ENERGY EFFICIENT ANIMAL SOUND RECOGNITION SCHEME IN WIRELESS ACOUSTIC SENSORS NETWORKS

Saad Al-Ahmadi and Badour AlMulhem

Department of Computer Science, King Saud University, Riyadh, Saudi Arabia

ABSTRACT

Wireless sensor network (WSN) has proliferated rapidly as a cost-effective solution for data aggregation and measurements under challenging environments. Sensors in WSNs are cheap, powerful, and consume limited energy. The energy consumption is considered to be the dominant concern because it has a direct and significant influence on the application's lifetime. Recently, the availability of small and inexpensive components such as microphones has promoted the development of wireless acoustic sensor networks (WASNs). Examples of WASN applications are hearing aids, acoustic monitoring, and ambient intelligence. Monitoring animals, especially those that are becoming endangered, can assist with biology researchers' preservation efforts. In this work, we first focus on exploring the existing methods used to monitor the animal by recognizing their sounds. Then we propose a new energy-efficient approach for identifying animal sounds based on the frequency features extracted from acoustic sensed data. This approach represents a suitable solution that can be implemented and used in various applications. However, the proposed system considers the balance between application efficiency and the sensor's energy capabilities. The energy savings will be achieved through processing the recognition tasks in each sensor, and the recognition results will be sent to the base station.

KEYWORDS

Wireless Acoustic Sensor Network, Animal sound recognition, frequency features extraction, energy-efficient recognition schema in WASN.

1. INTRODUCTION

WSNs have rapidly grown as a cost-effective solution for data collection and measurements. A significant advantage of WSN is the ease of deployment in challenging environments; because of the use of routing protocols that self-configures the network. But these sensors are powered by small batteries, which are typically limited power capabilities [1].

WASNs include a number of standard sensors nodes with integrated microphones. These microphones add more capabilities to the sensors by make sensors capable of acquiring and processing the audio signals from a region of interest [2], [3].

WASN can efficiently monitor the environment by a process the animal audio signals using sound recognition techniques. In general, the sound recognition process consists of first, recording the waveform using a microphone, second, pre-processing the audio signal, third, extracting features from the audio, fourth, classifying the extracted features, to determine the species/animal whose audio belongs to (Figure 1) [4], [5].

DOI:10.5121/ijwmn.2020.12402

The paper aims to propose a new energy-efficient approach for recognizing animal sound based on the frequency features extracted from acoustic sensed data. The aim had achieved by developing a system with an acceptable recognition accuracy rate in conjunction with a low-power scheme in wireless acoustic sensor nodes.

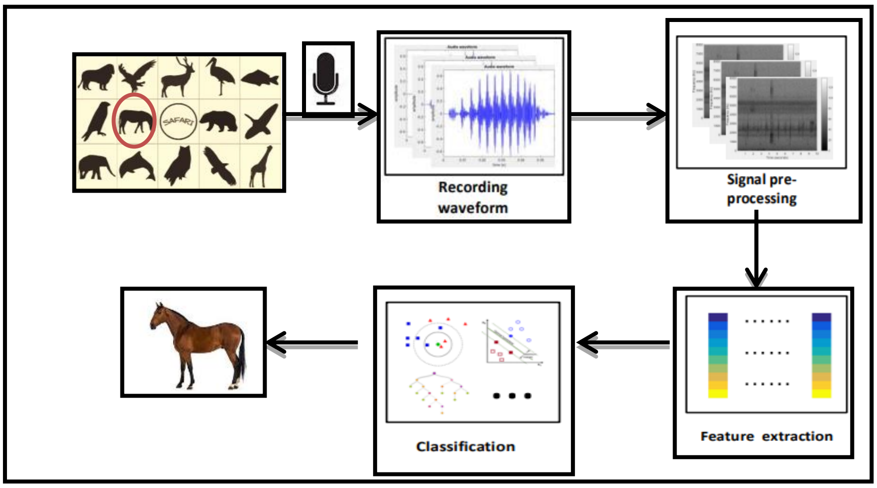

Figure 1. Animal sound recognition system: collecting data, pre-processing, feature extraction, and classification

1.1. Problem Statement

Although many acoustic monitoring applications had developed, the development of efficient acoustic monitoring, that deals with wireless acoustic sensor networks restrictions on computing and transmission power is still an open research issue. Until now, Most of the available researches depended on prerecorded animal sounds. A limited number of studies have been done in the field of animal sound recognition systems, using WASN, which is very helpful for bioacoustics and any audio analysis applications [3], [6].

