Applied Acoustics xxx (2011) xxx–xxx

Contents lists available at ScienceDirect

Applied Acoustics

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

Acoustic classification of Australian anurans based on hybrid spectral-entropy approach

Ng Chee Han [a,*], Sithi V. Muniandy [b], Jedol Dayou [a]

[a] Energy, Vibration and Sound Research Group (e-VIBS), Locked Bag 2073, School of Science and Technology, University of Malaysia Sabah, 88999 Kota Kinabalu, Malaysia
[b] Department of Physics, University of Malaya, 50603 Kuala Lumpur, Malaysia

ARTICLE INFO

Article history:
Received 28 August 2009
Received in revised form 27 January 2010
Accepted 7 February 2011
Available online xxxx

Keywords:
Feature extraction
k-NN classifier
Spectral centroid
Shannon entropy
Rényi entropy

ABSTRACT

A new hybrid method for automated frog sound identification, using spectral centroid, Shannon entropy and Rényi entropy is proposed. The advantage of using entropy based information theoretic approach for analyzing complexity of bioacoustics signals in animal vocalization is discussed. Sound samples from nine species of Microhylidae frogs are first segmented into syllables. Fourier spectral centroid, Shannon entropy and Rényi entropy of the syllables are then determined. Finally, nonparametric *k*-th nearest neighbour (*k*-NN) classifier is used to recognize the frog species based on these three extracted features. Result shows that the *k*-NN classifier based on these selected features is capable to identify the species of the frogs with an average accuracy of 98%. It is found that the accuracy reduces significantly only when the noise levels higher than −20 dB.

1. Introduction

Recognition of animals from their calls has been studied for a very long time [1]. Animal sound productions normally can be divided into two categories namely the non-incidental sounds, which are used for communication purpose, and the incidental sounds that result as the by-product of their activities. Quite naturally, the animal species could be identified according to their sound productions. Nevertheless, manual classification of bioacoustics signals can be very ambiguous and most often rely heavily on the surveyor's expert knowledge of the group under investigation. Automated identification or recognition of animal species based on their sounds offers many advantages especially for rapid biodiversity assessment and eco-system monitoring using sensor networks, which also may bring the joy of discovering of new species of animals by amateurs equipped with a smart species recognition system. When many sensors are available, localization of particular species could also possible for other purposes, for example for exact species mapping or location identification. Biodiversity assessment has become a central and urgent task in conservation biology, not only to determine species richness but also to evaluate differences between communities occupying different areas or to determine the temporal change in the population [2].

Species identification using computational methods involves pattern recognition in which an unknown specimen is placed into one of a number of possible classes depending on the features extracted from the measurements on the species [1]. Many techniques have been introduced for bioacoustics signal detection and analysis. Most of these approaches rely on time domain and/or frequency domain analyses. Time domain approach for signal processing may include features such as frame energy, silence ratio, volume root mean square (RMS), volume dynamic ratio (VDR), total energy and zero-crossing ratio [3]. Fourier transform based power spectrum, wavelet transform and linear prediction coding (LPC) coefficients are examples methods used to extract relevant frequency (or time–frequency) contents for frequency (or time–frequency) domain techniques [4]. Tyagi et al. [5] studied bird sound based on the average spectrum over time and classified species using template matching method. Vilches et al. [6] has introduced the pulse-by-pulse basis as feature extraction and used data mining techniques for classification. In the related development, Chesmore [1,3] has introduced application of time domain signal coding (TDSC) approach to extracting birds and orthoptera sound signals and successfully classified them by using artificial neural network (ANN). Degradation in the performance of this technique under high noise to signal ratio has also been pointed in the study.

Huang [7] has pointed out that most of the research works on recognition of animal calls have focused on animal species identification, such as for bird species identification [5,6]. Other than this, Chesmore [3] has carried out several studies on similar

* Corresponding author.
E-mail address: ng_chee_han@hotmail.com (N.C. Han).

0003-682X/$ - see front matter © 2011 Elsevier Ltd. All rights reserved.
doi:10.1016/j.apacoust.2011.02.002

classification such as for insects, bats, deer and killer whales. Studies on identification of other animal calls are relatively few [8].

Technique devoted to the development of automated frog sound recognition system is not widely known in the literature. Thus, it becomes essential to develop an efficient frog sound recognition system [7]. Frog sound can be seen as an organized sequence of brief sounds from a species-specific vocabulary. Those brief sounds are usually called syllables [9]. 'Fingerprint' of the bioacoustics signal can be constructed by extracting useful parameters of these syllables to form a feature vector. Selection of suitable features to be used in the recognition process plays a critical role for any automated recognition system. In this paper, we consider spectral-entropy method consists of Fourier spectral centroid, Shannon entropy and Rényi entropy as the bioacoustics features determined from a collection of sound syllables of nine frog species. It is shown that using this hybrid method, accuracy in frog species identification is superior compared to when they are used individually.

