Acoustic feature extraction using perceptual wavelet packet decomposition for frog call classification

Jie Xie, Michael Towsey, Philip Eichinski, Jinglan Zhang, Paul Roe
Faculty of Science and Technology, Queensland University of Technology, Brisbane, QLD 4001, Australia
Email: j3.xie@student.qut.edu.au
{m.towsey, philip.eichinski, jinglan.zhang, p.roe}@qut.edu.au

***Abstract*—Frog protection has become increasingly essential due to the rapid decline of its biodiversity. Therefore, it is valuable to develop new methods for studying this biodiversity. In this paper, a novel feature extraction method is proposed based on perceptual wavelet packet decomposition for classifying frog calls in noisy environments. Pre-processing and syllable segmentation are first applied to the frog call. Then, a spectral peak track is extracted from each syllable if possible. Syllable duration, dominant frequency and oscillation rate are directly extracted from the track. With k-means clustering algorithm, the dominant frequency of all syllables is clustered into k parts. Based on the centroids of the clustering result, wavelet packet decomposition (WPD) is applied to the frog calls for ensuring one node contains only one centroid. Based on the WPD coefficients, a new feature set named perceptual wavelet packet decomposition sub-band cepstral coefficients (PWSCC) is extracted. Finally, a k-nearest neighbour classifier is used for the classification. The experiment results show that these proposed features can achieve an average classification accuracy of 96.40% which outperforms Mel-frequency cepstral coefficients feature (MFCCs) (92.12%).**

***Index Terms*—frog call classification; wavelet packet decomposition; spectral peak track; k-means clustering**

I. Introduction

Currently, due to habitat loss, invasive species and climate change, global biodiversity is rapidly decreasing. Therefore, it is becoming ever more important to monitor biodiversity. Frogs have been widely used as an indicator of biodiversity because of their sensitivity to the environmental change, and as such monitoring of frog species must increase. [1]. Through the study of frog species distributions, we can then predict the state of the environment.

Due to the development of acoustic sensor techniques, lots of sensors have been widely deployed in nature for monitoring biodiversity, which produces large volumes of acoustic data. Compared with the traditional method [2] [3], acoustic sensor can help collect data across large areas for extended periods making them attractive in biodiversity monitoring [4]. With collected acoustic data, data analysis techniques are then incorporated to assist ecologists to study frogs.

Many studies have investigated the recognition or classification of animal calls. Classification systems are most commonly structured as follows: (1) Pre-processing, (2) Syllable segmentation, (3) Feature extraction, (4) Classification. Following this general classification workflow, frog call classification has been addressed in several papers. Huang et al combined spectral centroid, signal bandwidth, and thresholding crossing rate for the classification of frog calls with the k-nearest neighbour (k-NN) and support vector machine (SVM) classifier [5]. Huang's work studied the machine learning techniques for frog sound classification and developed online frog sound identification system. Han et al. introduced a k-NN classifier to classify frog calls with Fourier spectral centroid, Shannon entropy, and R*é*nyi entropy [6]. This method introduced the entropy information as the bioacoustic features for the animal sound classification. Since the time-varying information of frog calls has not been addressed for classification in prior work, a multi-stage average spectrum method was proposed by Chen et al. for frog classification [7]. Gingras et al. used a logistic regression model for classifying anurans. Three parameters, mean value for dominant frequency, coefficient of variation of root-mean square energy, and spectral flux, were combined for anuran classification [8]. Bedoya et al. developed a method for recognition of anuran species based on syllable identification. A fuzzy classifier and Mel-frequency cepstral coefficients were combined for the recognition [9], which can classify the species not presented in the training data. All those prior work extracted corresponding features from the short-time Fourier transform (STFT) results. However, there is a trade-off between time and frequency resolution of STFT, which restricts the discriminability of the features.

