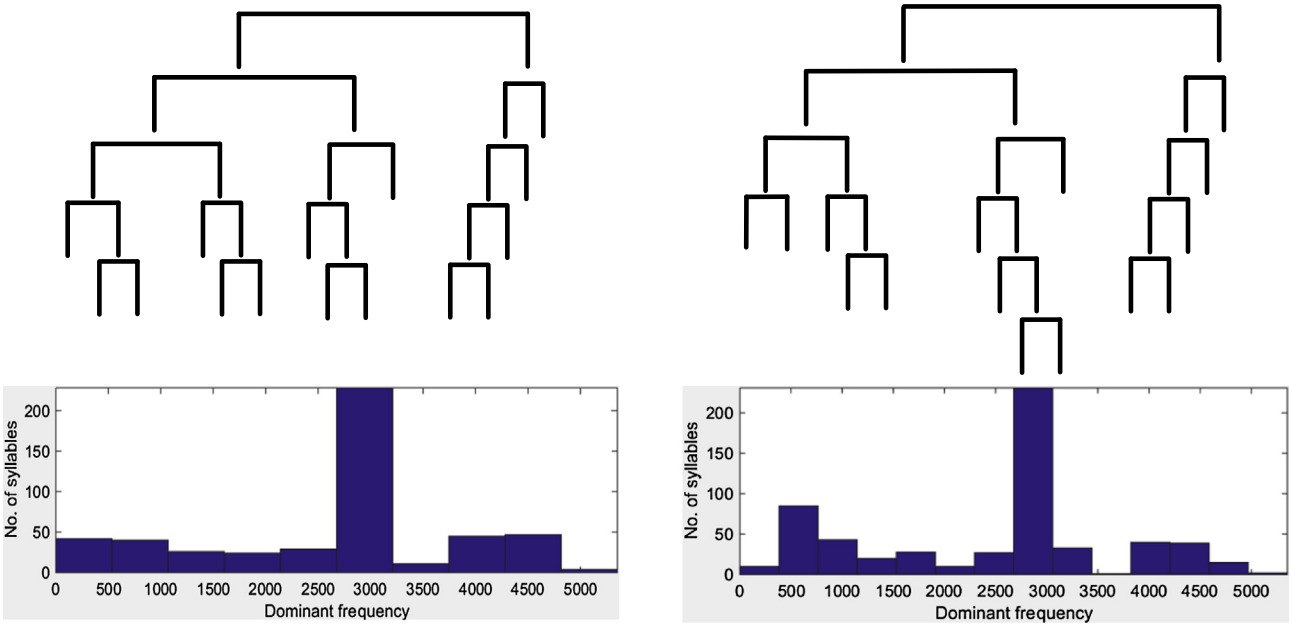


Fig. 8. Wavelet packet tree based on adaptive frequency scale for classifying 10 and 15 frog species.

L_l^n is the frog species label. Furthermore, the decision function for the classification problem based on SVM (Cortes and Vapnik, 1995) is defined by the training data as follows:

$$f(v) = sgn\left(\sum_{sv} \alpha_l^n L_l^n K(v, v_l^n) + b_l^n\right) \tag{13}$$

where $K(.,.)$ is the kernel function, α_l^n is the Lagrange multiplier, and b_l^n is the constant value.

3. Experiment result and discussion

Several experiments are described for evaluating our proposed frog call classification system. First, the parameter tuning is discussed based on the reference data set. Then, the comparisons between all proposed features are studied. Finally, the classification results under different SNR are described.

3.1. Parameter tuning

Five modules for parameter tuning are syllable segmentation, spectral peak track, feature extraction, and classification (Fig. 1).

For syllable segmentation, the window size and overlap are 512 samples and 25%, however, the intensity threshold is 10 dB and 5 dB for the commercial recordings and the JCU recordings respectively.

In the spectral peak track determination, there are seven parameters (see in Table 3). The parameter settings are shows in Table 4.

