ENERGY, VIBRATION AND SOUND RESEARCH GROUP (e-VIBS)
SCHOOL OF SCIENCE AND TECHNOLOGY
UNIVERSITI MALAYSIA SABAH

BIOACOUSTICS SIGNAL MODELING USING TIME-FREQUENCY DISTRIBUTION

JEDOL DAYOU, NG CHEE HAN, HO CHONG MUN, ABDUL HAMID AHMAD, MOHD NOH DALIMIN & SITHI V. MUNIANDY

FINAL REPORT
FUNDAMENTAL RESEARCH GRANTS
No. FRGS0031-ST-1/2006
SEPTEMBER 2011

ABSTRACT

BIOACOUSTICS SIGNAL MODELING USING TIME-FREQUENCY DISTRIBUTION

Biodiversity is one of the major studies in bio-conservation, which enable to evaluate the quality of ecosystem in a specific area, especially for protected area. In order to monitor the quality of the ecosystem structure, a long term rapid diversity assessment is needed. In term of that, bioacoustics has been introduced as a beneficial method for local species richness estimation. However, this method is still in the infancy state and many improvements are needed for more practical purposes. This research is carried out to develop new bioacoustics species identification method with the improvement in the identification accuracy. The method which developed in this research is based on entropy principles, and implements on Fourier transform (FT) and wavelet transform (WT) of bioacoustics signal. Several entropy principles including Shannon, Rényi and Tsallis, are investigated which representing measurement of richness of the information contents and complexity of a bioacoustics signal.

To evaluate the new identification system, nine frog species from Microhylidae family was selected as test samples. Ten syllables were segmented from each frog sounds and characteristic of each syllables was extracted with the corresponding features which carried out in this research. All of the test samples were then sent into the *k*-nearest neighbour (*k*-NN) classifier for classification purpose. The *k*-NN classifier compared the test samples with the training data set in order to recognize and identify the frog species.

To establish a base trial data, the widely used spectral centroid (SC) and wavelet centroid (WC) were used as reference. The SC and WC of the syllables for each species were determined. It is found that, in terms of the average of classification accuracy for all test samples, the SC method has shown slightly better compared to the WC method. The classification results in average were 88.89% for the SC features and 86.67% for the WC features.

The entropy alone, if implemented on raw bioacoustics signal shows reduction rather than improvement in the identification accuracy compared to the reference (SC and WC). It is found for example that the average identification accuracy for Shannon entropy (SE), Rényi entropy (RE) and Tsallis entropy (TE) were 76.67%, 75.56% and 83.33%, respectively.

Due to the poor classification results of the entropy alone approaches, alternative methods were proposed in this work called wavelet entropy (WE). WE is a combination of interdisciplinary concepts between wavelet transform and entropy. In order to archive this, the entropies (SE, RE and TE) were extracted from three types of wavelet transforms, namely continuous wavelet transform (CWT), discrete wavelet transform (DWT) and wavelet packet decomposition (WPD), of a bioacoustics signal. In this work, two possible ways to extract the entropy from CWT were introduced, which were wavelet scale entropy (WSE) and wavelet time entropy (WTE). Entropy that extracted from DWT and WPD were called as discrete wavelet entropy (DWE) and wavelet packet entropy (WPE), respectively. The species identification results based on these WE features which extracted from the bioacoustics signals were then examined and compared. In term of SE approach, WPE has given the best classification result compared with others (WSE, WTE and DWE), which was over 98% of accuracy. However, WSE was the best method for the RE approach with the accuracy of 92%. Based on TE approach, WPE has shown the best result with the classification accuracy of 100%. In term of that, this research work has proven that the WPE is the best method in the TE approach for species identification on bioacoustics signals.

In conclusion, this work has successfully developed the species identification system based on bioacoustics signals by using the concept of WE. By comparing to the reference methods (conventional or classical methods), this study has proven that the performance of the bioacoustics species identification system can be improved by using the entropy approach with association of WT. Since the bioacoustics species identification system that proposed in this study is based on entropy approach, the computer algorithms is much easier (less complex) compared to the conventional methods, particularly based on spectrogram and sonogram. The proposed method can reduce the energy and time consumptions in terms of data processing.

ABSTRAK

PERMODELAN ISYARAT BIOAKUSTIK MENGGUNAKAN TABURAN FREKUENSI-MASA

Kepelbagaian biologi merupakan salah satu kajian utama dalam pemuliharaan biologi, bagi membolehkan untuk menilai kualiti ekosistem di tempat tertentu, terutamanya kepada kawasan yang dilindungi seperti hutan simpanan. Untuk pengawalan kualiti struktur ekosistem, kaedah penilaian secara jangka panjang dan pantas amat diperlukan. Dengan itu, kaedah bioakustik diperkenalkan sebagai kaedah yang memanfaatkan pengangaran kekayaan spesis tempatan. Akan tetapi, kaedah ini masih di tahap permulaan. Tujuan utama kajian ini adalah untuk memperkenalkan kaedah baru sistem pengecaman spesis bioakustik dengan meningkatkan ketepatan pengecaman. Kaedah yang dibangunkan dalam kajian ini adalah berdasarkan prinsip entropi, dan digunapakai ke atas transformasi Fourier dan transformasi wavelet isyarat bioakustik. Beberapa kaedah entropi termasuk Shannon, Rényi dan Tsallis dikaji dan digunakan sebagai sifat ('feature') untuk mengambarkan kandungan maklumat dan juga kerumitan ('complexity') isyarat bioakustik.

Sembilan spesis katak daripada keluarga Microhylidae dipilih sebagai sampel kajian bertujuan untuk menilai kecekapan sistem pengecaman spesis baru dalam kajian ini. Sepuluh suku kata ('syllable') adalah ditemberengkan ('segmented') daripada setiap bunyi spesis katak tersebut dan seterusnya sifat-sifat suku kata adalah direntapkan ('extracted') dengan kaedah seperti yang dicadangkan. Selepas kerentapan sifat ('feature extraction'), semua sampel kajian dikelaskan dengan menggunakan pengkelas k-NN. k-NN membandingkan sifat-sifat sampel kajian dengan set data latih ('traning data') untuk mengenalpasti spesis haiwan.

Kaedah yang biasa digunakan seperti 'spectral centroid' (SC) dan 'wavelet centroid' (WC) digunakan sebagai kaedah rujukan bagi tujuan membandingkan kaedah yang dicadangkan dalam kajian ini. SC dan WC daripada suku kata setiap spesis dikenalpasti. Dengan mengambil kira purata ketepan pengkelasan, didapati bahawa kaedah SC (88.89%) lebih tinggi berbanding kepada kaedah WC (86.6%).

Dengan mengunakan kaedah entropi sahaja ke atas isyarat bioakustik mentah, keputusan menunjukkan penurunan ketepatan pengecaman jika dibandingkan dengan kaedah rujukan (SC dan WC). Keadaan ini dapat dibuktikan dengan mengambil kira keputusan pengekelasan berdasarkan kaedah entropi Shannon (SE), entropi Rényi (RE) and entropi Tsallis (TE) adalah 76.67%, 75.56% dan 83.33%.