In this paper, we propose a novel frog call classification method using perceptual wavelet packet decomposition (WPD). Rather than applying WPD based on the particular levels and some auditory scales like equivalent rectangular bandwidth (ERB) scale [10], Mel-scale [11], Bark-scale [12], the scale used for the WPD is based on the dataset here. After pre-processing and segmentation, spectral peak track is first extracted from each syllable if possible. Then track duration, dominant frequency and oscillation rate are extracted from spectral peak track to make syllable features. For extracting perceptual wavelet packet decomposition sub-band cepstral coefficients (PWSCC), dominant frequency is first extracted and clustered into k parts with k-means clustering algorithm. Then WPD is applied to the frog calls for ensuing that one node contains only one centroid. Finally, PWSCC is extracted based on the WPD coefficients. A k-NN classifier is used for the classification and PWSCC achieves higher classification accuracy (96.4%) than MFCCs (92.12%) including syllable duration, dominant frequency and oscillation rate.

The rest of this paper is organized as follows. Section II reviews related work. Section III describes the spectrogram analysis. Section IV introduces the proposed system. Section V reports the experiment results. Conclusions are drawn in Section VI.

II. Related work

Wavelet analysis has been widely used for the analysis of audio data due to its better ability in time and frequency resolution. Selin et al. introduced WPD for the recognition of bird calls. WPD was first used for the signal decomposition and construction of the time-frequency representation. Then four features were calculated: maximum energy, position, spread and width. These were combined with two neural networks for classification [13]. Based on WPD, Zhang et al. developed a modified feature set named Mel-scaled wavelet packet decomposition sub-band cepstral coefficient for bird sound detection [11]. Sahu et al. proposed the auditory ERB like admissible wavelet packet features for the TIMIT phoneme recognition. Based on the wavelet packet tree, energy, delta and acceleration features per frame were obtained for the final recognition [14].

Auditory scales used for WPD are all derived for different reasons. Mel scale is a perceptual scale of pitches judged by listeners to be equal in distance from one another. Bark scale is proposed for the analysis of psychoacoustical. ERB scale is also used in psychoacoustics, which gives an approximation to the bandwidths of the filters in human hearing. However, for frog call classification, it is important to find the suitable scale which is suitable for WPD.

For the frog, the advertisement calls of closely related species are more similar than those of distant species, hence the dominant frequency that strongly correlated with the advertisement call can be utilised for analysing frog calls [8]. Using spectral peak track extraction method, we can achieve the dominant frequency of all syllables. Then k-means clustering algorithm is applied to the dominant frequency for getting prior information, which can be further used for WPD.

III. Data description and Spectrogram analysis

In this study, 10 frog species which are widely spread in Queensland, Australia are selected for the experiment (Table I). All the recordings were made by David Steward, and have a sample rate of 44.1 kHz. Each recording only includes one frog species, and the minimal and maximal duration for those recordings are 21 and 55 seconds.

We first manually inspected spectrograms of three randomly selected examples of calls for each of the frog species. Three parameters were measured for each of the three examples and averaged, as listed in Table II. Those parameters are used as priori information for further analysis. It is worth to mention that those selected example calls are excluded from the dataset for experiment.

IV. System framework

In this study, frog call classification system consists of four subsections: pre-processing, syllable segmentation, feature extraction, classification (Fig.1). Detailed information of each stage is described in the following sections.

TABLE I: Summary of frog scientific name, common name, and code

Scientific name	Total syllable	Common name	Code
Crinia parinsignifera	32	Eastern sign-bearing frog	CPA
Litoria caerulea	65	Whites tree frog	LCA
Litoria chloris	31	Red-eyed tree frog	LCS
Litoria latopalmata	171	Broad-palmed frog	LLA
Litoria nasuta	73	Striped rocket frog	LNA
Mixophyes fasciolatus	32	Great barred frog	MFS
Mixophyes fleayi	27	Fleay's barred Frog	MFI
Neobatrachus sudelli	22	Painted burrowing frog	NSI
Uperoleia fusca	39	Dusky toadlet	UFA
Uperoleia laevigata	24	Smooth toadlet	ULA

TABLE II: Averaged frog parameters based on the visual inspection of the spectrogram, an asterisk denotes that frog species need spectrogram smoothing

Species code	Averaged syllable duration (millisecond)	Averaged peak frequency (Hz)	Averaged oscillation rate (cycle/second)
CPA	250	4300	350
LCA	500	500	50
LCS	800	1700	220
LLA	30	1400	2100
LNA	100	2800	160
MFS	200	1200	140
MFI	50	1000	140
NSI	480	1200	20
UFA*	550	2300	40
ULA*	450	2400	150

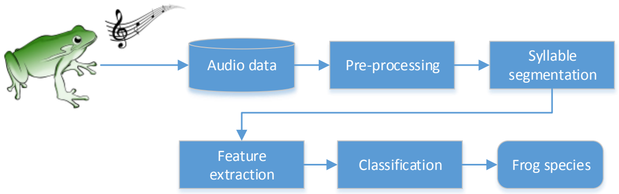


Fig. 1: Flowchart of frog call classification system.