Nine frog species from Australian Microhylidae family were selected for this investigation. The Microhylidae is a family of firmisternal frogs, which have broad sacral diapophyses, one or more transverse folds on the surface of the roof of the mouth, and a unique slip to the abdominal musculature. Almost all Australian Microhylids are small (snout to vent length less than 35 mm), and all have procoelous vertebrae, are toothless and smooth-bodied, with transverse grooves on the tips of their variously expanded digits. The terminal phalanges of fingers and toes of all Australian microhylids are T-shaped or Y-shaped with transverse grooves [10].

Normally, Microhylidae frogs make their call when they are ready to mate which is during the rainy season. Some calls are made with single or groups of ringing notes, while some have harsher voices. Many males have a bag of skin, called a vocal sac, on the throats. The vocal sac fills with air and deflates when they call (http://animals.jrank.org/).

2. General method of frog sound identification

The frog sound identification system proposed in this work basically consists of three processes, namely the syllable segmentation, feature extraction and classification.

2.1. Syllable segmentation

A syllable is basically a sound that a frog produces with a single blow of air from the lungs [7]. The rate of events in frog vocalization may be so high that the separation of individual syllables is difficult in a natural environment due to reverberation. Once the syllables have been properly segmented, a set of features can be calculated to represent each syllable [7]. Fig. 1 shows four examples of the syllables waveforms for frog species of *Cophixalus bombiens*. From this example, basically, we can see the waveforms from each syllables are actually looked similar.

Depending on the species, the number of syllable in a call varies from as low as 12 syllables and as high as 96 syllables. The duration for each syllable is between 3.58×10^{-3} s to 0.065 s, whereas the duration for each call is in the range of 0.536 s to over 3 s. The dominant frequency of the call for each species also varies from 1.7 kHz to 4.9 kHz with different active frequency range. Interested readers could refer to the table in Appendix A for some time domain and frequency-domain characteristics of the call for each species.

In this study, a signal segmentation software tool named '*Raven Lite*' has been used for all of the frog species sound samples. The segmented syllables for each species are digitized in 24-bit WAV format with sampling frequency of 44.1 kHz. Matlab software is used for further analysis, which is the feature extraction and classification.

Fig. 2 shows an example of spectrograms of a complete call for four different species which were also obtained by using Raven Lite