With a random parameter setting start, an iterative loop is performed for a fixed range of each parameter base on Table 1 to optimise those parameters.

For feature extraction, the window size and overlap are the same for MFCCs, MWSCCs, and AWSCCs using Hamming window, which are 128 samples and 90%, respectively. The dimensions of MFCCs, MWSCCs and AWSCCs are 12. For SFs and delta-AWSCCs, the dimensions are 3 and 24, respectively.

Following prior work (Huang et al., 2009; Han et al., 2011; Xie et al., 2015), the distance function used for k-NN is the Euclidean distance, and k is set as 3. As for the SVM classifier, the Gaussian kernel is used. Parameters α and v are selected independently for each feature set by grid-search using cross validation (Hsu et al., 2003).

3.2. Feature evaluation

All experiments are carried out in Matlab R2014b. Performance statistics are estimated with ten-fold cross validation. Totally, five feature sets including SFs, MFCCs, MWSCCs, and averaged AWSCCs, and delta-AWSCCs, are adopted to two classifiers, which are the k-NN and SVM classifiers. Both k-NN and SVM classifiers are run ten times for evaluating the feature robustness. Due to the non-uniform distribution of the number of syllables for different frog species in the commercial recordings, a weighted classification accuracy is define as follows

$$weighted\ \ Acc = \sum_{i=1}^{N} Acc(i) * \frac{n_i}{N} \tag{14}$$

where n_i is the number of syllables for frog species i, N is the number of syllables for all frog species, Acc is the classification accuracy for that particular frog species.

3.3. Comparison between different feature sets

The classification accuracy comparison for 18 frog species using five feature sets and two classifiers are shown in Table 5.

In this experiment, the best classification accuracy is 99.6%, which is achieved by the delta-AWSCCs with the SVM classifier. Compared with the average AWSCCs, the delta-AWSCCs are slightly improved. One may conjecture that the delta-AWSCCs can capture the dynamic information of the frog calls. For MWSCCs, the averaged classification accuracy of both classifiers is about 2% lower than averaged AWSCCs and delta-AWSCCs with 96.3%. The improvement shows that our proposed adaptive frequency scale can capture more information of frog calls than Mel-scale (Fig. 7).

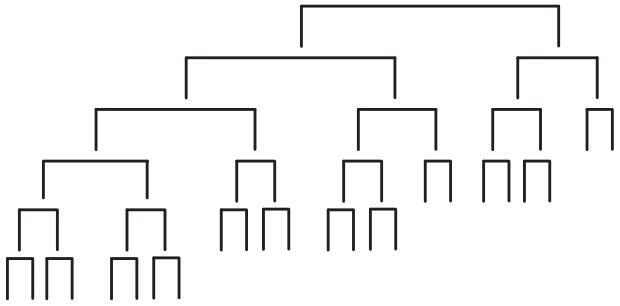


Fig. 9. Mel-scaled wavelet packet tree for frog call classification.

As for SFs and MFCCs, the averaged classification accuracy is much lower than AWSCCs, which is 83.2% and 91.8%, respectively. To explore the reason for the improvement of our proposed feature, the frog call classification accuracy of all frog species is shown in Table 6. However, only the features that use the SVM classifier is shown, because averaged accuracy of the k-NN classifier (93.2%) is lower than the SVM classifier (94.64%).

