Energy-aware Scheme for Animal Recognition in Wireless Acoustic Sensor Networks

Afnan Algobail, Adel Soudani and Saad Alahmadi
Department of Computer Science, College of Computer and Information Science, King Saud University, Riyadh, Saudi Arabia

Keywords: WASN, Object Recognition, Acoustic Sensing, Feature Extraction, Low-power Recognition.

Abstract: Wireless Acoustic Sensor Networks (WASN) have drawn tremendous attention due to their promising potential audio-rich applications such as battlefield surveillance, environment monitoring, and ambient intelligence. In this context, designing an approach for target recognition using sensed audio data represents a very attractive solution that offers a wide range of deployment opportunities. However, this approach faces the limited resource's availability in the wireless sensor. The power consumption is considered to be the major concern for large data transmission and extensive processing. Thus, the design of successful audio based solution for target recognition should consider a trade-off between application efficiency and sensor capabilities. The main contribution of this paper is to design a low-power scheme for target detection and recognition based on acoustic signal. This scheme, using features extraction, is intended to locally detect a specific target and to notify a remote server with low energy consumption. This paper details the specification of the proposed scheme and explores its performances for low-power target recognition. The results showed the hypothesis' validity, and demonstrate that the proposed approach can produce classifications as accurate as 96.88% at a very low computational cost.

1 INTRODUCTION

Biodiversity, protection and conservation of ecosystems are a worldwide concern. Thus, developing an automated monitoring strategy for animal recognition and tracking is highly demanded. In this context, since animals are more often heard rather than seen, acoustic surveying offer a more efficient solution for monitoring animals. Compared to traditional methods, wireless sensors can provide a cost-effective solution to record acoustic data across larger spatiotemporal scales. The recorded acoustic signal provides a rich source of information for context awareness. However, unlike traditional scalar sensors, acoustic sensors generate large volumes of data at relatively high sampling rates, resulting in a large increase in energy consumption when the full-recorded signal is transmitted to a remote server. An extensive transmission of the entire raw audio signal might question the lifetime of the communication system.

One of the interesting ideas to reduce the volume of data is to locally detect and recognize a particular target using the extracted acoustic signature from the received audio signal. The remote server will be then informed using small size notification packets. This scheme can significantly contribute to decrease the number of required data transmissions, and consequently leads to overall improvement in the network lifetime. However, the success of such scheme in achieving the desired results depends on its ability to process the acoustic signal at low-cost in acoustic motes within limited resources and capabilities.

In this paper, we focus on the specification of a low-complexity acoustic sensing scheme for target recognition using two feature extraction methods: Zero-Crossing Rate (ZCR) and Root Mean Square (RMS). A Minimum Mean Distance (MMD) classifier is used for the recognition. Our main objective is to demonstrate the trade-off between power consumption and recognition accuracy.

2 RELATED WORKS

Recognition of animals by their sounds has motivated many research efforts. Han et al. (Han,

Algobail, A., Soudani, A. and Alahmadi, S.
Energy-aware Scheme for Animal Recognition in Wireless Acoustic Sensor Networks.
DOI: 10.5220/0006604100310038
In *Proceedings of the 7th International Conference on Sensor Networks (SENSORNETS 2018)*, pages 31-38
ISBN: 978-989-758-284-4

Muniandy and Dayou, 2011) extracted spectral centroid, Shannon entropy, and Rényi entropy features. They used k-Nearest Neighbour (k-NN) classifier to recognize nine frog species gaining just a 75.6% accuracy. Wavelet transform was combined with Mel-Cepstral Fourier Coefficients (MFCCs) in (Colonna et al., 2012) to classifiy nine species of anuran using k-NN classifier, whereas Xie (Xie et al., 2015) used perceptual wavelet packet decomposition sub-band cepstral coefficients for identifying ten frog species. All the previous works used wavelet analysis approach due to its ability to achieve a high accuracy rate compared to time and frequency domain features. However, these schemes are very demanding both in computational power and buffer size.