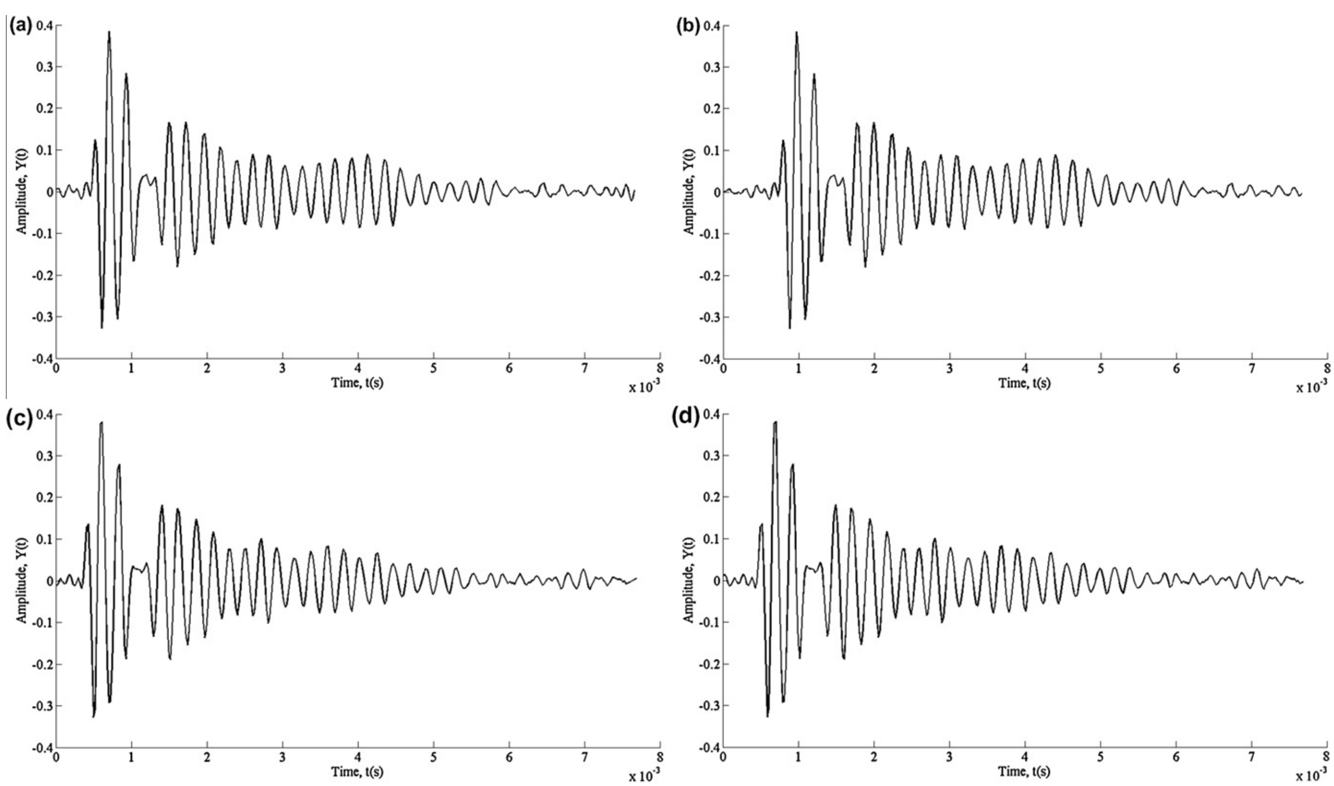

Fig. 1. Four examples of segmented syllables of *Cophixalus bombiens*. (a) Syllable 1, (b) Syllable 2, (c) Syllable 3 and (d) Syllable 4.

Please cite this article in press as: Han NC et al. Acoustic classification of Australian anurans based on hybrid spectral-entropy approach. Appl Acoust (2011), doi:10.1016/j.apacoust.2011.02.002

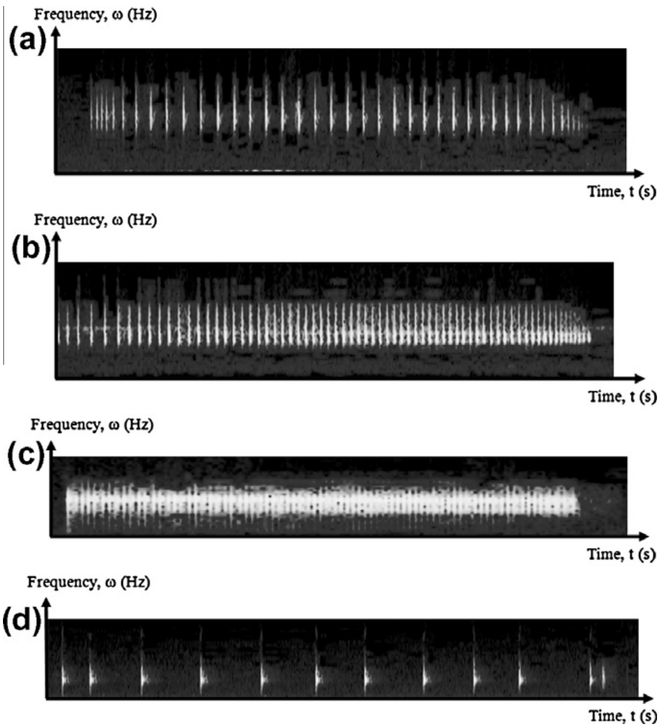

Fig. 2. Spectrogram examples of complete call for four species in this study. (a) *Cophixalus hosmeri*, (b) *Cophixalus infacetus*, (c) *Cophixalus neglectus* and (d) *Cophixalus saxatilis*.

software. Figs. 3 and 4 show the examples of segmented syllable from these species in waveforms and their corresponding spectrogram (using mesh method in MATLAB).

2.2. Feature extraction

Three features are extracted from the sound syllables, namely the spectral centroid, Shannon entropy and Rényi entropy. Spectral centroid is a commonly used feature in the pattern recognition studies [7]. The main contribution of the present study would be the inclusion of information theoretic concepts such as the Shannon entropy and the Rényi entropy to form a spectral-entropy hybrid system that is shown to improve species identification.

2.2.1. Spectral centroid

Spectral centroid is the center point of spectral distribution. In terms of human audio perception, it is often associated with the brightness of the sound. Brighter sound is related to the higher centroid [7]. The spectral centroid f_c of a power spectrum distribution is defined as the weighted mean of the frequencies present in the signal, given by

$$f_c = \frac{\sum_{k=0}^{N-1} f_k P_k}{\sum_{k=0}^{N-1} P_k}, \tag{1}$$

where P_k is the magnitude of the spectrum of bin number k and f_k represents the center frequency of the respective bin.

2.2.2. Shannon entropy

Shannon entropy H is the expected information content of a sequence or signal $X = \{x_1, x_2, x_3, \ldots, x_n\}$. Shannon entropy describes the average of all the information contents I weighted by their probabilities p_i, namely

$$H(X) = E[I(p)] = \sum_{i=1}^{n} p_i I(p) = -\sum_{i=1}^{n} p_i \log_2(p_i), \tag{2}$$

where $E[\,]$ denotes expectation value. The continuous version of the Shannon entropy is called the differential entropy written as