Oleh kerana keputusan pengkelasan kaedah entropi adalah kurang memuaskan, kaedah alternatif diperkenalkan dalam kajian ini dan bernama 'wavelet entropy' (WE). WE merupakan combinasi konsep daripada WT dan entropi. Dengan itu, entropi-entropi (SE, RE dan TE) adalah direntapkan daripada tiga jenis WT isyarat bioakustik, iaitu 'continuous wavelet transform' (CWT), 'discrete wavelet transform' (DWT) dan 'wavelet packet decomposition' (WPD). Dalam projek ini, dua cara perentapan entropi daripada CWT dijalankan, iaitu 'wavelet scale entropy' (WSE) dan 'wavelet time entropy' (WTE); dimana entropi yang direntapkan daripada DWT dan WPD dinamakan sebagai 'discrete wavelet entropy' (DWE) dan 'wavelet packet entropy' (WPE). Keputusan pengecaman spesis berdasarkan sifat-sifat WE adalah direntapkan daripada isyarat bioakustik dinilai dan dibandingkan. Dari segi kaedah SE, WPE memberikan keputusan yang terbaik berbanding kaedah lain (WSE, WTE dan DWE), iaitu ketepannya melebihi 98%. Walaubagaimanapun, WSE adalah kaedah yang terbaik jika diambil kira dari segi kaedah RE, iaitu 92% ketepatan. Dari segi TE, WPE memberikan keputusan yang terbaik, iaitu 100%. Dengan itu, adalah dibuktikan bahawa WPE merupakan kaedah yang terbaik dengan pendekatan TE dalam pengecaman spesis melalui isyarat bioakustik.

Kesimpulanya, kajian ini telah berjaya membangunkan sistem pengecaman spesis berdasarkan isyarat bioakustik dengan menggunakan konsep WE. Dengan membandingkan dengan kaedah rujukan, kajian ini telah membuktikan bahawa kecekapan sistem pengecaman bioakustik spesis dapat dipertingkatkan dengan menggunakan pendekatan entropi melalui bantuan WT. Didapati bahawa kaedah entropi ini adalah kurang rumit dari segi algoritma komputer berbandingkan kepada kaedah sedia ada, terutamanya yang berdasarkan kepada 'spectrogram' dan 'sonogram'. Kaedah yang dicadangkan ini dapat mengurangkan penggunaan tenaga dan masa dalam pemprosesan data.

CONTENTS

CHAPTER 1

INTRODUCTION

1.1 Introduction

Species identification using electronic instruments is defined as an application of general pattern recognition in which an unknown (specimen) is placed into one of a number of possible classes depending on features extracted from measurements on the species (Chesmore, 2004). One of the most difficult aspects of performing research in bioacoustics species recognition by machine is its interdisciplinary nature, and the tendency of most researchers to apply a monolithic approach to individual problems.

Generally, the development of automated bioacoustics recognition studies can be viewed from three aspects, which are: (1) the feature extraction, (2) classification method and (3) the animal species.

Features used in sound recognition applications are usually chosen such that they represent some meaningful characteristics (Huang *et. al.*, 2009). Selection of actual features used in recognition is a critical part for the recognition system. In term of bioacoustics recognition system, there are a lot of features already been discovered in the literatures. These features which introduced in the literatures are generally inspired from the work of speech recognition studies. These features normally can be categorized into two groups, time domain and frequency domain. 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; and 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. The selection of the classification tool is also can be seen as an important step to solve the pattern recognition problem. In term of bioacoustics classification, since not many studies in this field, only several pattern recognition methods can be found in the literatures, such as artificial neural networks (ANN) (Chesmore, 2001), data mining techniques, template matching method, artificial, *k*-nearest neighbour (*k*-NN) (Huang *et. al.*, 2009), fuzzy-*k*-nearest neighbour (Dietrich *et. al.*, 2003) and hidden Markov model (HMM) (Milone *et. al.*, 2009); again, these classifiers which introduced in the literatures for bioacoustics identification also influenced a lot by speech recognition methods (Rabiner and Juang, 1993). It is found that, similar to classification tools, there are only several animals are being studied for bioacoustics classification. The majority of the studies on bioacoustics for species identification can be found in literatures are mainly focused on some animal species, such as birds (Chesmore, 2001), frog (Huang *et al*, 2009), insects (Chesmore, 2004), whales (Clemins and Johnson, 2006), dolphin (Houser *et. al.*, 1999) and bats (Vaughan *et. al.*, 1996).

1.2 Research Objectives

So far, it is difficult to find an ideal bioacoustics species identification system. Each of the existing methods found in the literature has their own limitations. This study is carried out to develop new bioacoustics species identification with the improvement in the identification accuracy by using entropy approach. Several entropy properties including Shannon, Rényi and Tsallis, are investigated as a feature to characterize a bioacoustics signal. These entropy algorithms are implemented on several wavelet transforms of the signal to increase the accuracy. Fourier spectral centroid and wavelet centroid are also investigated and used as reference.

1.3 Contribution of the Research Work

As mentioned earlier, the development of bioacoustics species identification system helps in species identification, identification of individuals within a species and detection of the presence of animals. In other words, this system provides an opportunity to detect the appearance of new species in specific area and also the migration of certain animal species.

Besides that, it is believed that the bioacoustics species identification system can be used to improving the quality of ecosystem monitoring system with the properties of long term, long distant, low cost and rapid diversity assessment; and without invasion to the protected area during the activity of monitoring.

It is noted that the existing bioacoustics species identification systems have their own limitations. This research work provides an alternative tool to improve the efficiency of the bioacoustics species identification system. Besides that, this work explored the significant of entropy approach in bioacoustics signal analysis and pattern recognitions. Since the bioacoustics species identification system which proposed in this work is based on entropy approach, and the computer algorithms based on this method is much easier (less complex) compared to the existing methods, particularly based on spectrogram and sonogram. The proposed method can reduce the consumption of energy and time, and provide the opportunity to implement the application in hand-held or portable computer.

CHAPTER 2

OVERVIEW OF BIOACOUTICS SIGNAL PROCESSING

2.1 Signal Processing

2.1.1 What is Signal?

To define signal precisely is a difficult task, but in general, signal can be regarded that anything which carries information. Smith (1997) pointed out that a signal is a description of how one parameter is related to others. From communications, signal processing and electrical engineering point of views, signal is defined as any
time-varying or spatial-varying quantity. Signal can also be defined as any quantity that is measurable through time or over space. In information theory, a signal is defined as a codified message, which is the sequence of states in a communication channel that encodes a message. Examples of signals are human voice, chirping of birds, hand gestures (sign language), etc.

Signals can be classified based on their nature and characteristics in the time domain. They are broadly classified as 'deterministic' signal and 'nondeterministic' signal (Stearns and David, 1996). A deterministic signal is a signal that each value is uniquely specified by a mathematical expression, a table of data, or a rule of some sort. In other words, deterministic signals are functions that are completely specified in time. The nature and amplitude of such a signal at any time can be predicted; and, the pattern of the signal is regular and can be characterized mathematically (Salivahanan *et. al.,* 2000). The classes of deterministic signals are as follow:

1. **Periodic signals**: completely described as being specified within a signal components.
2. **Finite-duration signals**: are defined during some finite time interval and are undefined for all time outside that interval.
3. **Transient signals**: are those that are nonzero only within some finite time interval or vary over a short interval and then decay to a constant value.
4. **Almost periodic signals**: signals that composed of sums of sinusoids that are not harmonically related are almost periodic. This type of signal is not usually encountered in practical engineering applications.

One could argue that recorded data can never be truly deterministic due to unforeseen factors that affect the experiment. In many practical applications, however, a mathematical model of the recorded signal may be used to simplify the analysis with very little loss of accuracy. Contrary to deterministic signal, a nondeterministic signal is one whose occurrence is random in nature and its pattern is quite irregular. Unlike deterministic signal, nondeterministic signal (sometimes also known as stochastic or random signal) is not so nice. Nondeterministic signal cannot be characterized by a simple, well-defined mathematical equation and their future values cannot be predicted. In other words, nondeterministic or random signals is not completely predictable and is therefore described in terms of averages and other statistical properties (Stearns and David, 1996). A typical example of a nondeterministic signal is thermal noise in an electrical circuit. The behaviour of such a signal is probabilistic in nature and can be analysed only stochastically. Another example which can be easily understood is the number of accidents in a year. One cannot exactly predict what would be the figure in a particular year and this varies randomly (Salivahanan *et. al.,* 2000).