Among all mentioned features, MFCC feature set has been widely used because it provides more powerful discrimination capability. The authors in (Evangelista et al., 2014) computed a total of 64 MFCCs to recognize bird species using Support Vector Machine (SVM) classifier. In (Dong et al., 2015), MFCC and a spectral ridge were applied on 24 bird sound samples, achieving an accuracy of 71%. A combination of MFCC and LPC were adopted in (Yuan and Ramli, 2013) to classify eight frog species using k-NN with an identification rate of 98%. While in (Noda, Travieso and Sánchez-Rodríguez, 2016; Colonna et al., 2016), high recognition rates have been attained for the classification of anuran species using MFCC and other features such as energy. However, most of the previous works are not well suited for WASNs as they involve performing some kind of transformation and extracting a large set of features. In such studies, it is common to prioritize recognition accuracy even if it results in higher energy and storage costs.

Other studies developed an automatic recognition system based on syllable features. The idea behind such approach is to segment the stream of sound into syllable units and then derive syllable features. Colonna et al. (Colonna et al., 2015) applied ZCR and an energy entropy on 896 syllables of seven different anuran species. Alternatively, Xie et al. (Xie et al., 2016) introduced a method to recognize twenty-four frog species using a set of different frequency based features in addition to syllable, Linear Predictive Coding (LPC), and MFCC features and using a five different machine learning algorithms. In (Noda, Travieso and Sánchez-Rodríguez, 2016), a syllable feature used in conjunction with MFCC, and LFCC features for the classification of 199 anuran spices using SVM classifier. The results showed that the proposed set outperforms the approaches that used only MFCC, LFCC, or both of them. Also, in (Xie et al., 2015; Xie et al., 2016), the authors proved that using syllable features provide higher classification accuracy compared to MFCC under different levels of noise contamination.

Most of the presented schemes in reviewed studies didn't address the complexity and the power consumption to prove algorithms' efficiency. Such works have also been conducted in the field of species classification, but the recognition of different types of animals is not widely known in the literature. In addition, many current feature extraction systems are designed for particular applications, and hence are only appropriate for identifying a particular type of species. Thus, it becomes essential to develop an efficient recognition scheme to identify different animal sounds.

3 GENERAL APPROACH FOR ACOUSTIC-BASED LOW-POWER SENSING

We considered the implementation of an object recognition scheme in a WASN environment. The network is composed of a set of smart microphone nodes that are capable to sense and process audio streams and scalar data. At the set-up process, the end-user loads those sensors with the target audio-feature's descriptor. We assume that each sensor node will acquire a new acoustic signal in a periodic manner (over a time interval t). Then, the energy of the recorded signal will be computed and compared with a predefined threshold to generate a detection decision. Upon detection of the presence of a new object, the sensor node should extract a set of representative feature descriptors that can uniquely describe the object.

In the proposed scheme, instead of flooding the network with a large amount and perhaps useless transmitted data, a local classifier at each sensor node will first make a decision on the type of detected objects based on their extracted feature vector and the reference descriptor. When the target object is detected, a notification will be processed and sent to the base station according to the end-user requirements, which can be either a detection notification or vector of features.

The development of such a scheme that meets the constraints of WASNs is considered a big challenge. In fact, the design of this scheme requires

the use of low complexity processing methods for audio signal processing enabling target identification with minimum operating power and storage capacity. In our scheme, we focused on reducing the dimensionality of the feature vector to be represented by a minimum number of bytes, while maintaining a sufficient number of features to enable robust object recognition. This approach can reduce the amount of data to be exchanged, and as per consequence reduce the transmission energy consumption.

The success of such scheme in achieving the desired results depends mainly on two factors:

- The accuracy of the extracted features to abstract the target.
- The size of the dataset used to represent this feature.

Efficient features should be capable to discriminate between animals. However, the huge overlapping in the feature's vectors of different animals for the same feature makes almost impossible to accurately identify a specific animal. This problem is illustrated in Figure 1, where Root Mean Square (RMS) and Zero Crossing Rate (ZCR) values for different animals are showing a high overlapping intervals. In addition, the small set of training records is not sufficient enough to develop a model with high productivity. In specific, the learning algorithm doesn't have enough data to capture the underlying trends in the observed data in order to define the class boundaries properly.Thus, we proposed a multi-classification method in which one animal will be assigned to two class labels. The assignment will be based on finding the two classes that have the same value intervals for each feature.