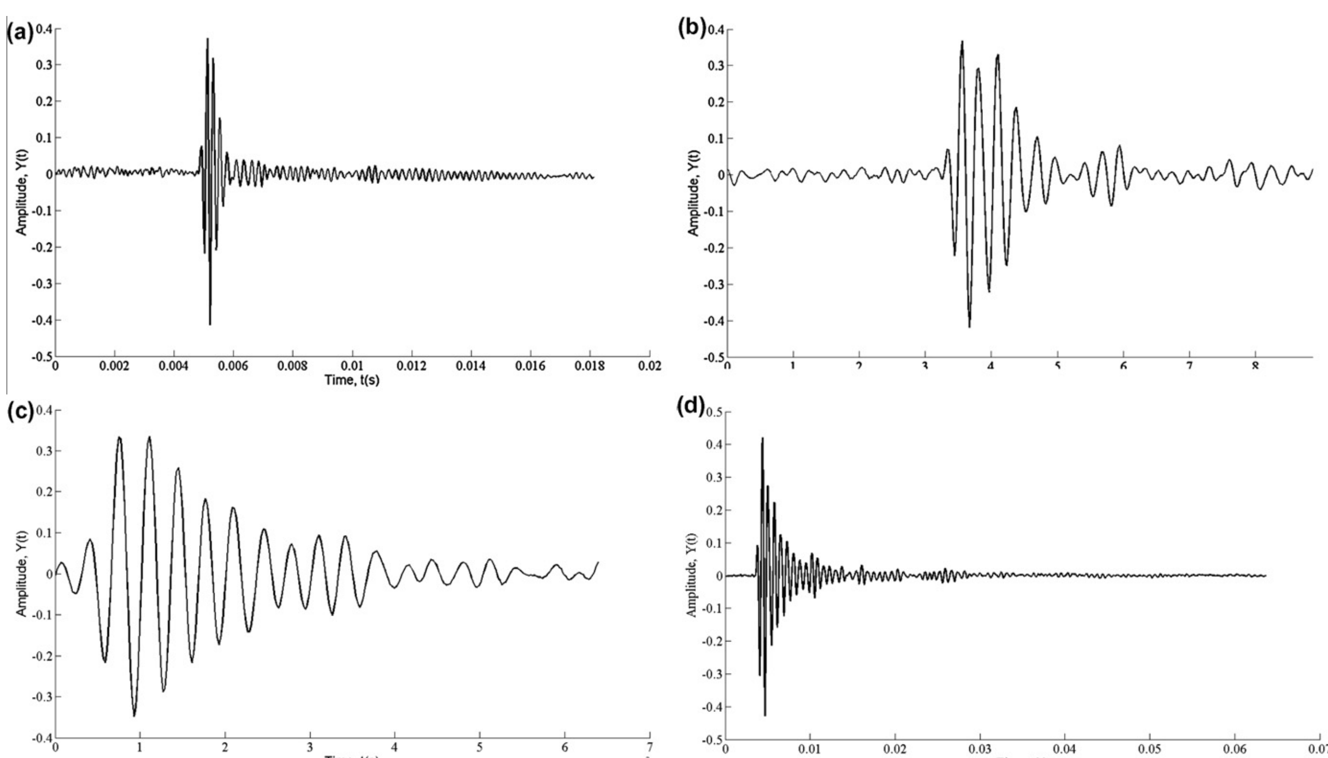

Fig. 3. Four examples of segmented syllable waveforms for four species in this study. (a) *Cophixalus hosmeri*, (b) *Cophixalus infacetus*, (c) *Cophixalus neglectus* and (d) *Cophixalus saxatilis*.

Please cite this article in press as: Han NC et al. Acoustic classification of Australian anurans based on hybrid spectral-entropy approach. Appl Acoust (2011), doi:10.1016/j.apacoust.2011.02.002

