

# Feature Extraction Based on Bandpass Filtering for Frog Call Classification

Jie Xie<sup>(✉)</sup>, Michael Towsey, Liang Zhang, Jinglan Zhang, and Paul Roe

Electrical Engineering and Computer Science School,  
Queensland University of Technology, Brisbane, Australia  
{j3.xie,168.zhang}@qut.edu.au,  
{m.towsey,jinglan.zhang,p.roe}@hdr.qut.edu.au  
<https://www.ecosounds.org/>

**Abstract.** In this paper, we propose an adaptive frequency scale filter bank to perform frog call classification. After preprocessing, the acoustic signal is segmented into individual syllables from which spectral peak track is extracted. Then, syllable features including track duration, dominant frequency, and oscillation rate are calculated. Next, a k-means clustering technique is applied to the dominant frequency of syllables for all frog species, whose centroids are used to construct a frequency scale. Furthermore, one novel feature named bandpass filter bank cepstral coefficients is extracted by applying a bandpass filter bank to the spectral of each syllable, where the filter bank is designed based on the generated frequency scale. Finally, a k-nearest neighbour classifier is adopted to classify frog calls based on extracted features. The experiment results show that our proposed feature can achieve an average classification accuracy of 94.3 % which outperforms Mel-frequency cepstral coefficients features (81.4 %) and syllable features (88.1 %).

**Keywords:** Frog call classification · Spectral peak track · k-means clustering · Filter bank · k-nearest neighbour

## 1 Introduction

Recently, frog biodiversity has been threatened due to human activity and climate change [1]. Therefore, frog monitoring is becoming ever more important. Compared with traditional monitoring methods such as field observation, acoustic sensors can extend the monitoring into larger spatiotemporal scales [2]. Correspondingly, the use of acoustic sensor generates large volumes of acoustic data, which makes it essential to develop automatic acoustic data processing techniques.

Several papers have already described automated methods for detection and classification of animal calls. Since an elementary unit for frog call classification is one syllable [3], the first step of one frog call classification system is often syllable segmentation. In prior work, different features have already been explored for syllable segmentation, including energy [4, 5], zero-crossing rate (ZCR) [4, 5],

amplitude [3], spectrogram [6,7]. Compared with energy, ZCR, and amplitude, syllable segmentation based on spectrogram is more robust to the background noise [8]. With segmented syllables, feature extraction is the next crucial step for the performance of classification system. Lee et al. used Mel-frequency cepstrum coefficients (MFCCs) for classifying frog and cricket calls with linear discriminant analysis [9]. Chen et al. developed a method for frog call classification based on syllable duration and a multi-stage average spectrum [4]. Bedoya et al. used MFCCs as the feature for the recognition of anuran species with a fuzzy clustering technique [10]. Jie et al. explored image features for frog call classification with a k-nearest neighbour classifier [11]. All the previous work achieves a high accuracy rate in recognition and classification of frog calls. However, most features used are transplanted from speech processing directly, which might be not suitable for studying frog calls.

In this paper, one novel feature based on an adaptive frequency scale band-pass filter bank is proposed for frog call classification. Following our prior work [7], spectrogram is first investigated for segmentation. Then, spectral peak track is extracted from each segmented syllable for feature calculation: track duration, dominant frequency and oscillation rate. Next, a frequency scale is constructed by applying a k-means clustering technique to the dominant frequency of segmented syllables. Furthermore, a bandpass filter bank is designed based on the frequency scale, and applied to the spectral of each frog call syllable for extracting bandpass filter bank cepstral coefficients (BFCCs). Finally, a k-nearest neighbour (k-NN) classifier is used for frog call classification with extracted features. The experimental results show that our proposed feature can achieve the highest classification accuracy for classifying frog calls, which outperforms MFCCs and syllable features (SFs).

## 2 Materials and Methods

Our frog call classification system consists of four steps including pre-processing, syllable segmentation, feature extraction and classification. Detailed description of each step is shown in the following parts.

### 2.1 Data Description and Pre-processing

In this study, eighteen frog species which are widely spread in Queensland, Australian are selected for the experiment. All the recordings are obtained from David Stewart's CD [12]. Each recording includes one frog species, with duration ranged from 8 to 55 s. For pre-processing, human voice are first excluded from the recordings. Then, all the recordings are re-sampled at 16 kHz and mixed to mono.