A. Pre-processing

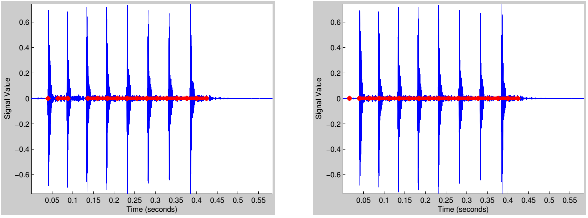
We first re-sampled recordings at 16 kHz per second and mixed them to mono in order to reduce computational burden. A spectrogram was then generated by applying short-time Fourier transform (STFT) to each recording. Specifically, each recording was divided into frames of 128 samples with 85% frame overlap. A fast Fourier transform was then preformed on each frame with a Hamming window, which yielded amplitude values for 64 frequency bins, each spanning 125 Hz. The final decibels (dB) were generated using $dB = 10 * log_{10}(A)$, where A was the amplitude value. Here the spectrogram is generated for the spectral peak track extraction rather than for segmentation. Noise reduction here was performed by spectral subtraction [15], which is an essential step for improving the classification result.

```
Algorithm 1: Modified Spectral Subtraction
Data: $S = S(T, F)$, Original spectrogram.
Result: $S^{'} = S^{'}(T, F)$, Noise reduced spectrogram.
begin
    for $f \in F$ do
        1. calculate the histogram of the intensity value
        2. smooth the histogram array with a moving
        average window of size 7
        3. regard the modal noise intensity at the position
        of maximal bin in the left-side of the histogram
    Construct the array of the modal noise values for all
    frequency bins;
    Smooth the array with a moving average filter with
    window of size 5;
    for $f \in F$ do
        1. subtract the modal noise intensity
        2. truncated negative decibel values to zero
```

B. Syllable segmentation

The elementary unit of frog vocalizations is the *syllable*, which can be utilised for species recognition. In this study, H$\ddot{a}$rm$\ddot{a}$'s method was used for syllable segmentation [16]. This syllable segmentation method is based on the iterative amplitude-frequency information. the detailed description of the algorithm can be found in our previous paper [17] Here the intensity threshold used is 20 dB here.

Different from the original method , we add an optional processing step which is spectrogram smoothing before H$\ddot{a}$rm$\ddot{a}$'s method. For those frog species (such as Neobatrachus sudelli) that contain large gaps within one syllable, it is necessary to do the smoothing. Here, we use Gaussian filter (7 × 7) to smooth the spectrogram. The size 7 × 7 is selected based on the averaged oscillation rate information in Table II. The segmentation results with and without smoothing are shown in Fig.2.



(a) before smoothing

(b) after smoothing

Fig. 2: Syllable segmentation results marked with red line (one syllable)

C. Spectral peak track extraction

For the frog, the advertisement calls of related frog species are more similar that those of distant species, therefore, the dominant frequency that strongly high correlated with the advertisement call can be utilised for analysing frog calls [8]. In this study, spectral peak track (SPT) is explored to represent the dominant frequency trace of frog calls. There are two reasons for using SPT: (1) Isolate the desired signal from background noise, (2) Extract corresponding features based on SPT. The method for extracting SPT is a simplified version of the method by Roch et al [18]. Different from the original method, we use the linear regression to connect peaks into the track. Then, corresponding parameters are pre-defined to decide whether or not keep the track.