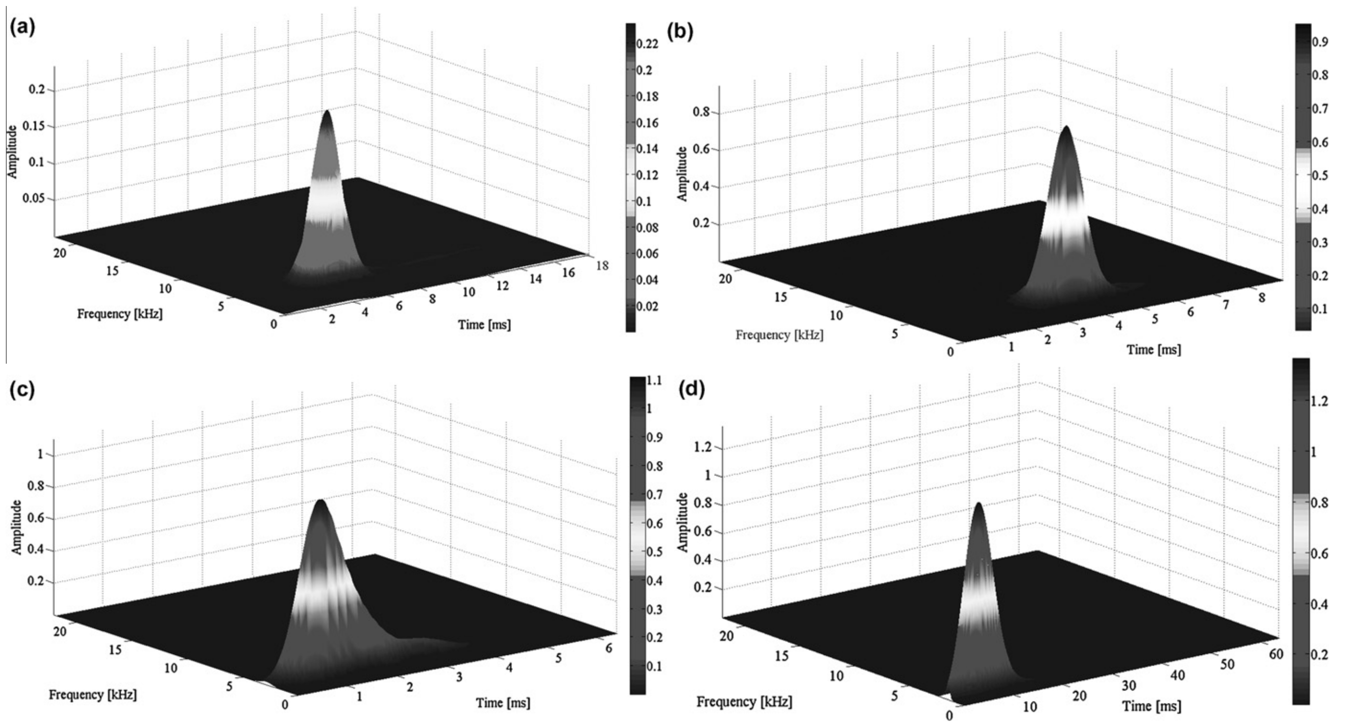

Fig. 4. Corresponding spectrogram for each segmented waveform given in Fig. 3. (a) *Cophixalus hosmeri*, (b) *Cophixalus infacetus*, (c) *Cophixalus neglectus* and (d) *Cophixalus saxatilis*.

$$H(X) = -\int_{-\infty}^{+\infty} p(x)\log_2[p(x)]dx. \tag{3}$$

In other words, Shannon entropy of a signal indicates the degree of predictability of the signal. Consider a d.c. signal at constant amplitude k. Its probability density function is then a unitary impulse located at k, i.e. $p_i = \delta(k)$, therefore its entropy or unpredictability is zero. In ecology diversity, Shannon entropy is the second most used index after species richness (number of species) [11]. It increases with the evenness of the categories numbers and frequencies. In ecology, categories are often species that differ by their relative richness in a habitat [12]. In biodiversity study, Sueur et al. [12] proposed the use of Shannon entropy on a time series sequence, where the categories are represented by time units and their frequencies are referred to the probability mass function of the amplitude envelope. Therefore, by using this measure, the time units with low probability mass function of the amplitude envelope characterize the acoustic diversity. Nevertheless, production of animal sounds in field will also influence the amplitude envelope at each time unit. In this study, we use the Shannon entropy as a measurement of richness of the information contents in frog sounds.

2.2.3. Rényi entropy

Rényi entropy of order $\alpha \geqslant 0$ is defined as [13]

$$H_\alpha(X) = \frac{1}{1-\alpha}\log_2\left(\sum_i^n p_i^\alpha\right) \tag{4}$$

where p_i is the probabilities of the occurrence $\{x_1, x_2, x_3, \ldots, x_n\}$ in the signal. Rényi entropy have been used in communication and coding theory [14], data mining, detection, segmentation, classification [15], characterization of signals and sequences [16], signal processing [17], image matching and registration [15]. Rényi information can be used to "obtain different averaging of probabilities" via the parameter α (see [18]). By considering $H\alpha$ as a function of α, the spectrum of the Rényi entropy is also of some interest in signal analysis. For example, Rényi information of order of $\alpha = 2$ is used as a measure of diversity in economics [19]. In the study of random signal, a lower bound of Rényi entropy at least in order of $\alpha = 2$ is often adopted. In the limit $\alpha \to 1$, the Rényi entropy approaches the Shannon entropy. In information theoretic works, measurement of Rényi entropy also refers to the estimation of noise when transferring a signal. In this paper, Rényi entropy is used to represent the noise content of the sound sample which also implies their complexity. One may expect that in this context, the 'highly ordered' frog call will produce sound of relatively low complexity. For this study, we choose $\alpha = 3$ for the estimation of Rényi entropy. This is the default value in Matlab software and is widely used in other studies

2.3. Classification of k-nearest neighbours

One of the most elegant and simplest classification techniques is the k-nearest neighbour (k-NN) rule [20], which is used to classify d-dimensional feature vector $x \in R^d$. At first, the classifier searches for the k-NN among a set of m prototypes for an input vector x. The k-NN metric is then calculated utilizing the L^p-norm, $p \in [1, \infty]$, where the special cases are the Manhattan distance ($p = 1$), the Euclidean distance ($p = 2$) and the maximum distance ($p = \infty$). The distance metric is defined as

$$d_j^p(x, x^j) = \|x - x^j\|_p = \left(\sum_{i=1}^d |x_i - x_i^j|\right)^{1/p} \tag{5}$$