Figure 2.1 shows the examples of a periodic, transient and random signals, where the signal amplitude, $x(t)$ is plotted versus time, t with $t = 0$ marks the starting point of the signal. The periodic signal illustrated in Figure 2.1 (a) is the sum of two sinusoids, where one at twice the frequency of the other. The damped sinusoid in Figure 2.1 (b) is typically categorized as a transient signal even though as defined, $x_2(t)$ requires an infinite amount of time to decay to zero. The sinusoid corrupted by noise in

Figure 2.1 (c) is an example of a periodic signal corrupted by random noise, since the noise, $y(t)$ can be described only in terms of its statistical properties (Stearns and David, 1996).The three signal in Figure 2.1 are called 'analogue' or 'continuous' signal, since both x and t are continuous variables.

2.1.2 Introduction to Signal Processing

Signal processing is one of the fields in electrical engineering, systems engineering and applied mathematics, where it mainly deals with operations on or analysis of signals. These operations can be performed on discrete or continuous time signals, the signals are including sound, images, time-varying measurement values and sensor data.

The operations and algorithms which usually used in signal processing are listed as below (Moon & Stirling, 2000):

1. **Filtering** – for tone controls and equalizers
2. **Smoothing / deblurring** – for image enhancement
3. **Adaptive filtering** – for echo-cancellation in a conference telephone, or denoising for aircraft identification by radar.
4. **Spectrum analysis** – for magnetic resonance imaging and tomographic reconstruction.
5. **Digitalization, reconstruction and compression** – for image compression, sound coding and other source coding.
6. **Storage** – for digital delay lines and reverb.
7. **Feature extraction** – for speech-to-text conversion.
8. **Modulation** – in modems.
9. **Wavetable synthesis** – in modems and music synthesizers.
10. **Prediction**
11. **System identification or classification**
12. **A variety of other operations**

It is believed that, the principles of signal processing were started in the classical numerical analysis techniques in 17th century. The 'digitalization' or digital refinement of these techniques were found in the digital control systems in 1940s and 1950s (Oppenheim & Schafer, 1975).

Basically, signal processing can be categorized into three, which are analogue signal processing, discrete time signal processing and digital signal processing. Analogue signal processing is normally used for signals that have not been digitized. For examples, signals which found in classical radio, telephone, radar and television systems. In order to process these kinds of signals, several linear and non-linear electronic circuits are needed, such as passive filters, active filters, additive mixers, integrators and delay lines, compandors, multiplicators (frequency mixers and voltage-controlled amplifiers), voltage-controlled filters, voltage-controlled oscillators and phase-locked loops.

Discrete time signal processing is used to sampled signals in discrete points in time or quantized in time, but not in magnitude. The main concept for this technique is to establish a mathematical basis for digital signal processing, without taking quantization error into consideration.

Digital signal processing is used to process digitized signals. Processing is normally done by computers or by digital circuits, including ASICs, filed-programmable gate arrays or specialized digital signal processors (DSP chips). There are numbers of arithmetical operations have been used for this processing purpose, such as fixed-point and floating-point, real-valued and complex-valued, multiplication and addition. The operations which supported by hardware are including circular buffers and look-up tables. Examples for these algorithms are including fast Fourier transform (FFT), finite impulse response (FIR) filter, infinite impulse response (IIR) filter, Wiener filter and Kalman filter.

There are many fields which related to signal processing, including:

1. **Statistical signal processing** – for analyzing and extracting information from signals and noise based on their stochastic properties.
2. **Audio signal processing** – for electrical signals representing sound, such as speech or music.
3. **Speech signal processing** – for processing and interpreting spoken words.
4. **Image processing** – in digital cameras, computers and various imaging systems.
5. **Video processing** – for interpreting moving pictures.
6. **Array processing** – for processing signals from arrays of sensors.
7. **Filtering** – used in many fields to process signals
8. **Seismic signal processing**
9. **Data mining**

2.2 Pattern Recognition

A key question in animal sound recognition is how bioacoustics patterns are compared to determine their similarity (or equivalently, the distance between patterns). Depending on the specifics of the recognition system, pattern comparison can be done in a wide variety of ways. The term 'pattern recognition' can be referred to as the task of placing some object to a correct class based on the measurement of the properties from the object. Usually this task is performed automatically with the help of computer. In this context, objects is recognised, measured and classified into possible classes. Webb (2002) defines that 'statistical pattern recognition' is a term used to cover all stages of an investigation from problem formulation and data collection through to discrimination and classification, assessment of results and interpretation. Some of the basic terminology is introduced as two complementary approaches to discrimination described. Thus, the system which makes the measurements on certain objects and thereafter classifies these objects is called a 'pattern recognition system'.

Pattern recognition as a field of study developed significantly in the 1960s and 1970s (Kuncheva, 2004). It is a very much an interdisciplinary subject, covering developments in the areas of statistics, engineering, artificial intelligence, computer science, psychology and physiology, among others. Some people entered the field with a real problem to solve. The large number of applications, ranging from the classical ones such as automatic character recognition and medical diagnosis to the more recent ones in data mining (such as credit scoring, consumer sales analysis and credit card transaction analysis), have attracted considerable research efforts, with many methods developed and advances made (Ullmann, 1973). For an example, a spam filter recognises automatically junk e-mails and send them in a different folder than the user's inbox. Table 2.1 shows a list of examples of pattern recognition applications. Other researchers were motivated by the development of machines with 'brain-like' performance that in some way could emulate human performance. There were many over-optimistic and unrealistic claims made, and to some extent exist strong parallels with the growth of research on knowledge-based systems in the 1970s and neural networks in the 1980s (Webb, 2002).

Nevertheless, within these areas significant progress has been made, particularly where the domain overlaps with probability and statistics, and within recent years there have been many exciting new developments, both in methodology and applications. These build on the solid foundations of earlier research and take advantage of increased computational resources readily available nowadays. These developments include, for example, kernel-based methods and Bayesian computational methods (Ullmann, 1973).

In general, the pattern recognition systems can be categorized into two types, which are 'supervised classification' and 'unsupervised classification'. In supervised classification (or discrimination), the classifier perform the work depend on the given feature vector in order to decide the class of the object. The given feature vector is represented as prototypes or training samples. Artificial neural networks (ANN) and k-NN are the examples of supervised classifiers. In contrast, unsupervised classification (also called clustering) which is no explicit teacher nor training samples. The classification work is done based on the similarity between the feature vectors in order to divide the objects in to group. *k*-means clustering is one of the popular unsupervised classification and has been employed to solve variety of pattern recognition problems (Rabiner and Juang, 1993).

2.2.1 Basic Structure of Pattern Recognition Systems

As mentioned earlier, pattern recognition system is to classify an object into a correct class based on the measurements on the particulars of the object (Kuncheva, 2004). It is noted that the possible classes are usually well-designed before the development of the pattern recognition system. Generally, there are three stages in many pattern recognition systems, including (1) sensing (or measurement), (2) feature extraction and (3) classification.

Sensing is referred to some measurement or observation about the object to be classified. For example, the data can consist of sounds or images and sensing equipment can be a microphone array or a camera. In some cases, pre-processing is needed to filtering the raw data for noise suppression and other operation performed on the raw data in order to improve the quality. Similar to segmentation, also may require on certain pattern recognition processes, the measurement data is partitioned in this process, so that each part represents exactly one object to be classified. The results from this process can be represented as a vector, or called 'representation pattern'.