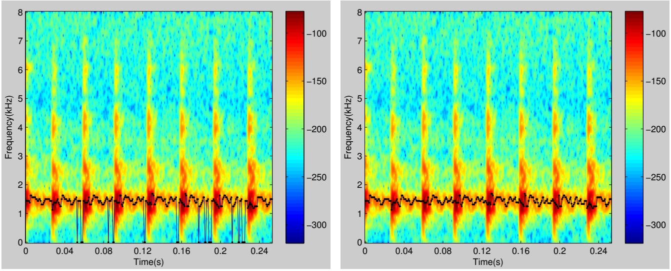
The SPT extraction algorithm requires seven parameters, which are explained in Table III. The process for selecting these parameters is explained in section V.

TABLE III: Parameters used to spectral peak extraction

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

The SPT algorithm is described as follows.

The SPT results are shown in Fig.3. During the detection of tracks, gaps in tracks are created where the minimal intensity value I is not reached. These gaps are filled by predicting the correct frequency bin using linear regression, as illustrated in Fig.3 (b).



(a) original spectral peak track

(b) adding predicted peaks using linear regression

Fig. 3: Spectral peak track extraction results

D. Perceptual wavelet packet decomposition

For frogs, dominant frequency has been used as an important parameter for frog call classification [8] [17] . Therefore, the frequency distribution can be a good feature for classifying frogs. Currently, there are several methods for calculating frequency distributions based on the scales [11] [14]. However, those scales are not designed for frogs. Here we design a method for calculating the frequency distribution based on the frog information (dominant frequency).

1) wavelet packet decomposition: Wavelet packet decomposition (WPD) is a wavelet transform where the discrete-time (sampled) signal is passed through more filters than the discrete wavelet transform. It can be used to decompose

Algorithm 2: Spectral Peak Track Extraction

```
Data: $S' = S'(T,F)$, $I$, $T_c$, $T_s$, $f_c$, $d_{min}$, $d_{max}$, $\beta$.
Result: $Track(N) = \{t_s, t_e, f_t(t_s \le t \le t_e)\}$, Spectral
        peak track.
begin
    Step 1: find maximum intensity value of each frame
    and produce the peak matrix $M(T,F)$
    for $t \in T$ do
        select the maximum intensity value $v$
        if $v \le I$ then
            $M(t,f) = 0$
        else
            $M(t,f) = I$
    Step 2: produce initial spectral peak track
    while $t_i \le T$ do
        while $t_j \le T$ do
            if $t_i - t_j \le I$ then
                $Track(1) = \{t_i, t_j, f_t(t_i \le t \le t_j)\}$
                break
    Step 3: spectral peak track extraction
    for $t \in T$ do
        1. repeat linear regression algorithm to
        recalculate the next predicted peak using at most
        the last 10 included peaks
        2. stop the iterative process until $t - t_e \ge T_s$
        3. calculate the duration $d$ and density $y$
        4. if $d \ge d_{min}$, $d \le d_{max}$ and $y \ge \beta$ then
            save current track to the track list $Track(N)$
        else
            discard the track
```

a signal into sub-bands with low frequency (approximation parts) and high frequency (detail parts) simultaneously [19]. Both the detail and approximation coefficients are decomposed to create the full binary tree. Therefore, the WPD has the same frequency bandwidth for each resolution.

Dominant frequency is an important parameter for recognising frog species, along with frequency distributions. For better capturing the frequency information, we decompose the frog call using the derived dominant frequency information in section E. According to the number of frog species we try to classify, we repeat K-mean clustering algorithm 10 times to generate the information for WPD. Here K is 10, the distance function is $city\,block$ function. 10 centroids ($C_i(i = 1 : 10)$) of the clustering result are saved for generating the scale for WPD.

2) Perceptual model: Based on the clustering result, we proposed an automated wavelet packet decomposition method (Algorithm 3). Different from the fixed frequency band scale, the frequency band scale of our WPD is motivated by the dataset, which means better discriminative ability and more robust in the complex environment. The wavelet packet decomposition result is show in Fig.4.

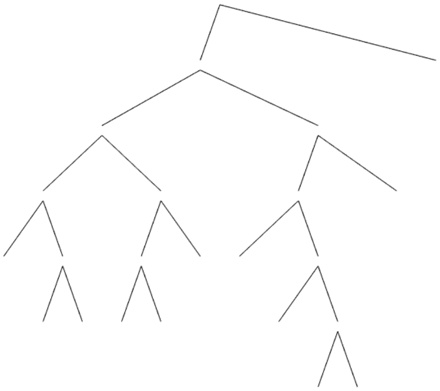


Fig. 4: Tree of perceptual WPD.