Wireless sensors nodes are limited energy supply because they are not connected to any wired energy sources. They are powered by a battery that had a limited lifetime (several months). The data transmission requires much higher energy consummation compared to data processing. So, the amount of transmission data needs to be reduced; by such as avoid transmitting a large stream of data to the base station before processing these data locally in each sensor by using data compression, feature extraction, etc. [7].

This research worked on a significant goal to design an energy-efficient animal sound recognition scheme to be used in the context of WASNs. We explored and integrated knowledge from multiple disciplines, particularly on sound recognition, machine learning, and WASN, this knowledge had been used to optimize energy consumption to extend the network lifetime while guaranteed the acceptable rate of recognition accuracy [3], [6].

The scheme possessed the limited energy supply challenge that this work had address it by proposing a solution that significantly decreases the data stream required for transmission. The energy-saving will be achieved in the proposed solution through processed the recognition tasks in each sensor, in which only the recognition results will be sent to the sink.

1.2 Motivation

Every animal within the biological community has a vital role on the planet. If one species of animal is extinct due to some imbalance, it can have significant cascading effects on the ecological balance, and cause critical danger to the global biodiversity. Monitoring animals, especially those that are becoming endangered, can assist biology researchers' efforts and improve our understanding of their interactions and, therefore, their impact on the ecosystem [4], [8].

Developing surveillance techniques is becoming more important; to gain insights about animals. Acoustic monitoring in WASN is one of the best detection mechanisms that gave good results in the detection of the Red Palm Weevil threat early. Red Palm Weevil is one of the most dangerous threats; that cause critical losses to the palm growers in Saudi Arabia [9].

A few types of research in acoustic monitoring focused on finding an adequate solution for WASNs constraints, such as limited power supply [4], [8]. Our work had been focused on proposing a new energy-efficient schema for recognizing animal sound based on the features extracted from the frequency domain. This approach represents a suitable solution that can be applied and used in a wide range of different applications. However, the proposed method considers the balance between application efficiency and sensor energy capabilities.

2. LITERATURE REVIEW

Animal sound recognition is precious for biological research and environmental monitoring applications. Furthermore, most of the animal vocalizations have evolved to be species-specific. Therefore, using animal sounds to recognize the animal species automatically is an adequate method for ecological observation, environment monitoring, biodiversity assessment, etc. [7]. WASN with the machine learning algorithms have been provided a low-cost solution for long-term monitoring, at difficult and vast areas, rather than traditional surveillance, which usually involves ecologists to visit locations to gather the environmental data physically, which is a time and cost consuming process. Over the past few years, a few studies have been focused on recognizing different types of animal sounds and finding an adequate solution by using WASN.

2.1. Signal pre-processing

For the animal sound classification, signal pre-processing is the first process after collected the acoustic data. Signal pre-processing includes signal processing, noise reduction, and syllable segmentation [4], [8].

2.1.1. Signal Processing

Signal processing in the animal sound classification is the transformation of the sounds signals from the waveform representation (time-domain) to spectrogram representation (time-frequency domain) using such as Short-Time Fourier Transform (STFT) and Wavelet Transform (WT) techniques [10][11]–[13], or to spectrum representation (frequency domain) using Fast Fourier transform (FFT) technique (Figure 2) [14], [15].