In the second stage, feature extraction is the process to search the information in order to characterize the data for classification, where the result of this stage is called the 'feature pattern' (Kuncheva, 2004); the space of all possible feature vectors is namely 'feature space'. Feature extraction is the key for identification. It is the most important component in designing the intelligent system. A feature extractor should reduce the vector (i.e., the original waveform) to a lower dimension, which contains most of the useful information from the original vector (Avci, 2009). In face recognition, for example, the principal component analysis (PCA) has widely been used to reduce the number of features. PCA is a statistical technique to reduce the dimensionality of a data vector while retaining most of the information that the data vector contains. In general, the link between feature extraction and classification is fuzzy. The task of the feature extractor is to produce a 'fingerprint' about the data in order to classify the object easily. Besides that, the task of the classifier is to perform the best classification accuracy based on the given features.

The classifier is used to measure the input data based on the given feature vector which extracted from the object to be classified. The object is then placed to the most appropriate class. Note that the classifier is unable to distinguish between two objects with the same feature vector. Last but not least, decision in term of pattern recognition system is to be made upon an action based on the classification results (Rabiner and Juang, 1993).

The pattern classification process described above is illustrated in Figure 2.2, where it is grossly oversimplified the pattern classification procedure. Data may undergo several separate transformation stages before a final outcome is reached. These transformations (sometimes termed pre-processing, feature selection or feature extraction) operate on the data in a way that usually reduces its dimension (reduces the number of features), removing redundant or irrelevant information, and transforms it to a form more appropriate for subsequent classification. The term 'intrinsic dimensionality' refers to the minimum number of variables required to capture the structure within the data (Webb, 2002).

Terminology varies between authors. Sometimes the term 'representation pattern' is used for the vector of measurements made on a sensor with the term 'feature pattern' being reserved for the small set of variables obtained by transformation (by a feature selection or feature extraction process) of the original vector of measurements. In some problems, measurements may be made directly on the feature vector itself. In these situations there is no automatic feature selection stage, with the feature selection being performed by the investigator who 'knows' (through experience, knowledge of previous studies and the problem domain) those variables that are important for classification. In many cases, however, it will be necessary to perform one or more transformations of the measure data (Webb, 2002).

2.2.2 Application of Pattern Recognition in Bioacoustics Studies

Bioacoustics recognition and animal sound analysis methods are inspired from the idea of human speech recognition. Automatic recognition of speech by machine has been a goal of research for more than four decades and has inspired such science fiction wonders as the computer HAL in Stanley Kubrick's famous movie 2001 – *A Space Odyssey* and the robot R2D2 in the George Lucas classic *Star Wars* series of movies (Rabiner and Juang, 1993). There are many useful algorithms for automatic speech recognition that have been developed like linear prediction, overlap-and-add synthesis, Mel frequency cepstral coefficients (MFCC), and audio codec like code-excited linear prediction. These algorithms have significantly influenced our lifestyle, especially in our communication media systems. Some more there are several studies have applied the application of speech recognition system into the robotic sensor industries. While the analyzing human speech methods are getting advanced, analysis for bioacoustics signals (mainly for animal sound signals or animal calls in this study) has remained in the static state.

There are some research activities have been done in the bioacoustics field, including sound signal structure analysis, bioacoustics classification and species identification based on vocalization. In the animal sound signal structure analysis, investigations are mainly carried out on the sound signals by using time domain and frequency domain approaches. Researchers were trying to analysis the sound signals and extracted out the features (or characteristics) from it, namely feature extraction. These extracted features serve as an identity or attributes, which enable us to classified the sound signals and find the uniqueness between them (some time also called fingerprint). This is how the animal species can be identified from their calling.

Chesmore (2004) has discussed that, species identification by electronic means is an application in general pattern recognition in which an unknown (sound sample in this context) is placed into one of a number of possible classes depending on features extracted from measurements on the species.

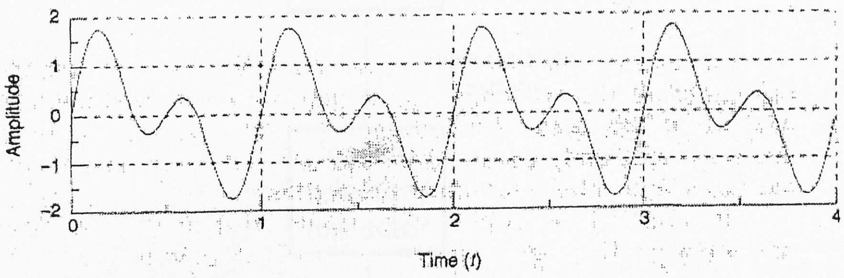
Several feature extraction methods have been introduced in literatures, as mentioned before, which are basically can be categorized into two groups, which are time domain and frequency domain. For time domain method, the feature extractions are including pulse distance density (Dietrich, 2003), temporal structure of the pulses, pulse length and zero-crossing rate (Chesmore, 2004). Frequency domain is the favourite analysis method for researchers, and this is because frequency domain is easier to be visualized and understanding the behaviour of the sound signals, particularly FFT and wavelet transform. For further extraction on the relevant frequency there are including frequency contour of pulses, energy contour of pulses, pulse frequency, linear prediction coding (LPC). On the other hand, time domain techniques are preferred for remote monitoring system, and this is because of frequency domain techniques are computationally intensive and difficult to implement on low cost microcontroller-based systems (Chesmore, 2001). It is noticed that the remotely-sited PC-based system only able to record about 75% of the time, and 25% is devoted to signal processing (Chesmore, 2001). Frequency domain and time domain can be combined (hybrid method) in order to get the better result for characterization sound signal (Dietrich, 2003; Huang, 2009).

The purpose of classification is to differentiate and to identify the specimen from a set of samples when the characteristics or extracted features are given. Several classification methods have been discovered and proposed in literatures, particularly discriminant function analysis (DFA) classifier commonly used in the bat literatures; Gaussian mixture models (GMM) classifier, hidden Markov models (HMM) classifier used by Milone *et. al.*, (2009), *k*-Nearest Neighbour (*k*-NN) for automated frog sound classification (Huang *et. al.*, 2009) and cricket species (Dietrich, 2003); and artificial neural networks (ANN) implemented by Chesmore (2004) which tested on bird and insect sound signals.

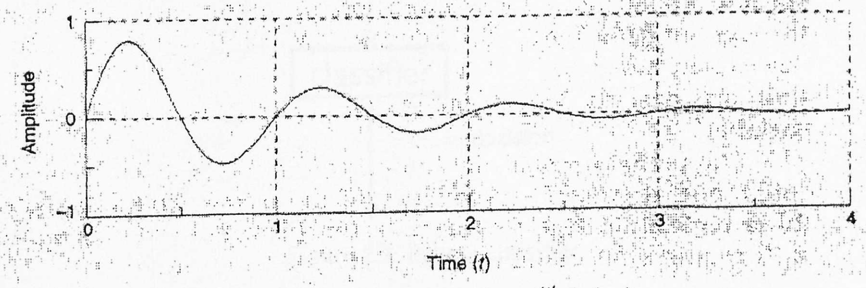
Dietrich has noted general problem in various pattern recognition applications when several classifiers were combined (Dietrich, 2003). In his study, combination of different classifiers, combination of classifiers trained on different data sets and combination of classifiers trained on different feature subsets are tested. He concluded that it is possible to improve the combination classifiers performance under certain conditions. Although most of the existing studies have shown a high recognition rate from the developed classifier for bioacoustics signals, there are still several occurrences of weaknesses. For example, Dietrich (2003) has discussed that the best result for cricket species classification by using single feature and combination of all suggested local features (as discussed in the literature) is 28% error rate and 6-10% error rate, respectively. Yet, Chesmore's (2001) study has shown a better result by using time domain (feature extraction) and ANN (classifiers) and achieved 100% success and zero misidentification for cricket sounds signals. However, the suggested method is only applicable under low noise condition.