### 2.2 Syllable Segmentation

After pre-processing, each recording consists of continuous frog calls, which is made up of multiple syllables. Here, one syllable is an elementary unit of frog

vocalizations for species detection [3]. For syllable segmentation, the iterative amplitude-frequency information is explored based on Härmä’s method [6]. The amplitude-frequency information is generated by applying STFT to the frog calls, where the window function is Kaiser window with the size and overlap being 512 samples and 25 %. A Gaussian filter ( $7 \times 7$ ) is optionally used before applying Härmä’s method for segmentation. The filter size used is set taking into account a trade-off between connecting gaps within one syllable and separating adjacent syllables.

### 2.3 Spectral Peak Track Extraction

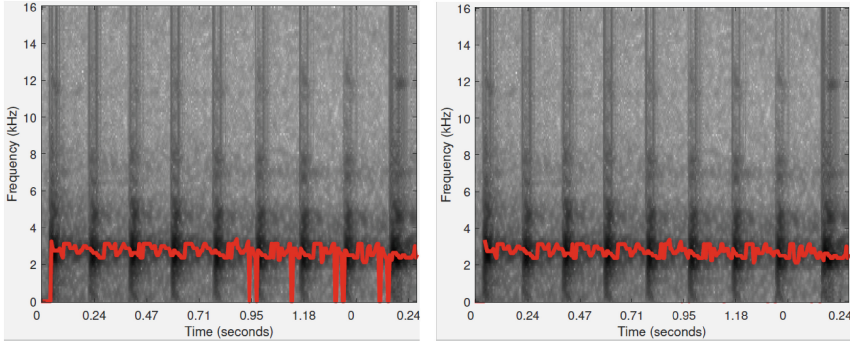
For frogs, related species often share more similar advertisement calls than distant species [13]. Applying STFT to those advertisement calls, each frog species is found to occupy one particular frequency band. Therefore, we explore the spectral peak track (SPT) to represent the dominant frequency trace of frog calls. The reasons for using SPT is (1) Isolate the desired signal from background noise; (2) Extract corresponding features based on SPT. The SPT extraction method used is briefly summarized here, with further details provided in [7]. In this SPT extraction algorithm, seven parameters need to be pre-defined (Table 1). The process for selecting those parameters is explained in Sect. 3.

**Table 1.** 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
$\beta$ (0 1)	Minimum density value

Before applying SPT extraction algorithm, each syllable is transformed to the spectrogram by dividing it into frames of 128 samples with 85 % overlap. For the generated spectrogram, selecting the maximum intensity (real peak) from each frame with a minimum required value  $I$  is the first step. Then, the time and frequency domain intervals between two successive peaks are calculated for satisfying  $T_c$  and  $f_c$ . If so, one initial track will be generated, then linear regression is applied to the generated track for calculating the position of next predicted peak. Next, the time and frequency domain intervals between predicted peak and the real peak are recalculated for satisfying  $T_c$  and  $f_c$ . If so, the real peak will be added to the initial track. This iterative process continues until  $T_s$  is no longer satisfied. After one track stops growing, comparing the duration and

density of the track with  $d_{min}$ ,  $d_{max}$ , and  $\beta$  is the next step. 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. 1. During the process of track extraction, time domain gaps of the track 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. 1.



**Fig. 1.** Spectral peak track extraction results for *Neobatrachus sudelli* (left: selected peaks whose frequency are zero mean that the peaks do not satisfy the intensity threshold  $I$  and are set to zero, right: spectral peak track with predicted peaks using linear regression).

Based on each spectral peak track, syllable features are calculated including track duration, dominant frequency and oscillation rate [7]. Here, track duration is the length of track; dominant frequency is calculate by averaging the frequency of the peak within the track; oscillation rate represents the pulse rate within one track.

## 2.4 Bandpass Filtering for Feature Extraction

After calculating the dominant frequency for all frog species, some frog species are found to have similar dominant frequency but different spectral distribution. In this study, we explore the bandpass filtering technique for capturing the spectral information. First, frequency scale is generated by applying k-means clustering algorithm to the dominant frequency of all frog syllables. Here,  $k$  is selected as 18, which is the same with the number of frog species to be classified; the distance function is city block function. After applying the k-means clustering, 18 centroids ( $C_i, (i = 1, \dots, 18)$ ) are saved for design the bandpass filter.

**Bandpass Filter Design for Feature Extraction.** In this study, a cascade of a 20th-order equalizer and a band-pass filter (Butter-worth filter) is used to design a filter bank for feature extraction.