Algorithm 3: Perceptual WPD method

```
Data: $C_i(i = 1 : 10)$, $F_s$.
Result: Perceptual wavelet packet decomposition process
begin
    Step 1: sort the centroid $C$ and calculate the
    difference between the consecutive vectors of $C$, the
    result is saved in $D_j(j = 1 : 9)$
    Step 2: calculate the initial decomposition level $L$.
    $F_s/min(D) \ge 2^{L+1}$
    Here, L is the minimum integer.
    Step 3: do the wavelet packet decomposition
    for $l = 1 : L$ do
        1. calculate the frequency resolution of level l
        for $i = 1 : 10$ do
            1: put the $C_i$ into the right frequency band
            2: count the number of $C_i$ in each band ($n$)
        if $n \ge 2$ then
            do decomposition to that particular node
        else
            stop decomposition;
```

E. Feature extraction

1. Syllable duration

$$SD = (t_e - ts)/r_x \tag{1}$$

where r_x is the x-axis resolution, and it is 845.68 frame per second.

2. Dominant frequency

$$DF = \sum_{t=t_s}^{t_e} \frac{f_t}{N} \tag{2}$$

where N is the number of frames.

3. Oscillation rate

First, we calculate the power in the frequency domain boundary $[l, h]$, here $[l, h] = [max(f - 5, 1), f + 5]$, where f is the dominant frequency bin. Then we do the autocorrelation of the power and apply a discrete cosine transform to the mean subtraction of correlation result. Finally, the oscillation rate is calculated as

$$OR = \frac{p_{max}}{SD} * \gamma / r_x \quad (3)$$

Where γ is set as 0.5, p_{max} is the location information of the the higher power.

4. Perceptual wavelet packet decomposition sub-band cepstral coefficients (PWSCC)

Based on the perceptual wavelet packet decomposition, we extract perceptual wavelet packet decomposition sub-band cepstral coefficients for frog call classification, which is similar with the procedure of MFCC. Here MFCC is used as the baseline for comparison.

The steps for calculating the PWSCC are as follows:

Step 1. Add hamming window to each frog syllable.

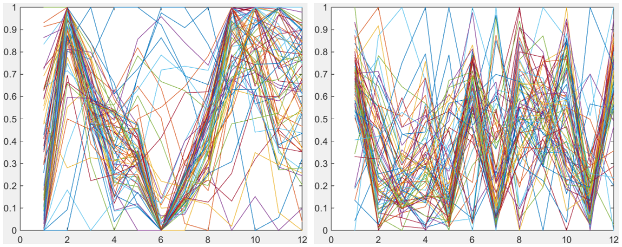
Step 2. Perform the perceptual WPD as described in subsection D and the wavelet base function used here is 'db4'.

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

Step 4. Normalise the energy for each sub-band.

Step 5. Apply DCT on the logarithm sub-band energy and select 12 coefficients as the final feature PWSCC.

The comparison of PWSCC and MFCC is shown in Fig.5 .



(a) MFCC

(b) PWSCC

Fig. 5: Feature comparison of MFCC and PWSCC.

V. Experiments and Results

In this part, several experiments are made for evaluating our proposed approach. First, the validation set is used for parameter tuning. Then, we compare the frog call classification accuracy between syllable features (SF) including syllable duration, dominant frequency and oscillation rate, MFCC and PWSCC. We also study the classification accuracy under different signal to noise ratio (SNR).

A. Parameter tuning

There are three modules including syllable segmentation, spectral peak track extraction, and feature extraction, whose parameters need to be discussed.

For syllable segmentation, the window size and overlap are 512 samples and 0.25, The window function is $Kaiser$ window. The intensity threshold for stopping criteria used is 20 dB, the segmentation result is sensitive with this value, which needs to be tuned.

Spectral peak track extraction algorithm has seven parameters, and all those parameters are pre-defined based on the manually inspection result of Table II. Here minimal duration and maximal duration are 40 ms and 1000 ms. The density value is 0.8, which describes the integrity of one frog call syllable. The minimal intensity value is 3 dB. The maximal time interval for connecting peaks is 1.5 ms, the minimal time interval for stopping growing tracks is 4 ms. The maximal frequency interval is 520 Hz. Here seven parameters need to be pre-defined, then the algorithm can work well.