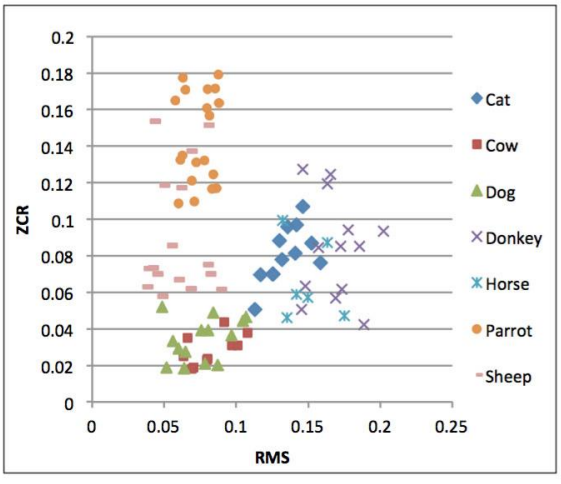

Figure 1: Two-dimensional feature space for animal dataset.

4 SPECIFICATION OF THE SENSING SCHEME FOR TARGET RECOGNITION

The main contribution of this paper is to design a low-complexity scheme for object recognition in order to be used in the context of WASN. However, achieving this goal is directly dependent on the appropriateness of the feature extraction and classification algorithms. Two main factors should be taken into consideration in this task, which are:

- Network resource constraints: wireless sensor networks are generally characterized by their limited resources, such as small memory size, limited power supplies, and low communication bandwidth. Therefore, it is important to keep the feature vector that represents the object descriptor as short as possible. In addition, the classification algorithm should rely on a model that is capable of making recognition decisions with a minimum number of data set.
- Computational complexity: The power consumption level in the sensor node is directly affected by the complexity of the feature extraction and classification method. This comes from the fact that a large number of mathematical operations need a large number of clock cycles to be executed, and hence increase the demands in energy consumption. Therefore, it is essential to adopt a low-complexity method with as less number of arithmetic operations as possible.

The proposed approach is composed of two sequentially phases Figure 2: Reference vectors extraction and object recognition. In the first phase, different feature vectors will be extracted from each training records and then the mean for each feature and each object will be computed. Based on these, a unique descriptor (one vector per animal) will be constructed and stored. These descriptors are loaded into the sensor memory during the set-up process. Afterward, in the second phase, these descriptors will be compared with the detected object feature vector for the classification purpose. In general, each stage is composed of multiple steps, as detailed in the following sections.

4.1 Object Detection

The detection of a new object is based on a certain received signal power threshold level. The purpose of detection is to differentiate the acoustic events from the background noise, without determining

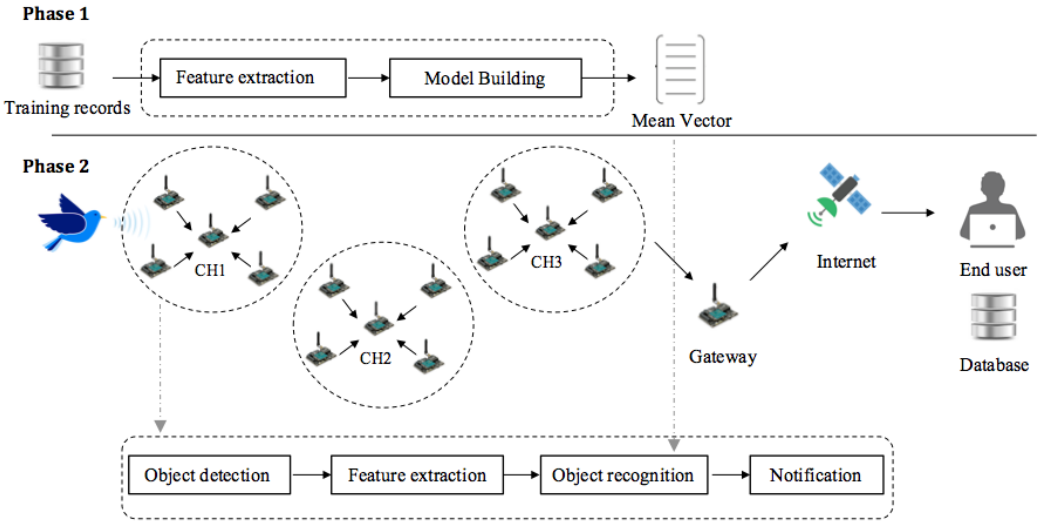

Figure 2: Flowchart of the overall object recognition scheme.