**Algorithm 1.** Design of a bandpass filter bank

**Data:**  $C_i (i = 1 : N)$ , where  $C_i$  is the centroid of the clustering results;  $N$  is the number of frog species to be classified.

**Result:** Bandpass filter bank

**begin**

**Step 1:** Sort the centroid  $C$  and calculate the difference between the consecutive vectors of  $C$ , save the results as  $d_j (j = 1 : N - 1)$

**Step 2:** Design a filter bank with  $N$  bandpass filters based on the centroids. Here,  $k = 1 : N$  is the index of bandpass filter; LF and HF represent the low and high cutoff frequency.

**if**  $k = 1$  **then**

$$\text{HF}(k) = C_k + \frac{d_k}{2};$$

$$\text{LF}(k) = C_k - \frac{d_k}{2};$$

**else if**  $k = N$  **then**

$$\text{HF}(k) = C_k + \frac{d_{k-1}}{2};$$

$$\text{LF}(k) = C_k - \frac{d_{k-1}}{2};$$

**else**

$$\varepsilon = \min\left(\frac{d_{k-1}}{2}, \frac{d_k}{2}\right)$$

$$\text{HF}(k) = C_k + \varepsilon;$$

$$\text{LF}(k) = C_k - \varepsilon;$$

With the generated bandpass filter bank, we apply it to the spectrum of each frog syllable  $x(n)$ . Detailed steps for calculating bandpass filter bank cepstral coefficients (BFCCs) are described as follows:

**Step 1:** Apply bandpass filter bank to  $X(k)$

Filter  $X(k)$  with the generated filter bank, and save the filtered results of each bandpass filter as  $B(i, j), j = 1, \dots, J$ . Where  $X(k)$  is the result after applying fast Fourier transform to the windowed signal  $x(n)$ ,  $i$  is the number of coefficients for each bandpass filter,  $j$  is the index of the filter.

**Step 2:** Calculate the energy of filtered result for each frequency band

$$E_{i,j} = \sum_{i=1}^{M_i} [B(i, j)]^2 \quad (1)$$

where  $M_i$  means the number of coefficients after bandpass filtering.

**Step 3:** Perform discrete cosine transform on the logarithm energy and obtain the feature BFCCs for each windowed signal

$$BFCCs(d) = \sum_{i=1}^I \log E_{i,j} \cos\left(\frac{d(i-0.5)}{I} \pi\right) \quad (2)$$

where  $d = 1, 2, \dots, D$ , and  $D$  is the dimension of BFCCs and set as 12.  $i$  is the index of energy for each bandpass filter.

**Step 4:** Average BFCCs over the temporal direction

$$BFCC_s = \frac{\sum_{f=1}^F BFCC_s(d, f)}{F} \quad (3)$$

where  $f$  is the index of windowed signal,  $F$  is the number of windowed signal after windowing.

## 3 Experiments and Discussion

### 3.1 Parameter Tuning

In this study, parameters of two parts need to be discussed: spectral peak track extraction and feature calculation. For spectral peak track, seven parameters were determined empirically by applying various combinations of thresholds to a small randomly selected syllables. Here, minimum and maximum duration are 60 ms and 1000 ms. The density value is 0.8, which describes the integrity of one syllable. The minimum intensity value is 3 dB. The maximum time interval for connecting peaks is 1.5 ms, and the maximum frequency interval is 500 Hz. For feature extraction, MFCCs are used as the baseline [9], where the window size and overlap are 512 samples and 50%.

### 3.2 Classification

The k-NN classifier has been successfully employed for classifying bioacoustic signal [3, 7, 14]. In this experiment, the k-NN classifier is used to learn a model on the training examples with 10-fold cross-validation for frog classification. Since the k-NN classifier is sensitive to the local structure of the data as well the initial cluster centroids, we run the k-NN classifier for 10 times based on different initial points. The feature performance is evaluated by the classification accuracy, which is defined as

$$Accuracy(\%) = \frac{N_c}{N_t} * 100\% \quad (4)$$

where  $N_c$  is the number of syllables that are correctly classified, and  $N_t$  is the total number of syllables for one frog species. Three features are put into the classifier: syllable features (SFs), MFCCs, and BFCCs. The averaged classification accuracy is shown in Table 2.