For the distances d_j^p,there is a sequence $(\tau_i)_{i=1}^m = \in (1, 2, \ldots, m)$, with the property of $d_{\tau_1}^p \leqslant \ldots \leqslant d_{\tau_m}^p$. The k-NN of x is then defined by $N_k(x) := \{x^{\tau_1}, \ldots x^{\tau_k}\}$ with $k \leqslant m$. Let $c_i = c(x_{\tau_i}), i = 1, \ldots, k$ be the classes of the k-NN and $N_k^j(x) = \{y \in N_k(x) | c(y) = j\}$ be the subset of nearest neighbour of class j. The classification for input x is defined through majority voting, i.e. j^*: = argmax $(j = 1, \ldots, l | N_k^j(x)|)$, where j^* is the class which most often occurs among the k-NN. Here l is

Table 1
Comparison of the accuracy of the classifier with and without entropy (Shannon entropy and Rényi entropy).

No.	Scientific name	Number of syllable	k-NN (with entropy)		k-NN (without entropy)	
			Accuracy (%)	Correct syllable	Correct syllable	Accuracy (%)
1	*Cophixalus bombiens*	6	6	100.00	3	50.00
2	*Cophixalus concinnus*	6	6	100.00	3	50.00
3	*Cophixalus exiguus*	6	6	100.00	3	50.00
4	*Cophixalus hosmeri*	6	5	83.33	2	33.33
5	*Cophixalus infacetus*	6	6	100.00	3	50.00
6	*Cophixalus monticola*	6	6	100.00	6	100.00
7	*Cophixalus neglectus*	6	6	100.00	3	50.00
8	*Cophixalus ornatus*	6	6	100.00	6	100.00
9	*Cophixalus saxatilis*	6	6	100.00	6	100.00

the number of classes [21]. k-NN has been applied in various sound analysis problems [7]. Given as a set of parameters, it finds the nearest neighbour among training data and uses the categories of the neighbour to determine the class of a given input. The spectral centroid, Shannon entropy and Rényi entropy are the input parameters for the k-NN classifier used in this study.

3. Experimental results

In this work, a database that consists of nine frog species found in Australia, as listed in Table 1 (obtained from http://www.Frogs-australia.net.au/frogs), are used. Sound files in the form of single syllable for nine frog species are segmented from the original recordings. For each species, a total of six segments were prepared.

The following classification accuracy is used to examine the performance of the proposed work:

$$A = \frac{N_c}{N_s} \times 100, \tag{6}$$

where N_c is the number of syllables which were recognized correctly and Ns is the total number of test syllables.

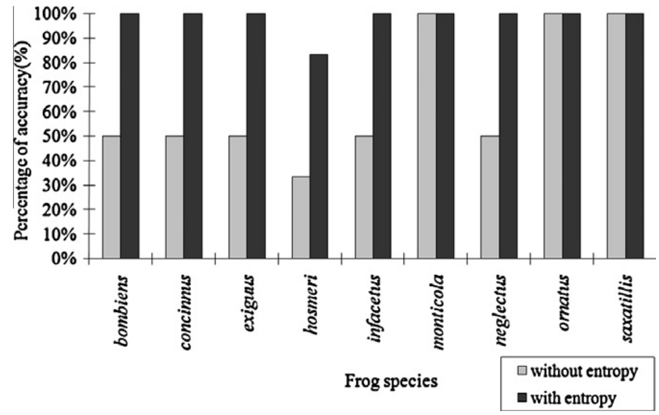

Fig. 5. Comparison of the performance of spectral and hybrid spectral-entropy based classification of nine Microhylidae frog species.

The experimental results are shown in Fig. 5 where a total of 54 syllables (six syllables for each species) are used for testing. The k-NN classifier is used in two models, which are with spectral centroid alone and with spectral centroid, Shannon entropy and Rényi entropy (hybrid method). Their accuracies are then compared to assess the effectiveness of entropy as an additional feature for frog sound classification. Euclidean distance is chosen for the k-NN classifier with the number of neighbour, $k = 11$.

The results of classification with spectral centroid, Shannon entropy and Rényi entropy for nine Microhylidae frog species are shown in Figs. 6–8, respectively. Finally, the sensitivity of the k-NN classifier with Shannon and Rényi entropies is tested for different noise levels. We used the Gaussian white noise at signal to noise ratio of −40 dB, −30 dB, −20 dB and −10 dB, respectively and the results are shown in Fig. 9.