whether the target object is present or not. The sensor will acquire a new acoustic signal in a periodic manner (over a time interval T). Then, the received signal energy (E) during that time interval will be measured using the root mean square feature. The energy should be checked, if it is greater than a predetermined threshold value (T*thre*), then a new object is supposed to be detected. The detection function D can be defined as follows (1)

$$D = \begin{cases} 1 & E > Tthre \\ 0 & E \leq Tthre \end{cases} \quad (1)$$

4.2 Feature's Extraction

The feature extraction can play an important role in saving sensor resources in terms of energy, memory requirements and processing time. In addition, it is an essential step in any audio based-recognition process or classification task. In fact, the capability of the sensor to discriminate the target object from different auditory objects is a challenging task, especially when it involves objects with relatively similar sound waves. Therefore, these characteristic features offer precise information that can maximize discrimination between different objects, and hence increase the recognition accuracy. The proposed feature extraction process will be based on the following sequential steps:

I. Divide the captured audio stream into (T) frames; each frame has 1024 sequential samples in which 50% of them are overlapped between two successive frames.
II. Calculate the characteristic's features for each frame. In our approach, we have focused on two temporal feature extraction methods: RMS and ZCR (Lerch, 2012). RMS is an important measurement that provides information about the amplitude of a signal, while ZCR demonstrates the characteristics of dominant frequency.
III. Obtain the global feature vector by computing the mean for each features' dimension d = 1,2,..,D of N-dimensional features vector F that is obtained from all the frames.

4.2.1 Root Mean Square Feature

It is a time domain feature denotes the average signal strength, which can be expressed as

$$RMS = \sqrt{\frac{1}{N}\sum_{i=1}^{N} x_i^2} \quad (2)$$

where x_i is the *i*th sample value of the frame, and N denoted as the frame length (Lerch, 2012).

4.2.2 Zero Crossing Rate Feature

This is a time domain feature that measures the number of times the signal changes its sign within a time window, from positive to negative and vice versa. It is one of the most commonly used methods for computing the frequency content of the signal (Lerch, 2012), which can be calculated according to the following equations (3)

$$ZCR = \frac{1}{2(N-1)}\sum_{m=1}^{N-1} |sgn[x(m+1)] - sgn[x(m)]| \quad (3)$$

where x(m) is the value of the m^{th} sampled signal and sgn[] is the sign function, which defined as:

$$sgn[x(n)] = \begin{cases} 1 & x(n) \geq 0 \\ -1 & x(n) < 0 \end{cases} \quad (4)$$

4.3 Classification

There are several classification techniques that can be used for acoustic target recognition. These techniques can be grouped into two general categories: Supervised (e.g. Maximum Likelihood, MMD, k-NN) and machine learning (e.g. artificial neural network, classification tree, SVM) algorithms. The machine learning classification methods have been widely and successfully used in many studies in the literature to recognize a single target. However, most these classification methods cannot be directly applied in the area of WASNs because they significantly need large resources and computation power than available on a low-power sensor node. To overcome these limitations, the developed algorithm must be computationally light and its classification model should consume limited storage space.