Table 2.1: A list of example of pattern recognition applications.

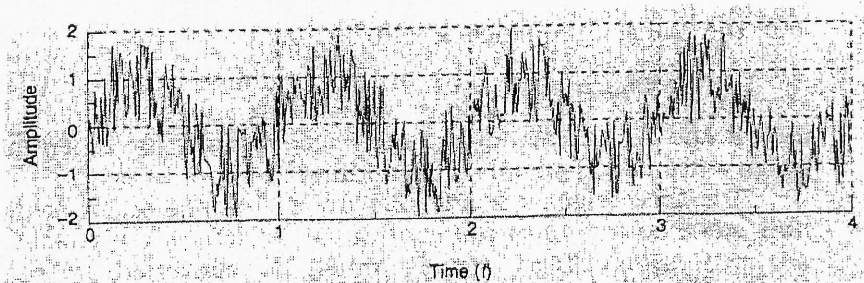
Problem Domain	Application	Input Pattern	Pattern Classes
Document image analysis	Optical character recognition	Document image	Characters, words
Document classification	Internet search	Text document	Semantic categories
Document classification	Junk mail filtering	Email	Junk/non-junk
Multimedia database retrieval	Internet search	Video clip	Video genres
Speech recognition	Telephone directory assistance	Speech Waveform	Spoken words
Natural language processing	Information extraction	Sentences	Parts of speech
Biometric recognition	Personal identification	Face, iris, fingerprint	Authorized users for access control
Medical	Computer aided diagnosis	Microscopic image	Cancerous/healthy cell
Military	Automatic target recognition	Optical or infrared image	Target type
Industrial automation	Printed circuit board inspection	Intensity or range image	Defective/non-defective product
Industrial automation	Fruit sorting	Images taken on a conveyor belt	Grade of quality
Remote sensing	Forecasting crop yield	Multispectral image	Land use categories
Bioinformatics	Sequence analysis	DNA sequence	Known types of genes
Data mining	Searching for meaningful patterns	Points in multidimensional space	Compact and well-separated clusters



(a) Periodic signal: $x_1(t) = \sin(\omega t) + \sin(2\omega t)$



(b) Transient signal: $x_2(t) = e^{-\omega t} \sin(\omega t)$



(c) Random signal: $x_3(t) = \sin(\omega t) + y(t)$

Figure 2.1: Examples of basic types of signals (Stearns and David, 1996).

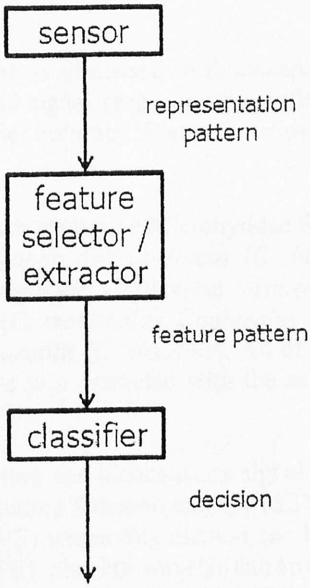


Figure 2.2: Pattern classifier.

CHAPTER 3

METHODOLOGY

The methodology of this research work is discussed in this chapter. The experimental work basically consists of five stages, which are sound signal preprocessing, syllable segmentation, feature extraction, classification and assessment of classifier accuracy. Figure 3.1 shows the flow chart for this experimental work.

First, a collection of frog sounds from nine Microhylidae frogs were prepared. The selected frog species in this study are including *Cophixalus bombiens* (*C. bombiens*), *Cophixalus concinnus* (*C. concinnus*), *Cophixalus exiguus* (*C. exiguous*), *Cophixalus hosmeri* (*C. hosmeri*), *Cophixalus infacetus* (*C. infacetus*), *Cophixalus monticola* (*C. monticola*), *Cophixalus neglectus* (*C. neglectus*), *Cophixalus ornatus* (*C. ornatus*) and *Cophixalus saxatilis* (*C. saxatilis*). All of the frog sounds were then segmented into syllables and each of the syllables was extracted with the selected features in this work (refer to Appendix A to details on segmentation).

In order to develop new feature for bioacoustics signal identification, several features were investigated in this research work, including Shannon entropy (SE), Rényi entropy (RE), Tsallis entropy (TE), continuous wavelet entropy (CWE) where this method can be divided into wavelet scale entropy (WSE) and wavelet time entropy (WTE), discrete wavelet entropy (DWE) and wavelet packet entropy (WPE); where spectral centroid (SC) and wavelet centroid (WC) were used as reference methods in order to compare the performances of the classifiers.

After the feature extraction, all of the test samples were sent into the *k*-NN classifier to train the recognition system (refer to Appendix B to details on *k*-NN classifier). The *k*-NN classifier compared the test samples with the training data set in order to recognise and identify the frog species. The performance of classifier in recognizing the species in term of accuracy was then measured. The following classification accuracy is used to examine the performance of the proposed work:

$$Accuracy = \frac{N_c}{N_T} \times 100, \tag{3.1}$$

where N_c is the number of syllables which were recognised correctly and N_T is the total number of test syllables.

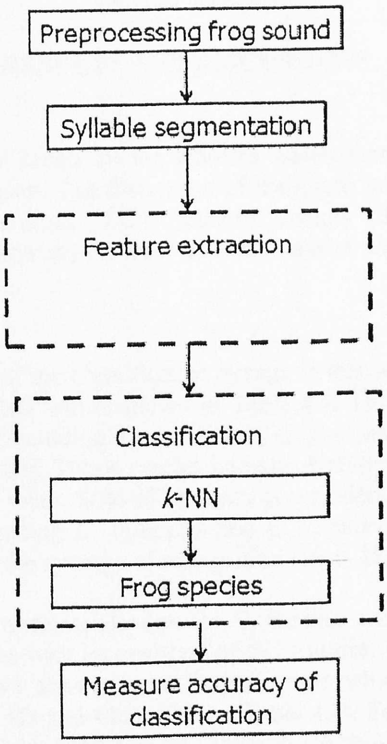


Figure 3.1: Architecture of the frog sound classification system.

CHAPTER 4

RESULTS AND DISCUSSIONS

4.1 Introduction

The performances of the classifier based on the features which mentioned in the earlier chapter are discussed and compared in this chapter. The discussion of the results are based on the features, which are spectral centroid (SC), wavelet centroid (WC), Shannon entropy (SE), Rényi entropy (RE), Tsallis entropy (TE), continuous wavelet entropy (CWE), discrete wavelet entropy (DWE) and wavelet packet entropy (WPE).

4.2 Spectral Centroid

Figure 4.1 shows the performance of the classification system in this work where the classification work is based on the selected training data which shown in Table 4.1. The results reveal that seven species were identified, where they were including *C. bombiens*, *C. concinnus*, *C. exiguus*, *C. monticola*, *C. neglectus*, *C. ornatus* and *C. saxatilis*. These species had successfully recognised by the classifier with more than 80% of accuracy. In this work, 80% of accuracy is considered success. In contrast, two species are misclassified in this case, including *C. infacetus* and *C. hosmeri* with less than 80% of accuracy. Based on the classification results, the average of recognition rate is 88.89%.

Two reasons explain the misclassification for *C. hosmeri* and *C. infacetus*. First, the classifier unable to recognise to those species with inconsistent of SC features. As can be seen, the SC values for 10 syllables from *C. hosmeri* have shown six different values which are 4553.10 Hz, 4575.80 Hz, 4650.10 Hz, 4683.20 Hz, 4747.60 Hz and 4753.70 Hz. (Table 4.2). For the case of *C. infacetus*, the SC values for 10 syllables are 3553.60 Hz, 3589.70 Hz, 3598.80 Hz, 3607.10 Hz, 3617.00 Hz and 3681.20 Hz. These values are not unique and therefore misclassified. Secondly, the features which varied in large range and overlapped with other species features. For examples, the classifier unable to differentiate between *C. hosmeri* and *C. bombiens* with the SC values at 4650.10 Hz and 4651.20 Hz, respectively. Similar to the case of *C. infacetus* with the SC values at 3617.00 Hz which simply closed to the SC values of *C. exiguus* at 3617.60 Hz. These species has a very similar SC value and therefore misclassified.