4. Discussions and conclusions

In this study, we have incorporated two concepts of entropy, namely Shannon and Rényi entropy, as features for species classification in the k-NN classifier. Our results showed that the classifier based on spectral centroid without the entropy failed to classify the following species: *C. bombiens*, *Cophixalus concinnus*, *Cophixalus exiguus*, *Cophixalus hosmeri*, *Cophixalus infacetus* and *Cophixalus neglectus*. For these species, the accuracy of classification is less than 50%. This could be explained by referring to Fig. 6, where the spectral centroid of the *C. bombiens and C. hosmeri* falls into the same level of frequency. Similar explanation applies for *C. exiguus* and *C. infacetus*. Besides, *Cophixalus concinuus*, *Cophixalus*

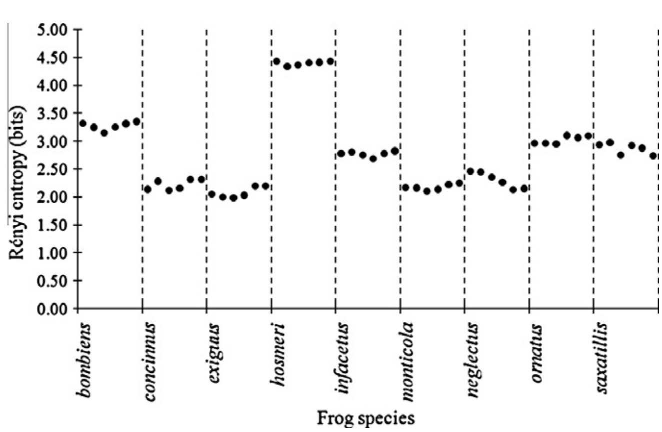

Fig. 6. Classification of nine Microhylidae frog species based on spectral centroid.

Please cite this article in press as: Han NC et al. Acoustic classification of Australian anurans based on hybrid spectral-entropy approach. Appl Acoust (2011), doi:10.1016/j.apacoust.2011.02.002

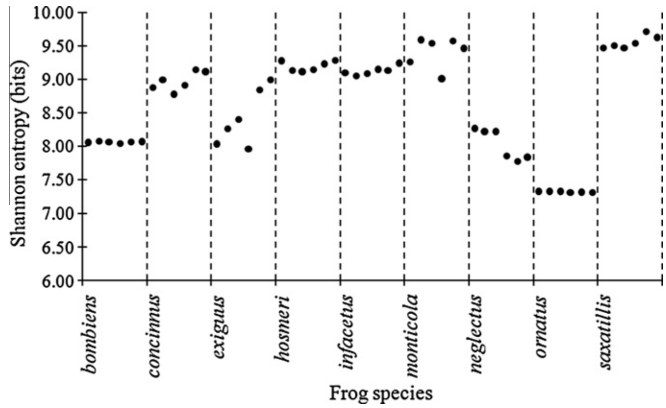

Fig. 7. Classification of nine Microhylidae frog species based on spectral centroid combined with Shannon entropy.

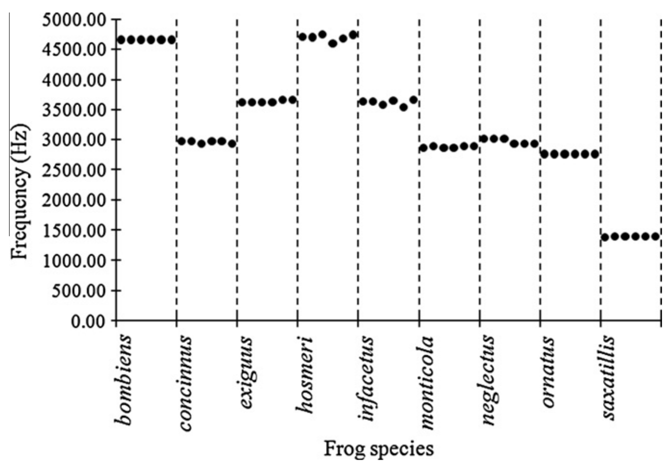

Fig. 8. Classification of nine Microhylidae frog species based on spectral centroid combined with Rényi entropy.

monticola, *C. neglectus* and *Cophixalus ornatus* exhibit spectral centroid frequencies very close to each other. This explained why the *k*-NN classifier is unable to identify or misclassify the frog species which share similar frequencies. However, there are some species such as the *C. monticola*, *C. ornatus* and *Cophixalus saxatilis* which classifier able to differentiate despite having spectral centroid frequencies close to each other (see Fig. 6). As for *C. saxatillis*, the spectral centroid frequency is very distinctive from other species, thus making it the easiest species to classify.

By introducing the entropy into the classifier, we have seen marked improvement in the accuracy of classification. The entropy based classifier managed to identify nine of the Microhylidae frogs with average accuracy more than 98%. As a comparison, the hybrid classifier with entropy approach has successfully identified several species which failed in the previous case, such as *C. bombiens*, *C. concinnus*, *C. exiguus*, *C. hosmeri*, *C. infacetus* and *C. neglectus* (see Table 1).