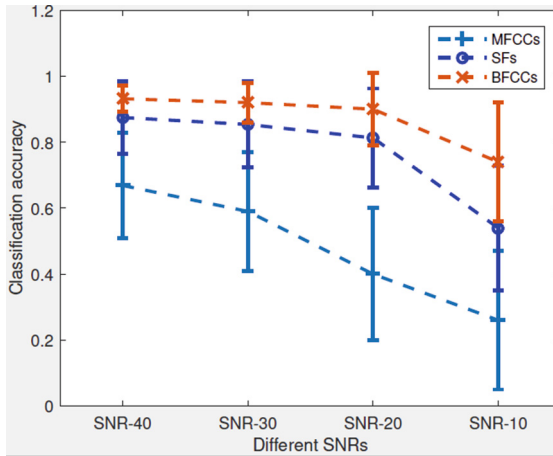
In this experiment, the averaged classification accuracy for MFCCs and SFs is 81.4% and 88.1% respectively. Our proposed feature achieves the highest classification accuracy (94.3%). For MFCCs, the classification accuracy of *Neobatrachus sudelli* and *Phyloria kundagungan* is 100%, because their spectrum is different from other frog species. Compared with MFCCs and SFs, the classification accuracy of all frog species using BFCCs are higher than 90% except *Mixophyes fleayi*. Since the spectrum of *Mixophyes fleayi* and *Limnodynastes terraereginae* are similar, the classification accuracy of them are relatively low.

However, the duration and oscillation rate between *Mixophyes fleayi* and *Limnodynastes terraereginae* are different, which leads to a higher classification accuracy for SFs. For MFCCs, the spectrum is extracted based on the Mel-scale filter bank, which is designed based on the human auditory rather than the character of frog calls. With derived dominant frequency which has shown its ability for discriminating frog species [7, 13], the designed bandpass filter bank is more suitable for the frequency scale of frog species to be classified. Compared with SFs, the use of bandpass filter bank captures not only the information of dominant frequency but also the distribution of the frog calls through all frequency bands.

**Table 2.** Classification accuracy (mean and standard deviation) for 18 frog species with three different features.

Scientific name	No. of syllables	Classification accuracy		
		MFCCs	SFs	BFCCs
<i>Assa darlingtoni</i>	36	23.3 ± 26.0	85.0 ± 16.6	100.0 ± 0.0
<i>Crinia parinsignifera</i>	32	93.3 ± 13.3	76.7 ± 26.0	100.0 ± 0.0
<i>Crinia signifera</i>	22	60.0 ± 32.7	50.0 ± 22.4	100.0 ± 0.0
<i>Limnodynastes tasmaniensis</i>	16	50.0 ± 40.0	35.0 ± 22.9	90.0 ± 30.0
<i>Limnodynastes terraereginae</i>	54	86.7 ± 12.5	96.0 ± 8.0	95.0 ± 7.6
<i>Litoria chloris</i>	35	76.7 ± 15.3	77.5 ± 17.5	90.0 ± 15.3
<i>Litoria latopalmata</i>	172	98.2 ± 2.7	97.6 ± 3.9	100.0 ± 0.0
<i>Litoria nasuta</i>	74	85.0 ± 9.4	87.1 ± 13.5	98.8 ± 3.8
<i>Litoria revelata</i>	130	93.8 ± 5.8	99.2 ± 2.3	100.0 ± 0.0
<i>Litoria rubella</i>	37	93.3 ± 13.3	90.0 ± 16.6	96.7 ± 10.0
<i>Litoria tyleri</i>	71	90.0 ± 10.9	98.6 ± 4.3	98.8 ± 3.8
<i>Litoria verreauxii verreauxii</i>	28	80.0 ± 33.2	100.0 ± 0.0	100.0 ± 0.0
<i>Mixophyes fasciolatus</i>	33	95.0 ± 10.0	90.0 ± 15.3	92.5 ± 16.0
<i>Mixophyes fleayi</i>	16	70.0 ± 45.8	100.0 ± 0.0	40.0 ± 29.0
<i>Neobatrachus sudelli</i>	22	100.0 ± 0.0	90.0 ± 20.0	100.0 ± 0.0
<i>Philoria kundagungan</i>	22	100.0 ± 0.0	100.0 ± 0.0	100.0 ± 0.0
<i>Uperoleia fusca</i>	39	95.0 ± 10.0	87.5 ± 12.5	97.5 ± 7.5
<i>Uperoleia laevigata</i>	25	75.0 ± 33.5	100.0 ± 0.0	100.0 ± 0.0
Averaged accuracy	864	81.4 ± 18.0	88.1 ± 13.7	<b>94.3 ± 7.9</b>

For testing the robustness of the syllable features, a Gaussian white noise signal, with signal to noise ratio (SNR) of 40 dB, 30 dB, 20 dB, and 10 dB was added to the original audio data. The results from running the classifier on audio data with artificially added background noise are shown in Fig. 2, which show the ability of our feature extraction method for dealing with background noise.