In our approach, the classification process will be based on MMD classifier (Rudrapatna and Sowmya, 2006). The main idea of this approach is to check the objects' similarity by finding the class mean, which has the minimum distance from the detected object feature vector. In spite of its simplicity, some animal acoustic classification approaches have adopted MMD classifier because it can provide an acceptable degree of accuracy in most situations (Huang et al., 2009; Luque et al., 2016). However, since the overlapping between animals is large in our case as previously explained (figure 1), we used two-classes for the classification approaches in the proposed scheme, in which an object will yield an estimated probability of belonging to two classes. The classification process is composed of two phases:

I. Classification (learning) model construction: The mean of the feature vectors of training records will be calculated to construct the descriptor vector for each specific object. Each descriptor vector Refi = $\{\mu_{RMS}, \mu_{ZCR}\}$, contains two values, which are the mean of (RMS and ZCR) features. This vector will be stored in the database and will be loaded into the sensor memory to be used later in the similarity matching stage.
II. Similarity matching: The similarity between the newly extracted unknown object feature vector and the reference vectors Ref_i will be measured and stored in a distance vector $D_i = \{d_i,\ldots d_N\}$ using the Euclidean distance metric (Xie et al., 2016); The Euclidean distance d between two vectors can be calculated as follows:

$$d = \sum_{i=1}^{D} |f_i - f_i'|^2 \quad i = 1,..,D \quad (5)$$

where f_i is the detected object vector, and f_i' is the mean vector of classes. Then, we will find the two minimum values in the distance vector D_i, which represent the classes that the test object belongs to.

4.4 Notification to the End User

When the target object is detected in a certain cluster, a notification will be processed by each node and will be sent to the cluster head. The notification will be processed at the node level according to the end user requirements, which can be either: few bits' notification or vector of features. The cluster head will find the packet with the highest RMS value and pass it to the end user for further classification. Then, at the end user, the feature vector will be used to assign the target object to a single animal class.

5 IMPLEMENTATION AND PERFORMANCES ANALYSIS

5.1 Performance Analysis at the Application Level

The capabilities of the proposed scheme to classify a specific target object to two classes are tested and evaluated using MATLAB tool.

5.1.1 Pre-processing

Collected audio records have to be pre-processed before being used. The pre-processing involves data cleaning, audio segmentation, and audio re-sampling. Data cleaning process aims to remove the records that have a bad quality sound or contains high noises. Then, the records that contain mixed animal sounds should be segmented to produce a record that has only single animal call. Each record is made up of either one or multiple continuous calls of one animal, with the duration ranging from one to two seconds. Finally, the audio records need to be re-sampled into a unified sampling frequency in order to not affect the recognition accuracy. After the pre-processing, the dataset will be divided into two sets: training set and test set. Almost 70% of the total collected samples are used as training data to build the classification model, while the rest are used

for testing.

5.1.2 Dataset

The dataset for object recognition was built from audio collected from Animals & Birds Sound Effects CD (Sound-ideas.com, 2017). The rest of the records were collected from different animal sound library websites. The audio recordings are stored in WAV format with a sample rate of 44.1 kHz and 16 bits per sample. The dataset contains seven animals with 114 samples: Cat, Cow, Dog, Donkey, Horse, Parrot, and Sheep. A typical example of two different dog bark sounds is depicted in Figure 3.

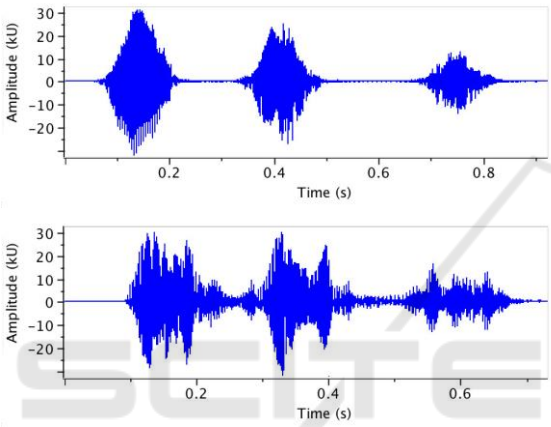

Figure 3: Audio sample (wave form) for dog bark.

5.1.3 Performance Evaluation

The main metric used to examine the performance of the proposed scheme is the successful recognition rate. The ratio of success represents how many objects that the algorithm was able to classify correctly, which can be calculated as follows

$$Accuracy = \frac{N_C}{N_S} x\ 100 \qquad (6)$$

where N_c is the number of records which were recognized correctly and N_s is the total number of test samples.