In order to determine the classifier's response to varying noise levels, Gaussian white noise was added to all sound samples. It is evident from Fig. 9 that the identification accuracy improves under low noise conditions, except for *C. hosmeri*. The overall performance of entropy based *k*-NN classifier is found to be better than the purely spectral based approach. The former did well in classifying *C. bombiens, C. exiguus, C. monticola, C. ornatus* and *Cophixalus*

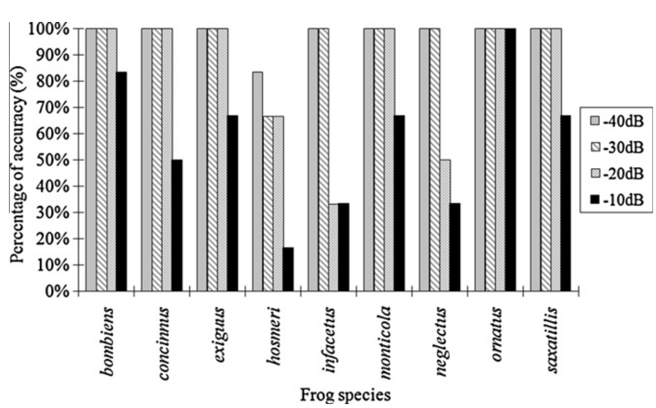

Fig. 9. Sensitivity of the spectral-entropy based classifier for different levels of noise contamination.

saxatillis species. This study has shown that performance of the classifier can be enhanced by using entropy approaches. By combining different definitions of entropy, we have shown the advantage of spectral-entropy based *k*-NN classifier that incorporates both of the spectral harmonics as well as signal predictability and complexity measures for species classification.

Acknowledgements

This study was supported by the Fundamental Research Grants No. FRGS0031-ST-1/2006. S.V. Muniandy thank the Malaysian Ministry of Science, Technology and Innovation (MOSTI) for financial support under e-Sciencefund 16-02-03-6008.

Appendix A. Supplementary material

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.apacoust.2011.02.002.

References

[1] Chesmore ED. Automated bioacoustic identification of species. An Acad Bras Ciênc 2004;76(2):435–40.
[2] Purvis A, Agapow PM, Gittleman JL, Mace GM. Nonrandom extinction and the loss of evolutionary history. Science 2000;288:328–30.
[3] Chesmore ED. Application of time domain signal coding, artificial neural networks to passive acoustical identification of animals. Appl Acoust 2001;62: 1359–74.
[4] Reby D, Lek S, Dimopoulos I, Joachim J, Lauga J, et al. Artificial neural networks as a classification method in the behavioural sciences. Behav Process 1997;40:35–43.
[5] Tyagi H, Hegde RM, Murthy RHA, Prabhakar A. Automatic identification of bird calls using spectral ensemble average voiceprints. In: Proceedings of the 13th European signal processing conference, Florence, Italy; 2006. p. 1–5.
[6] Vilches E, Escobar IA, Vallejo EE, Taylor CE. Data mining applied to acoustic bird species recognition. In: Proceedings of the 18th international conference on pattern recognition, Hong Kong, vol. 3; 2006. p. 400–3.
[7] Huang CJ, Yang YJ, Yang DX, Chen YJ. Frog classification using machine learning techniques. Exp Syst Appl 2009;36.
[8] Guo G, Li Z. Content-based classification and retrieval by support vector machines. IEEE Trans Neural Network 2003;14:209–15.
[9] Duellman WE, Trueb L. Biology of amphibians. New York: McGraw-Hill; 1986.
[10] Burton C. In: Glasby CG, Ross GJB, Beesley PL, editors. Family microhylidae in fauna of Australia, vol. 2A. Canberra: AGPS; 1993.
[11] Buddle CM, Beguin J, Bolduc E, Mercado A, Sackett TE. The importance and use of taxon sampling curves for comparative biodiversity research with forest arthropod assemblages. Can Entomol 2005;137:120–7.
[12] Sueur J, Pavoine S, Hamerlynck O, Duvail S. Rapid acoustic survey for biodiversity appraisal. PLoS ONE 2008;3(12):4065.
[13] Rényi A. On the dimension and entropy of probability distributions. Acta Math Acad Sci Hung 1959;10:193–215.
[14] Csiszár I. Generalized cutoff rates and Rényi's information measures. IEEE Trans Inform Theory 1995;41:26–34.
[15] Neemuchwala H, Hero A, Carson P. Image matching using alpha-entropy measures and entropic graphs. Signal Process 2005;85:277–96.
[16] Vinga S, Almeida JS. Rényi continuous entropy of DNA sequences. J Theory Biol 2004;231:377–88.
[17] Baraniuk R, Flandrin P, Janssen A, Michel O. Measuring time–frequency information content using the Rényi entropies. IEEE Trans Inform Theory 2001;47:1391–409.
[18] Kurths J, Voss A, Saparin P, Witt A, Kleiner HJ, Wessel N. Quantitative analysis of heart rate variability. Chaos 1995;1:88–94.
[19] Hart PE. Moment distributions in economics: an exposition. J Royal Stat Soc Ser A 1975;138:423–34.
[20] Gregorio AD, Iacus SM. On Rényi information for ergodic diffusion processes. Inform Sci 2009;179(3):279–91.
[21] Dietrich C, Palma G, Riedeb K, Schwenkera F. Classification of bioacoustic time series based on the combination. Pattern Recognit 2004;37:2293–305.

Please cite this article in press as: Han NC et al. Acoustic classification of Australian anurans based on hybrid spectral-entropy approach. Appl Acoust (2011), doi:10.1016/j.apacoust.2011.02.002