For MFCC and PWSCC, window size and overlap are the same, which are 128 samples and 0.85, the window function used is $Hamming$ window.

B. Classification

In this study, the k-NN classifier is used to learn a model on the training examples with 10-fold cross-validation. For evaluating the robustness of our proposed feature, the k-NN classifier is run 10 times for each classification task. The classification performance is defined as follows:

$$Classification(\%) = \frac{N_c}{N_t} \quad (4)$$

where N_c is the number of correctly classified instance, N_t is the total number of instance.

Following prior work [6] [5], the distance function of the k-NN classifier is Euclidean function, the number of neighbour, K is 5. The classification results using SF, MFCC, and PWSCC are displayed in Fig.6. Overall classifier accuracy for k-NN is 89.68%, 93.51% and 96.71% . With SF, the classification accuracy of MFCC and PWSCC is shown in Table.IV.

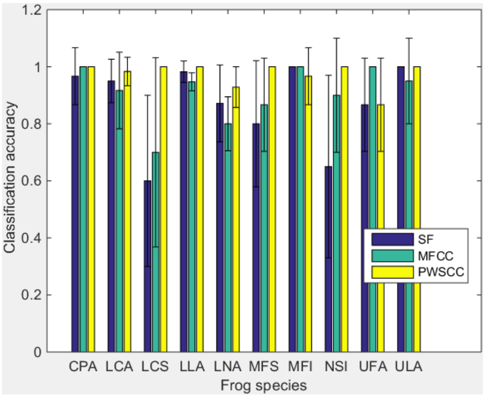


Fig. 6: Classification accuracy of 10 frog species.

TABLE IV: Overall classification accuracy.

Feature	***SF***	***MFCC***	***PWSCC***
Without SF	86.87%	90.80%	97.45%
With SF	NA	96.15%	97.95%

For further testing the robustness of PWSCC, a Gaussian white noise signal, with signal to noise ratio (SNR) of 40 dB, 30 dB, 20 dB, 10 dB , was added to the audio data. The results are shown in Fig.7. It is worth to mention that the noise was added to the signal after syllable segmentation. The classification accuracy of different SNRs shows the robust of our proposed feature PWSCC.

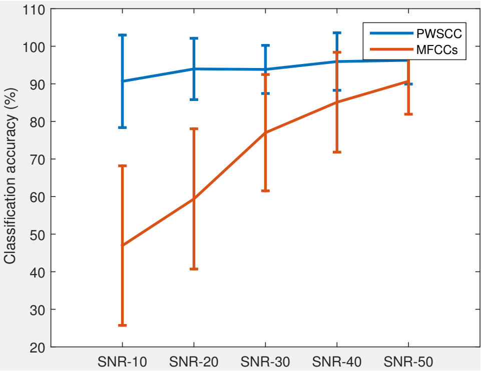


Fig. 7: Sensitivity of MFCC and PWSCC feature for different levels of noise contamination.

Here the classification accuracy of MFS is relatively lower than other frog species, because the dominant frequency of MFS is similar with LLA and MFI and NSI. For CPA, the classification accuracy is high due to the dominant frequency difference with others. For further improving classification accuracy, we add three syllable features. The averaged classification accuracy of MFCC and PWSCC are improved by 2.00% and 1.60% respectively. However, the classification accuracy of some particular frog species is descended due to the similarity of syllable features with other frog species, such as CPA and LNA. The results from running the classifier on audio data with added artificial noise show the ability of our proposed feature for addressing the background noise.

VI. Conclusion

We propose a novel frog call classification method based on perceptual wavelet packet decomposition. The audio data is first pre-processed and segmented into syllables. Then spectral peak track is extracted for getting the priori information, which can be used for wavelet packet decomposition. Finally, a new acoustic feature set named PWSCC is calculated for frog call classification with a k-NN classifier. Experiment results are promising with an average classification of 96.71% including syllable features. Future work will focus on a wider frog call database, including a larger number of frog species, and frog calls from different geographical and environment conditions. We will also extend this work for classifying other animal species such as birds, whales.