Table 4.1: Training data set of spectral centroid in *k*-NN classifier for patter recognition.

Frog species	Spectral centroid, f_c (Hz)
bombiens	4650
concinnus	2950
exiguus	3630
hosmeri	4670
infacetus	3600
monticola	2840
neglectus	3010
ornatus	2760
saxatilis	1390

Table 4.2: Spectral centroid of segmented syllables for nine Microhylidae frogs.

	Spectral centroid, f_c (Hz)								
No.	*bombiens*	*concinnus*	*exiguus*	*hosmeri*	*infacetus*	*monticola*	*neglectus*	*ornatus*	*saxatilis*
1	4651.20	2971.60	3746.80	4753.70	3673.30	2842.40	3014.60	2756.30	1388.90
2	4651.20	2971.60	3660.60	4575.80	3598.80	2842.40	3014.60	2756.30	1388.90
3	4651.20	2928.50	3617.60	4747.60	3617.00	2842.40	3014.60	2756.30	1388.90
4	4651.20	2928.50	3746.80	4683.20	3681.20	2842.40	3014.60	2756.30	1388.90
5	4651.20	2928.50	3617.60	4650.10	3553.60	2842.40	3014.60	2756.30	1378.10
6	4651.20	2928.50	3617.60	4553.10	3589.70	2842.40	2928.50	2756.30	1388.90
7	4651.20	2971.60	3617.60	4747.60	3607.10	2842.40	3014.60	2756.30	1388.90
8	4651.20	2928.50	3660.60	4553.10	3673.30	2842.40	3014.60	2756.30	1388.90
9	4651.20	2928.50	3617.60	4650.10	3598.80	2842.40	3014.60	2756.30	1378.10
10	4651.20	2971.60	3617.60	4683.20	3553.60	2842.40	3014.60	2756.30	1388.90

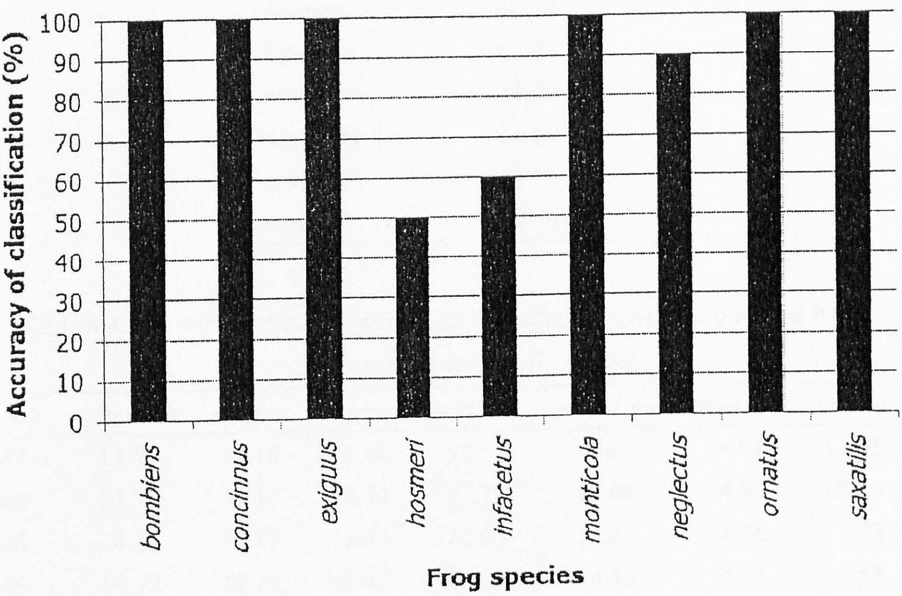


Figure 4.1: Performance of the *k*-NN classifier with spectral centroid for nine Microhylidae frog species.

4.3 Wavelet Centroid

By using the listed training data in Table 4.3, the classifier has successfully identified seven species, including *C. bombiens, C. concinnus, C. exiguus, C. infacetus, C. monticola, C. ornatus* and *C. saxatilis* with 100% of accuracy. In contrast, there are two species are misclassified, including *C. hosmeri* and *C. neglectus* with 40% of accuracy of recognition rate (see Figure 4.2)

The misclassification of *C. hosmeri* is expected in this work due to the WC features were exhibited in the same range with *C. bombiens* (see Table 4.4). For example, there were six WC features for *C. hosmeri* varied in the range of 12.44 – 12.47 and these features seem to be closer to the training data of *C. bombiens* (which has training data with 12.46). The same situation occurs to *C. neglectus*, where the misclassification of this species was due to the features values which were exhibited in the large range and overlapped to other species features. By referring to Table 4.4, it is proven that the WC for *C. neglectus* were overlapped to the features of *C. concinnus* and *C. monticola*. For examples, the *C. neglectus* WC features which varied at 14.28 – 14.37 and 14.90 – 14.91 are recognised as *C. monticola* and *C. consinnus*, respectively, seeing as these WC features are closer to the training data of *C. monticola* and *C. consinnus*.

Despite the pattern of energy distribution were similar (see Figure 4.3), somehow, by using the WC approach, the majority of the test samples are identified. Overall, the average of recognition rate in this work is 86.67%.

Table 4.3: Training data set of wavelet centroid in *k*-NN classifier for pattern recognition.

Frog species	**Wavelet centroid, χ_c (scale)**
bombiens	12.46
concinnus	14.85
exiguus	12.18
hosmeri	12.50
infacetus	12.30
monticola	14.40
neglectus	14.50
ornatus	15.15
saxatilis	22.90

Table 4.4: Wavelet centroid of segmented syllables for nine Microhylidae frogs.

	wavelet centroid, χ_c (Scale)								
No	*bombiens*	*concinnus*	*exiguus*	*hosmeri*	*infacetus*	*monticola*	*neglectus*	*ornatus*	*saxatilis*
1	12.47	14.77	12.18	12.44	12.31	14.40	14.91	15.08	22.88
2	12.48	14.77	12.18	12.44	12.32	14.40	14.90	15.16	22.93
3	12.46	14.81	12.18	12.45	12.33	14.42	14.28	15.23	22.84
4	12.46	14.92	12.19	12.47	12.34	14.42	14.28	15.18	22.84
5	12.46	14.99	12.20	12.52	12.34	14.39	14.28	15.13	22.88
6	12.46	14.91	12.18	12.48	12.32	14.39	14.37	15.14	22.91
7	12.46	14.77	12.18	12.45	12.33	14.39	14.49	15.23	22.84
8	12.48	14.91	12.18	12.48	12.31	14.39	14.51	15.18	22.84
9	12.45	14.99	12.18	12.52	12.34	14.42	14.45	15.13	22.88
10	12.46	14.77	12.20	12.47	12.33	14.40	14.47	15.14	22.91