Table 6 lists the classification accuracy of all 18 frog species with five features. It can be seen from the table that delta-AWSCCs have an accuracy greater than 95% for all frog species. Compared with averaged AWSCCs, the classification accuracy of *Pseudophryne coriacea* (PCA) and *Litoria verreauxii verreauxii* (LVV) are improved to 100%, it might be that the delta-AWSCCs include the dynamic information of frog calls. For *Litoria revelata* (LRA), both the classification accuracy of averaged AWSCCs and delta-AWSCCs are lower than 100%, it is because the dominant frequency is quite similar with multiple frog species including *Assa darlingtoni* (ADI), *Litoria nasuta* (LNA) and *Litoria verreauxii verreauxii* (LVV). However, the classification of *Litoria revelata* (LRA) is 100% using Mel-scale based techniques, because the Mel-scale has a better frequency resolution for *Litoria chloris* (LCS) within its dominant frequency range. In Table 8, the classification accuracy of SFs and MFCCs is lower than other three features, which is only 84.2% and 92.8%, respectively.

The statistical significance of the results is shown in Table 7. The classification accuracy of average AWSCCs is not significantly lower than the delta-AWSCCs. However, the classification accuracy of MWSCCs, MFCCs and SFs is significantly lower than delta-AWSCCs.

Since our wavelet packet tree for feature extraction is obtained based on the frog species to be classified, two more experiments are used for further evaluation. The first experiment is to classify first ten frog species (No.1–10); the second is to classify the first fourteen frog species (No.1–14) (see Table 1). The wavelet packet tree for classifying ten and fourteen frog species is shown in Fig. 8, which is different from the tree for classifying eighteen frog species. However, the Mel-scaled wavelet packet tree is the same for all experiments (see Fig. 9). The classification results are shown in Table 8. Since the classification accuracy

Table 8
Classification accuracy (%) for classifying different number of frog species with four feature sets.

Features	SFs	MFCCs	MWSCCs	Averaged AWSCCs
frog species	84.2 ± 10.5	92.8 ± 11.0	97.6 ± 5.7	99.0 ± 4.6
frog species	89.6 ± 9.7	94.4 ± 8.5	99.2 ± 2.6	100.0 ± 0.0
frog species	94.6 ± 8.7	95.8 ± 8.6	100.0 ± 0.0	100.0 ± 0.0

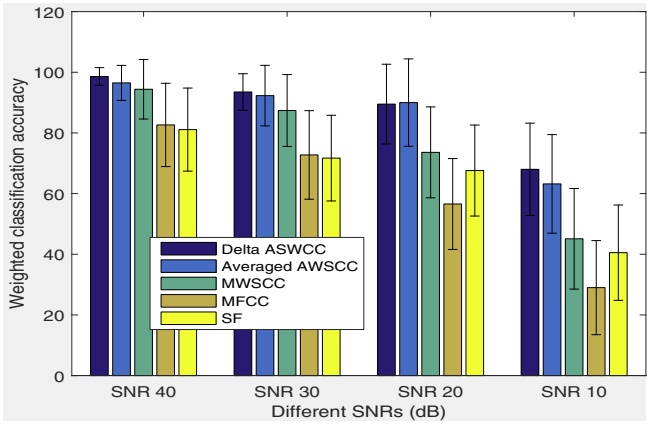


Fig. 10. Sensitivity of five features for different levels of noise contamination.

with averaged AWSCCs is very high for classifying 10 and 14 frog species, the delta-AWSCCs is not included in this experiment. Table 8 shows that averaged AWSCCs can achieve the highest classification accuracy for classifying different number of frog species. Since the averaged AWSCCs is adaptively extracted based on the data, more frog species do not cause a large decrease in the classification accuracy.

3.4. Comparison under different SNRs

To further evaluate the robustness of the proposed feature, a Gaussian noise signal, with SNR of 40 dB, 30 dB, 20 dB, and 10 dB, is added to the original signal. The noise is added after syllable segmentation, because this study focuses on the development of novel features for classification rather than the segmentation method. The classification accuracy with five features under different SNRs is shown in Fig. 10. Compared with MFCCs and MWSCCs, SFs has a stronger anti-noise performance, because the dominant frequency of SFs has a small variation under low SNR. Correspondingly, the adaptive frequency scale also has a small variation, because it is generated by means of applying the k-means clustering algorithm to the dominant frequency. Therefore, our proposed feature has a stronger anti-noise performance than other cepstral features (MFCCs and MWSCCs).