Acknowledgment

The authors would like to thank the China Scholarship Council (CSC) for the support. Any opinions, findings and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the CSC.

References

[1] J. G. Burkhart, G. Ankley, H. Bell, H. Carpenter, D. Fort, D. Gardiner, H. Gardner, R. Hale, J. C. Helgen, P. Jepson *et al.*, "Strategies for assessing the implications of malformed frogs for environmental health." *Environmental Health Perspectives*, vol. 108, no. 1, p. 83, 2000.

[2] C. J. Ralph, J. R. Sauer, and S. Droege, *Monitoring bird populations by point counts*. DIANE Publishing, 1998.

[3] R. Heyer, M. A. Donnelly, M. Foster, and R. Mcdiarmid, *Measuring and monitoring biological diversity: standard methods for amphibians*. Smithsonian Institution, 2014.

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

[5] 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.

[6] N. C. Han, S. V. Muniandy, and J. Dayou, "Acoustic classification of australian anurans based on hybrid spectral-entropy approach," *Applied Acoustics*, vol. 72, no. 9, pp. 639–645, 2011.

[7] W.-P. Chen, S.-S. Chen, C.-C. Lin, Y.-Z. Chen, and W.-C. Lin, "Automatic recognition of frog calls using a multi-stage average spectrum," *Computers & Mathematics with Applications*, vol. 64, no. 5, pp. 1270–1281, 2012.

[8] B. Gingras and W. T. Fitch, "A three-parameter model for classifying anurans into four genera based on advertisement calls," *The Journal of the Acoustical Society of America*, vol. 133, no. 1, pp. 547–559, 2013.

[9] C. Bedoya, C. Isaza, J. M. Daza, and J. D. López, "Automatic recognition of anuran species based on syllable identification," *Ecological Informatics*, vol. 24, pp. 200–209, 2014.

[10] P. Sahu, A. Biswas, A. Bhowmick, and M. Chandra, "Auditory erb like admissible wavelet packet features for timit phoneme recognition," *Engineering Science and Technology, an International Journal*, vol. 17, no. 3, pp. 145 – 151, 2014.

[11] X. Zhang and Y. Li, "Adaptive energy detection for bird sound detection in complex environments," *Neurocomputing*, vol. 155, no. 0, pp. 108 – 116, 2015.

[12] Y. Litvin and I. Cohen, "Single-channel source separation of audio signals using bark scale wavelet packet decomposition," *Journal of Signal Processing Systems*, vol. 65, no. 3, pp. 339–350, 2011.

[13] A. Selin, J. Turunen, and J. T. Tanttu, "Wavelets in recognition of bird sounds," *EURASIP Journal on Applied Signal Processing*, vol. 2007, no. 1, pp. 141–141, 2007.

[14] P. Sahu, A. Biswas, A. Bhowmick, and M. Chandra, "Auditory erb like admissible wavelet packet features for timit phoneme recognition," *Engineering Science and Technology, an International Journal*, vol. 17, no. 3, pp. 145–151, 2014.

[15] M. W. Towsey, B. Planitz, A. Nantes, J. Wimmer, and P. Roe, "A toolbox for animal call recognition," *Bioacoustics : The International Journal of Animal Sound and its Recording*, vol. 21, no. 2, pp. 107–125, February 2012.

[16] A. Harma, "Automatic identification of bird species based on sinusoidal modeling of syllables," in *Acoustics, Speech, and Signal Processing, 2003. Proceedings.(ICASSP'03). 2003 IEEE International Conference on*, vol. 5. IEEE, 2003, pp. V–545.

[17] 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*, April 2015, pp. 1–6.

[18] M. A. Roch, T. S. Brandes, B. Patel, Y. Barkley, S. Baumann-Pickering, and M. S. Soldevilla, "Automated extraction of odontocete whistle contours," *The Journal of the Acoustical Society of America*, vol. 130, no. 4, pp. 2212–2223, 2011.

[19] G. Bhatnagar, J. Wu, and B. Raman, "Fractional dual tree complex wavelet transform and its application to biometric security during communication and transmission," *Future Generation Computer Systems*, vol. 28, no. 1, pp. 254–267, 2012.