**Fig. 2.** Classification accuracy with MFCCs, SFs, and BFCCs under different levels of noise contamination.

## 4 Conclusion

This study presents a novel feature extraction method for frog call classification. After segmenting the audio data into syllables, the SPT algorithm is applied to each syllable. Then, syllable features are calculated including track duration, dominant frequency and oscillation rate. Based on the dominant frequency, a frequency scale is constructed with a k-means clustering algorithm for generating the bandpass filter bank. Finally, a feature set is extracted with generated filter bank for classifying frog calls using a k-NN classifier. The experimental results are promising with an average classification accuracy of 94.3% for BFCCs. Future work will include additional experiments that test a wider variety of frog calls from different geographical and environment conditions.

## References

1. Stuart, S.N., Chanson, J.S., Cox, N.A., Young, B.E., Rodrigues, A.S., Fischman, D.L., Waller, R.W.: Status and trends of amphibian declines and extinctions worldwide. *Science* **306**(5702), 1783–1786 (2004)
2. Wimmer, J., Towsey, M., Planitz, B., Williamson, I., Roe, P.: Analysing environmental acoustic data through collaboration and automation. *Future Gener. Comput. Syst.* **29**(2), 560–568 (2013)
3. Huang, C.-J., Yang, Y.-J., Yang, D.-X., Chen, Y.-J.: Frog classification using machine learning techniques. *Expert Syst. Appl.* **36**(2), 3737–3743 (2009)
4. Chen, W.-P., Chen, S.-S., Lin, C.-C., Chen, Y.-Z., Lin, W.-C.: Automatic recognition of frog calls using a multi-stage average spectrum. *Comput. Math. Appl.* **64**(5), 1270–1281 (2012)
5. Jaafar, H., Ramli, D.A.: Automatic syllables segmentation for frog identification system. In: 2013 IEEE 9th International Colloquium on Signal Processing and Its Applications (CSPA), pp. 224–228. IEEE (2013)



6. Harma, A.: Automatic identification of bird species based on sinusoidal modeling of syllables. In: Proceedings of the 2003 IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP 2003), vol. 5, pp. V–545. IEEE (2003)
7. Xie, J., Towsey, M., Truskinger, A., Eichinski, P., Zhang, J., Roe, P.: 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, April 2015
8. Colonna, J.G., Cristo, M., Salvatierra, M., Nakamura, E.F.: An incremental technique for real-time bioacoustic signal segmentation. *Expert Syst. Appl.* **4221**, 7367–7374 (2015)
9. Lee, C.-H., Chou, C.-H., Han, C.-C., Huang, R.-Z.: Automatic recognition of animal vocalizations using averaged MFCC and linear discriminant analysis. *Pattern Recogn. Lett.* **27**(2), 93–101 (2006)
10. Bedoya, C., Isaza, C., Daza, J.M., López, J.D.: Automatic recognition of anuran species based on syllable identification. *Ecol. Inform.* **24**, 200–209 (2014)
11. Xie, J., Towsey, M., Zhang, J., Dong, X., Roe, P.: Application of image processing techniques for frog call classification. In: 2015 IEEE International Conference on Image Processing (ICIP), pp. 4190–4194, September 2015
12. Stewart, D.: “Australian frog calls: subtropical east,” Audio CD (1999). [http://www.naturesound.com.au/cd\\_frogsSE.htm](http://www.naturesound.com.au/cd_frogsSE.htm)
13. Gingras, B., Fitch, W.T.: A three-parameter model for classifying anurans into four genera based on advertisement calls. *J. Acoust. Soc. Am.* **133**(1), 547–559 (2013)
14. Han, N.C., Muniandy, S.V., Dayou, J.: Acoustic classification of Australian anurans based on hybrid spectral-entropy approach. *Appl. Acoust.* **72**(9), 639–645 (2011)