5.1.4 Recognition Ratio

We studied the capability of two features (RMS and ZCR) to discriminate between different animals using MMD and different machine learning classifiers, namely, Decision Tree (DT), Gaussian Mixtures Model (GMM), and K-NN (Table 1). The machine learning classifiers were tested to decide on which classifier best suits the single animal classification at the end user. The results show that the proposed scheme was capable to correctly classify 31 out of 32 targets (accuracy of 96.88%). We also note from Table 1 that the GNN classifier was able to identify the target animal with better accuracy than the DT and K-NN classifiers, gaining 71.88% recognition accuracy. Nevertheless, the three classifiers performed poorly for Horse and Cow. Compared to other animals, the number of training set of these two animals is not enough for classifiers to enable complete learning, which caused under-fitting problem.

Table 1: The comparison of different classification algorithm results.

Animal	MMD	DT	GMM	KNN
Cat	100%	75%	75%	75%
Cow	100%	25%	25%	25%
Dog	100%	40%	60%	40%
Donkey	100%	80%	100%	100%
Horse	50%	0%	0%	0%
Parrot	100%	100%	100%	100%
Sheep	100%	66.67%	83.33%	66.67%
Average	96.88%	62.50%	71.88%	65.63%

5.2 Energy Efficiency of the Proposed Scheme for Sensor Node

We investigated the energy consumption of our scheme using AVRORA (Titzer, K. Lee and Palsberg, 2005). AVRORA is an instruction-level sensor network simulator that emulates the Mica family platforms (MICA2dot, MICA2, MICAz). These motes are based on a low-power ATmega128L micro-controller with 4KB of RAM. The main purpose of such framework is to execute TinyOS applications prior their deployment in actual WSN nodes. The AVRORA helps to estimate the number of clock cycles, power consumption, and processing time. We focused mainly in measuring the performance of a sensor mote during recognition and notification task. Table 2 summarizes the cost of the recognition task per sample, assuming a sampling frequency of 44.1 kHz. As shown in Table 2, feature extraction is the most energy requiring and time consuming steps of object recognition scheme, but nevertheless it is very essential to reduce the communication overhead and prolong the network lifetime. Such task involves performing several expensive computations on large sets of data compared to other steps.

Table 2: Evaluation of the recognition cost on MICA2.

Sound size	Per 1 second (44100 samples)		
Measured Attribute	Clock cycles	Time (ms)	Energy (mj)
New object detection	352	0.044	0.0009
Feature extraction	1527	0.19	0.004
Classification	12025	1.5	0.03
Whole scheme	13904	1.734	0.0349

In Table 3, we considered measuring the notification cost for three different scenarios, in which a sensor node can either send a packet that includes a detection notification, object feature vector, or unprocessed acoustic sample. Here we assumed using a 16-bit ADC acoustic sensor to collect audio signals at a rate of 44100 samples per second. The sensor will send data to the server side at a rate of 16 bits/sample * 44100 samples/second = 705600 bits/second. Thus, the power consumed during the radio transmission for a total of 705,600 bits per second / 8 = 88,200 bytes per second is 29106 mj. These latter results prove that the proposed scheme can dramatically reduce the total energy consumption in the network.

Table 3: Evaluation of the notification cost on MICA2.

Measured Attribute	Time (s)	Energy (mj)
Transmit detection notification (1 byte)	0.01	0.33
Transmit 2D feature vector (2 bytes per feature)	0.04	1.32
Transmit raw signal (2 bytes per sample)	0.02	0.66

6 CONCLUSIONS

We presented a low-complexity scheme for target recognition in an energy-constrained acoustic sensor network and applied it to audio-based animal's classification scenario. The proposed scheme uses low-complexity feature extraction techniques, in which the goal is to minimize the heavy processing burden and transmission overhead on the network. This was achieved by the detection of event of interest locally using RMS and ZCR features, and then reporting the event to the server with small-size packets. Results indicate that the adopted feature extraction methods were able to generate a unique signature, which was successfully used to discriminate between different auditory objects in the recognition process. Moreover, the performed simulations have demonstrated that the proposed scheme can save up to 70% of power consumption, while guaranteeing high recognition accuracy. As a future work, we are interested to study a possible real implementation of the proposed algorithm on MICA2 mote.