3.5. Proposed feature evaluation using the JCU recordings

Table 9 shows the classification accuracy comparison using our proposed feature to classify 8 frog species obtained from the JCU recordings. Since calls of some frog species in the JCU recordings do not have oscillation structure, SFs are not included for the comparison. Compared with other referred features, our proposed feature also achieves the best classification performance. Since the JCU recordings often have multiple calls from different frog species, spectral peak track occasionally can not capture the specific frog species (labelled species for that syllable) but other frog species to be classified, however, applying k-mean clustering to the dominant frequency calculated from the spectral peak track can reduce this deviation. Therefore, the frequency scale used for the WPD can be accurately achieved which still leads to a high classification accuracy with the proposed feature.

Table 9
Classification accuracy using the JCU recordings.

Feature set	Classification accuracy (%)	
	k-NN	SVM
MFCCs	67.5 ± 13.2	70.8 ± 14.1
MWSCCs	90.4 ± 9.2	91.6 ± 8.7
Averaged AWSCCs	94.1 ± 6.3	94.5 ± 5.8
Delta-AWSCCs	97.0 ± 5.2	97.4 ± 5.4

4. Conclusion and future work

In this study, a novel feature extraction method for frog call classification is developed using the adaptive frequency scaled wavelet packet decomposition. With segmented syllables, spectral peak track is first extracted from each syllable. Then, track duration, dominant frequency, and oscillation rate are calculated based on each track. Next, a k-means clustering algorithm is applied to the dominant frequency, which generates the frequency scale for WPD. Finally, a new feature set, AWSCCs, is calculated. Since our feature extraction method is developed based on the data itself, the wavelet packet tree differs according to the frog species to be classified. Compared with the Mel-scaled WPD tree, the proposed adaptive wavelet packet tree can better fit the dominant frequency distribution of the frog species to be classified. With the proposed frequency scale, the call character of those frog species to be classified can be enhanced, however, the background noise and calls from other animals will be suppressed. Therefore, our proposed feature sets can achieve a higher accuracy for the classification of frog calls than others. Meanwhile, since the frequency scale is calculated based on the dominant frequency of those frog species to be classified, our proposed wavelet tree structure is more accurate and efficiency in classifying the frog calls when compared with Mel-scale (Figs. 8 and 9).

As for the feature extraction algorithm, it is designed for classifying frog calls. For frog calls, the typical structure in a spectrogram is frequency contour (named spectral peak track in this study) that are within a given frequency range starting at a given time (Mellinger et al., 2011). For other organisms that have similar frequency contour structures such as the whistles of dolphins, chirps of birds (Chen and Maher, 2006), spectral peak tracks can also be extracted from the spectrograms of their calls. Based on those spectral peak tracks, dominant frequency can be calculated. For the subsequent analysis, we can calculate the features using the same process as described in this study. For those organisms without clear frequency contour structure, this proposed method can also be used by enhancing the frequency contour structure, which can be realized by applying a small window size and a large window overlap to the recording waveform.

For future work, the oscillation rate is calculated based on the spectrogram, which is generated by applying STFT to the waveform. However, when the temporal gap is smaller than the window size used for STFT, the oscillation structure will disappear. Therefore, finding new techniques for translating the 1-D signal to 2-D signal is our future direction. Since the frequency scale is generated based on the dominant frequency, this technique can be applied to other organisms that have clear frequency contour structure. Modifying this algorithm to those organisms without a clear frequency contour structure needs to be solved. We also plan to include additional experiments that test a wider variety of audio data from different geographical and environment conditions. Other animal calls such as birds, insects, and whales can also be studied. Furthermore, we will explore the idea of developing new features based on the data itself.