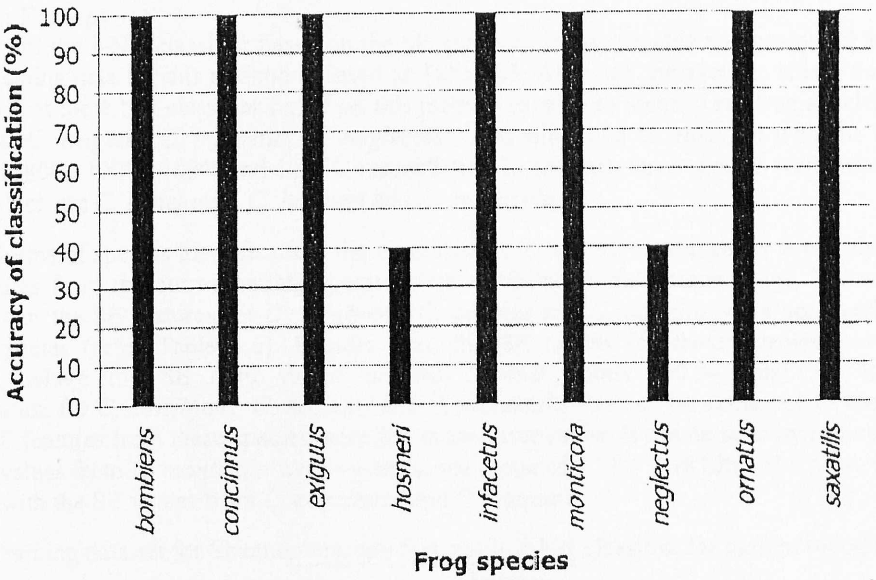


Figure 4.2: Performance of the *k*-NN classifier based on wavelet centroid approach for nine Microhylidae frog species.

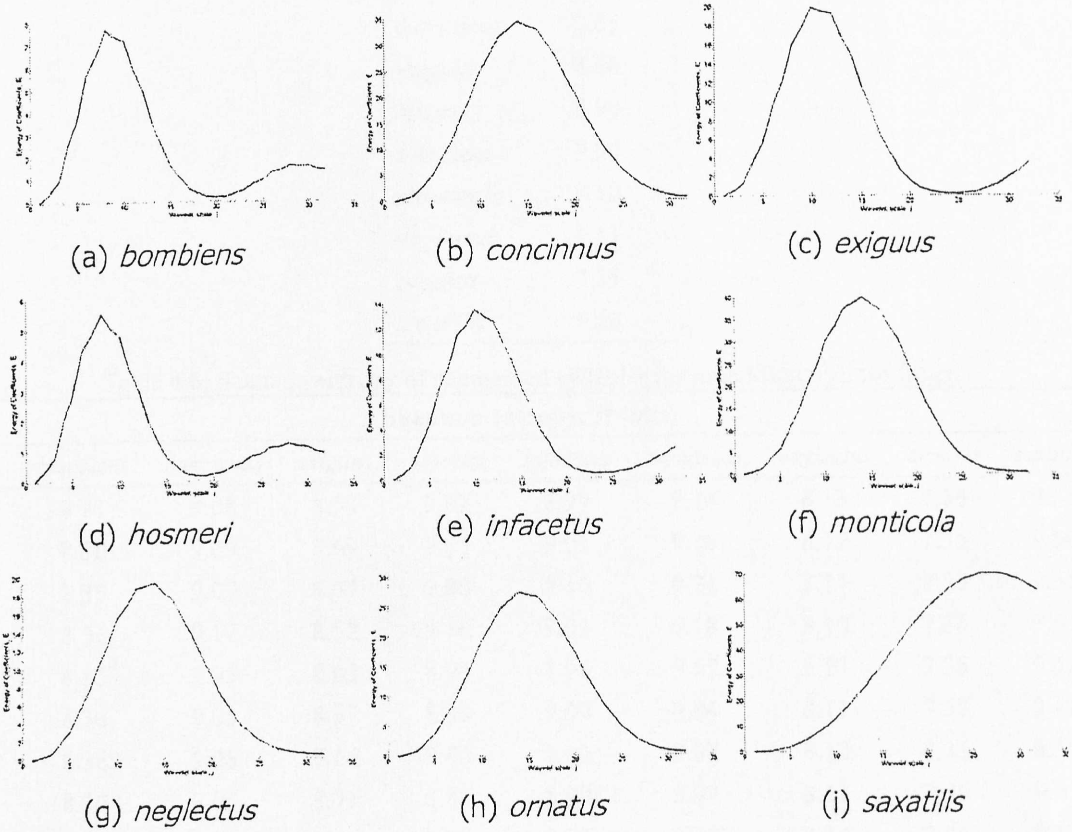


Figure 4.3: Examples of energy distribution of coefficients by scale of CWT on nine Microhylidae frogs segmented syllables.

4.4 Shannon Entropy

The accuracy of the *k*-NN classifier based on the SE approach is investigated and presented in Figure 4.4 where the training data for this method is listed as Table 4.5. Although the average is less than 80%, the results show that the *k*-NN classifier based on this method capable to identify six frog species, including *C. bombiens*, *C. exiguus*, *C. infacetus*, *C. neglectus*, *C. ornatus* and *C. saxatilis* with the accuracy at 100%, 100%, 80%, 100%, 100% and 100 %, respectively. In contrast, three species were misclassified in this work, which are *C. concinnus, C. hosmeri* and *C. monticola*.

In terms of species identification, the main reason for the successful cases is mainly because of the SE features from the corresponding species were exhibited in the unique range. For examples, as mention earlier, the SE features for *C. bombiens, C. ornatus* and *C. saxatilis* were not overlapped with any other species (refer Table 4.6). Besides that, the SE results for these species were exhibited consistently, where the SE only varied in two decimal points (±0.04 bits). In contrast, the misclassification for *C. concinnus, C. hosmeri* and *C. monticola* also can be explained in the same way, where the SE features from these species were fall in the same range. It can be seen by referring to Table 4.6, the SE values from *C. monticola* were varied in the range of 8.97 – 9.18 bits, which were also in the same range with the SE values from *C. concinnus* and *C. hosmeri.*

Table 4.5: Training data set for Shannon entropy features in *k*-NN classifier for pattern recognition.

Frog species	Shannon Entropy, H (bits)
bombiens	8.34
concinnus	9.05
exiguus	8.60
hosmeri	8.90
infacetus	9.00
monticola	9.10
neglectus	8.11
ornatus	7.35
saxatilis	9.50

Table 4.6: Shannon entropy of segmented syllables for nine Microhylidae frogs.

	Shannon entropy, H (bits)								
No.	*bombiens*	*concinnus*	*exiguus*	*hosmeri*	*Infacetus*	*monticola*	*neglectus*	*ornatus*	*saxatilis*
1	8.31	9.08	8.53	9.02	8.99	9.18	8.13	7.36	9.55
2	8.30	9.09	8.69	9.07	9.01	9.16	8.12	7.35	9.50
3	8.35	9.07	8.67	9.06	9.10	9.21	8.11	7.35	9.55
4	8.36	9.12	8.52	9.16	9.03	9.18	8.10	7.36	9.51
5	8.35	8.95	8.65	8.96	8.96	9.07	8.11	7.36	9.53
6	8.36	9.05	8.67	8.89	9.00	9.04	8.11	7.37	9.49
7	8.36	9.08	8.66	9.06	9.00	8.97	8.12	7.35	9.55
8	8.30	9.05	8.71	8.89	8.99	8.97	8.11	7.36	9.51
9	8.32	8.95	8.70	8.96	9.01	9.18	8.14	7.36	9.53
10	8.33	9.09	8.65	9.16	8.96	9.15	8.14	7.37	9.49