REFERENCES

Colonna, J. G., Ribas, A. D., Santos, E. M. d., and Nakamura, E. F. (2012). Feature sub-set selection for automatically classifying anuran calls using sensor networks. In: *International Joint Conference on Neural Networks (IJCNN)*. pp.1-8. IEEE.

Colonna, J., Peet, T., Ferreira, C., Jorge, A., Gomes, E. and Gama, J. (2016). Automatic Classification of Anuran Sounds Using Convolutional Neural Networks. In: *C3S2E '16 Proceedings of the Ninth International Conference on Computer Science & Software Engineering*. ACM.

Colonna, J., Cristo, M., Salvatierra, M. and Nakamura, E. (2015). An incremental technique for real-time bioacoustic signal segmentation. *Expert Systems with Applications*, 42(21), pp.7367-7374.

Dong, X., Xie, J., Towsey, M., Zhang, J. and Roe, P. (2015). Generalised Features for Bird Vocalisation Retrieval in Acoustic Recordings. In: *17th International Workshop on Multimedia Signal Processing (MMSP)*. IEEE.

Evangelista, T., Priolli, T., Silla Jr., C., Angelico, B. and Kaestner, C. (2014). Automatic Segmentation of Audio Signals for Bird Species Identification. In: *International Symposium on Multimedia*. IEEE.

Han, N., Muniandy, S. and Dayou, J. (2011). Acoustic classification of Australian anurans based on hybrid spectral-entropy approach. *Applied Acoustics*, 72(9), pp.639-645.

Huang, C., Yang, Y., Yang, D. and Chen, Y. (2009). Frog classification using machine learning techniques. *Expert Systems with Applications*, 36(2), pp.3737-3743.

Lerch, A. (2012). An Introduction to Audio Content Analysis: *Applications in Signal*. 1st ed. Somerset: John Wiley & Sons.

Luque, J., Larios, D., Personal, E., Barbancho, J. and León, C. (2016). Evaluation of MPEG-7-Based Audio Descriptors for Animal Voice Recognition over Wireless Acoustic Sensor Networks. *Sensors*, 16(5).

Noda, J., Travieso, C. and Sánchez-Rodríguez, D. (2016). Methodology for automatic bioacoustic classification of anurans based on feature fusion. *Expert Systems with Applications*, 50, pp.100-106.

Rudrapatna M., Sowmya A. (2006) Feature Weighted Minimum Distance Classifier with Multi-class

Confidence Estimation. In: *Advances in Artificial Intelligence*. Springer.

Sound-ideas.com. (2017). *HD – Animals & Birds Sound Effects*. [online] Available at: https://www.sound-ideas.com/Product/380/HD-–-Animals-Birds-Sound-Effects [Accessed 15 Jul. 2017].

Titzer, B., K. Lee, D. and Palsberg, J. (2005). Avrora: scalable sensor network simulation with precise timing. In: *Fourth International Symposium on Information Processing in Sensor Networks (IPSN)*. IEEE.

Xie, J., Towsey, M., Eichinski, P., Zhang, J. and Roe, P. (2015). Acoustic Feature Extraction using Perceptual Wavelet Packet Decomposition for Frog Call Classification. In: *11th International Conference on eScience*. IEEE.

Xie, J., Towsey, M., Truskinger, A., Eichinski, P., Zhang, J. and Roe, P. (2015). Acoustic classification of Australian anurans using syllable features. In: *Tenth International Conference on Intelligent Sensors, Sensor Networks and Information Processing (ISSNIP)*. IEEE, pp.7-9.

Xie, J., Towsey, M., Zhang, J. and Roe, P. (2016). Acoustic classification of Australian frogs based on enhanced features and machine learning algorithms. *Applied Acoustics*, 113, pp.193-201.

Xie J., Towsey M., Zhang L., Zhang J., Roe P. (2016). Feature Extraction Based on Bandpass Filtering for Frog Call Classification. In: *Image and Signal Processing (ICISP)*. Springer.

Yuan, C. L. T., Ramli D. A. (2013) Frog Sound Identification System for Frog Species Recognition. In: Context-Aware Systems and Applications (ICCASA). Springer.