Acknowledgement

Thanks to the QUT Eco-acoustics Research Group for providing the datasets used in this experiment, as well as to the support from the Wet Tropics Management Authority, Queensland, Australia. Thanks to the anonymous reviewers for their careful work and thoughtful suggestions that have helped improve this paper substantially.

All funding for this research was provided by the Queensland University of Technology and the China Scholarship Council (CSC).

References

Acevedo, M.A., Corrada-Bravo, C.J., Corrada-Bravo, H., Villanueva-Rivera, L.J., Aide, T.M., 2009. Automated classification of bird and amphibian calls using machine learning: A comparison of methods. Ecol. Inf. 4, 206–214.

Bedoya, C., Isaza, C., Daza, J.M., López, J.D., 2014. Automatic recognition of anuran species based on syllable identification. Ecol. Inf. 24, 200–209.

Biswas, A., Sahu, P., Chandra, M., 2014. Admissible wavelet packet features based on human inner ear frequency response for hindi consonant recognition. Comput. Electr. Eng. 40, 1111–1122.

Chen, Z., Maher, R.C., 2006. Semi-automatic classification of bird vocalizations using spectral peak tracks. J. Acoust. Soc. Am. 120, 2974–2984.

Chen, W.P., Chen, S.S., Lin, C.C., Chen, Y.Z., Lin, W.C., 2012. Automatic recognition of frog calls using a multi-stage average spectrum. Comput. Math. Appl. 64, 1270–1281.

Clauzel, C., Bannwarth, C., Foltete, J.C., 2015. Integrating regional-scale connectivity in habitat restoration: An application for amphibian conservation in eastern france . J. Nat. Conserv. 23, 98–107. http://dx.doi.org/10.1016/j.jnc.2014.07.001.

Colonna, J., Ribas, A., dos Santos, E., Nakamura, E., 2012. Feature subset selection for automatically classifying anuran calls using sensor networks. Neural Networks (IJCNN), The 2012 International Joint Conference on, pp. 1–8 http://dx.doi.org/10.1109/IJCNN.2012.6252794.

Cortes, C., Vapnik, V., 1995. Support-vector networks. Mach. Learn. 20, 273–297.

Farooq, O., Datta, S., 2001. Mel filter-like admissible wavelet packet structure for speech recognition. IEEE Signal Process Lett. 8, 196–198.

Gage, S.H., Axel, A.C., 2014. Visualization of temporal change in soundscape power of a michigan lake habitat over a 4-year period. Ecol. Inf. 21, 100–109.

Garcia, R.A., Cabeza, M., Rahbek, C., Araújo, M.B., 2014. Multiple dimensions of climate change and their implications for biodiversity. Science 344, 1247579.

Gingras, B., Fitch, W.T., 2013. A three-parameter model for classifying anurans into four genera based on advertisement calls. J. Acoust. Soc. Am. 133, 547–559.

Han, N.C., Muniandy, S.V., Dayou, J., 2011. Acoustic classification of australian anurans based on hybrid spectral-entropy approach. Appl. Acoust. 72, 639–645.

Harma, A., 2003. Automatic identification of bird species based on sinusoidal modeling of syllables. Acoustics, Speech, and Signal Processing, 2003. Proceedings.(ICASSP'03). 2003 IEEE International Conference on, IEEE., pp. V–545

Heller, J.R., Pinezich, J.D., 2008. Automatic recognition of harmonic bird sounds using a frequency track extraction algorithm. J. Acoust. Soc. Am. 124, 1830–1837. http://dx.doi.org/10.1121/1.2950085.

Hsu, C.W.W., Chang, C.C., Lin, C.J., et al., 2003. A practical guide to support vector classification.

Huang, C.J., Yang, Y.J., Yang, D.X., Chen, Y.J., 2009. Frog classification using machine learning techniques. Expert Syst. Appl. 36, 3737–3743.