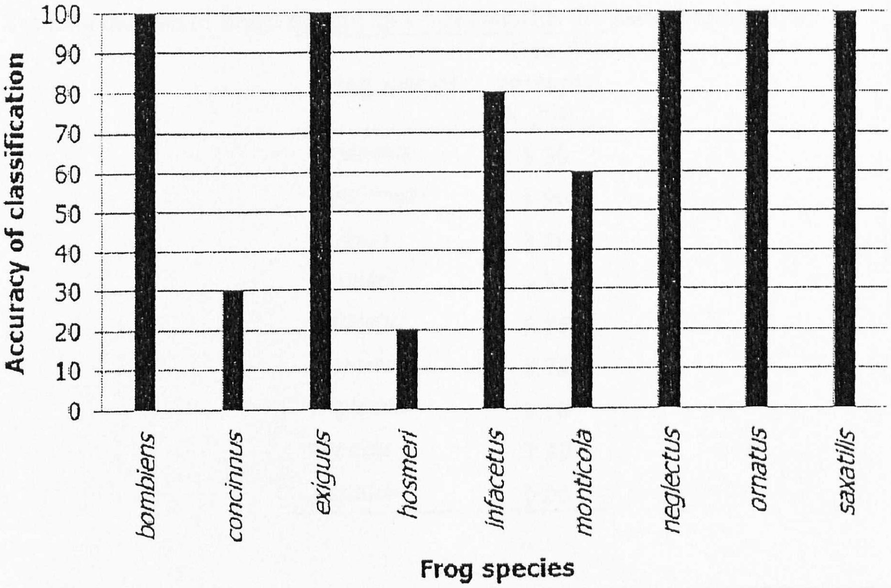


Figure 4.4: Performance of the *k*-NN classifier with Shannon entropy for nine Microhylidae frog species.

4.5 Rényi Entropy

The results of the frog sound signal based on RE feature are listed as Table 4.7. Based on the classification results, in term of 80% as success rate, there are only five species that correctly identified by using RE, which are *C. bombiens, C. exiguus, C. neglectus, C. ornatus* and *C. saxatilis*. Meaning to say that there is four species are not identified (less than 80% of accuracy) by the classifier, which are *C. concinnus, C. hosmeri, C. infacetus* and *C. monticola* (see Figure 4.5) where the training data is llisted as Table 4.8.

It is found that, all of the species which are not identified in this work are from the group of high complexity range. As mentioned earlier, the RE results from these species are exhibited similarly and also overlapped to each other. Meaning to say, based on the RE feature, the classifier in general unable to find the uniqueness between these species.

Table 4.7: Rényi entropy of segmented syllables for nine Microhylidae frogs.

	Rényi entropy (bits)								
No.	*bombiens*	*concinnus*	*exiguus*	*hosmeri*	*infacetus*	*monticola*	*neglectus*	*ornatus*	*saxatilis*
1	8.28	8.96	8.24	8.84	8.86	8.91	8.12	7.36	9.10
2	8.27	8.95	8.41	8.87	8.88	8.90	8.11	7.34	9.00
3	8.33	8.93	8.44	8.91	8.99	8.95	8.09	7.34	9.05
4	8.34	8.99	8.21	9.01	8.90	8.91	8.08	7.36	9.03
5	8.32	8.76	8.39	8.81	8.80	8.79	8.09	7.36	9.05
6	8.34	8.91	8.46	8.73	8.85	8.73	8.09	7.37	8.99
7	8.34	8.96	8.43	8.91	8.86	8.64	8.10	7.34	9.05
8	8.26	8.91	8.42	8.73	8.86	8.67	8.09	7.36	9.03
9	8.29	8.76	8.43	8.81	8.88	8.91	8.13	7.36	9.05
10	8.31	8.95	8.40	9.01	8.80	8.99	8.13	7.37	8.99

Table 4.8: Training data set of Rényi entropy in k-NN classifier for pattern recognition.

Frog species	Rényi entropy, R (bits)
bombiens	8.30
concinnus	8.90
exiguus	8.40
hosmeri	8.85
infacetus	8.86
monticola	8.75
neglectus	8.10
ornatus	7.35
saxatilis	9.00

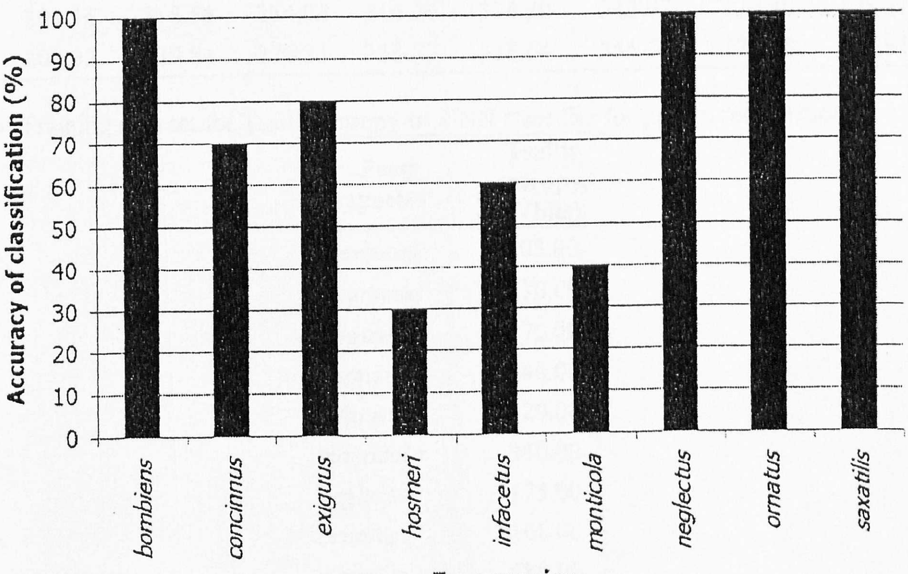


Figure 4.5: Performance of the k-NN classifier with RE features for nine Microhylidae frog species.

4.6 Tsallis Entropy

The TE results and the corresponding training in this study is listed in Table 4.9 and Table 4.10, respectively. Based on the classification results shows in Figure 4.6, the TE values from the species which exhibited in the small range are all identified in this work, which consists of *C. bombiens, C. monticola, C. ornatus* and *C. saxatilis*. Although the TE values from *C. exiguus* are exhibited in the large range, the classifier still able to identify it due to the property of the feature values which exhibited in the unique range. For the case of *C. neglectus,* where the corresponding TE values also exhibited in the wide range, it was successfully identified due to the consistency of the feature values (see Table 4.9). On the other hand, the misclassification for *C. concinnus* and *C. hosmeri* just simply because of their feature values are exhibited in the large range and overlapped to each other.

Table 4.9: Tsallis entropy of segmented syllables for nine Microhylidae frogs.

No. of syllable	Tsallis entropy, T (bits)								
	bombiens	*concinnus*	*exiguus*	*hosmeri*	*infacetus*	*monticola*	*neglectus*	*ornatus*	*saxatilis*
1	200.02	336.75	261.68	335.51	321.10	382.44	176.52	108.31	584.44
2	199.45	339.84	285.78	346.08	324.30	379.78	175.93	107.69	578.34
3	204.04	335.55	276.38	340.05	340.01	388.36	174.76	107.69	596.08
4	204.61	343.67	261.56	358.73	328.10	383.99	174.17	108.31	578.92
5	203.48	318.56	276.19	316.26	316.64	362.74	174.76	108.31	587.15
6	205.19	331.27	276.49	304.30	324.22	359.36	174.76	108.91	577.56
7	205.19	336.75	274.75	340.05	323.15	348.49	175.34	107.69	596.08
8	198.87	331.27	286.98	304.30	321.10	347.58	174.76	108.31	578.92
9	201.16	318.56	284.78	316.26	324.30	383.97	177.10	108.31	587.15
10	202.32	339.84	276.21	358.73	316.64	355.79	177.10	108.91	577.56

Table 4.10: Training data set for Tsallis entropy in k-NN classifier for pattern recognition.

Frog species	Tsallis entropy, T (bits)
bombiens	203.00
concinnus	330.00
exiguus	270.00
hosmeri	340.00
infacetus	320.00
monticola	350.00
neglectus	175.00
ornatus	108.00
saxatilis	580.00