Jancovic, P., Kokuer, M., 2015. Acoustic recognition of multiple bird species based on penalized maximum likelihood. IEEE Signal Process Lett. 22, 1585–1589. http://dx.doi.org/10.1109/LSP.2015.2409173.

Litvin, Y., Cohen, I., 2011. Single-channel source separation of audio signals using bark scale wavelet packet decomposition. J. Signal Proc. Sys. 65, 339–350.

Mellinger, D.K., Martin, S.W., Morrissey, R.P., Thomas, L., Yosco, J.J., 2011. A method for detecting whistles, moans, and other frequency contour sounds. J. Acous. Soc. Am. 129, 4055–4061.

Melter, R.A., 1987. Some characterizations of city block distance. Pattern Recogn. Lett. 6, 235–240.

Ren, Y., Johnson, M.T., Tao, J., 2008. Perceptually motivated wavelet packet transform for bioacoustic signal enhancement. J. Acoust. Soc. Am. 124, 316–327.

Roch, M.A., Brandes, T.S., Patel, B., Barkley, Y., Baumann-Pickering, S., Soldevilla, M.S., 2011. Automated extraction of odontocete whistle contours. J. Acoust. Soc. Am. 130, 2212–2223.

Sahidullah, M., Saha, G., 2012. Design, analysis and experimental evaluation of block based transformation in mfcc computation for speaker recognition. Speech Comm. 54, 543–565. http://dx.doi.org/10.1016/j.specom.2011.11.004.

Selin, A., Turunen, J., Tanttu, J.T., 2007. Wavelets in recognition of bird sounds. EURASIP J. Appl. Signal Proc. 2007, 141-141.

Shine, R., 2014. A review of ecological interactions between native frogs and invasive cane toads in australia. Austral Ecol. 39, 1–16.

Stewart, D., 1999. Australian frog calls: subtropical east. Audio CD . URL: http://www.naturesound.com.au/cd_frogsSE.htm.

Tanton, J.S., 2005. Encyclopedia of mathematics. Facts On File.

Towsey, M., Planitz, B., Nantes, A., Wimmer, J., Roe, P., 2012. A toolbox for animal call recognition. Bioacoustics 21, 107–125.

Wimmer, J., Towsey, M., Planitz, B., Roe, P., Williamson, I., 2010. Scaling acoustic data analysis through collaboration and automation. e-Science (e-Science), 2010 IEEE Sixth International Conference on, pp. 308–315 http://dx.doi.org/10.1109/eScience.2010.17.

Wimmer, J., Towsey, M., Planitz, B., Williamson, I., Roe, P., 2013. Analysing environmental acoustic data through collaboration and automation. Futur. Gener. Comput. Syst. 29, 560–568.

Xie, J., 2016. Widelife acoustic. http://www.wildlifeacoustics.com/products/song-meter-sm2-birds.

Xie, J., Towsey, M., Truskinger, A., Eichinski, P., Zhang, J., Roe, P., 2015. Acoustic classification of australian anurans using syllable features. 2015 IEEE Tenth International Conference on Intelligent Sensors, Sensor Networks and Information Processing (IEEE ISSNIP 2015), Singapore, Singapore.

Yen, G.G., Fu, Q., 2002. Automatic frog call monitoring system: a machine learning approach. AeroSense 2002, International Society for Optics and Photonics., pp. 188–199

Yen, G.G., Lin, K.C., 2000. Wavelet packet feature extraction for vibration monitoring. IEEE Trans. Ind. Electron. 47, 650–667.

Zhang, X., Li, Y., 2015. Adaptive energy detection for bird sound detection in complex environments. Neurocomputing 155, 108–116.

Zhang, J., Huang, K., Cottman-Fields, M., Truskinger, A., Roe, P., Duan, S., Dong, X., Towsey, M., Wimmer, J., 2013. Managing and analysing big audio data for environmental monitoring. Computational Science and Engineering (CSE), 2013 IEEE 16th International Conference on, pp. 997–1004.