Fast Fourier Transform had been used in this research to extract the frequency/spectral values in the signal source. In practice, they usually use a specific type of Fourier transform, which is DFT. But since WASN has an only limited lifetime, it had been inefficient to implement DFT directly on the sensor nodes. Instead, we applied FFT; an efficient algorithm used to calculate and reduces the complexity of the DFT computations for N samples from O (N2) to O (NlogN). FFT can be expressed as below equation Where $f(k)$ is the frequency domain signal (spectrum) [3].

$$f(k)=\int_{-\infty}^{\infty} f(t).e^{-2\pi ikt}.\mathrm{d}\tau \qquad (1)$$

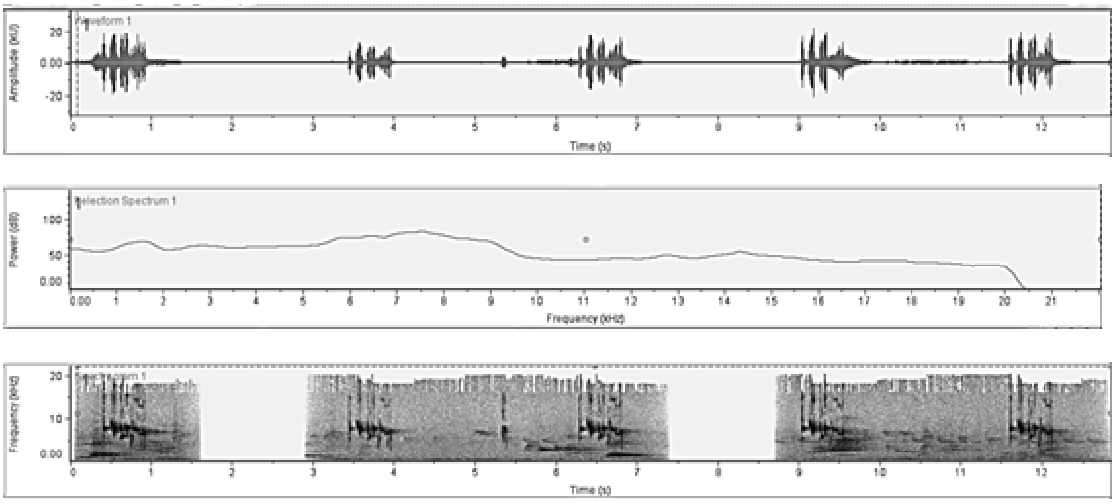

Figure 2. Waveform, spectrum, and spectrogram for bird sound sample

2.1.2. Noise Reduction

Noise reduction is an optional step for the animal sound classification, used for extracting the pure signal [4], [10]. Huang et al. [13] and [16] have been applied a de-noise filter, which is a well-known technique to remove the noise from the recorded sound. Bedoya et al. [17] introduced the spectral noise gating methodology, which used for estimation and suppression of the noisy components in the selected spectrum of the signal.

2.1.3. Syllable segmentation

Continuous call, emitted by any animal, is composed of some syllables repeated along a specific period. A syllable is a basic unit for classification. In particular, a set of features are all obtained from syllables, so the accuracy of syllable segmentation directly affects the classification accuracy [4], [18]. Huang et al. presented [19] iterative time-domain algorithm that used a time-domain feature amplitude for the segmentation of the audio signal. Also, authors in [13], [16], and [11] used an adaptive-end-point segmentation algorithm to extract syllables by measuring the positions of two endpoints of the waveforms. The peak amplitude will be in the middle of each syllable. Colonna et al. [18] proposed a new incremental segmentation approach that computes the Energy and the Zero Crossing Rate incrementally in the waveforms to extract syllables.

2.2. Extracted Features

Selecting proper acoustic features that show more considerable differences between species is the key to achieving effective classification performance. Several features can be used to describe the signal. For the animal's sound classification, acoustic features can be classified based on their domain of computation into six categories: time features, energy features, frequency (spectral) features, cepstral features, Biologically/Perceptually driven and time-frequency features [4], [5], [20]. Acoustic features and their categories are listed in Table 1.

2.2.1. Frequency (spectral) features

Spectral or Frequency features are extracted from the signal in the frequency domain. Frequency domain in the audio refers to the analysis of signals concerning the frequency, rather than time. The original signal can be converted from the time domain (waveform) to the frequency domain (spectrum) by transforming functions such as FFT, STFT, and WT. Frequency features include but limited to, spectral centroid, spectral flatness, spectral rolloff, signal bandwidth, spectral flux, and fundamental frequency. These features have a high computational cost as they are obtained by applying transform functions to the original signal [5], [19], [20]. To achieve a better sound classification performance, the frequency features are usually combined with other features such as time features, energy features, etc.

Time-based features performance is strongly influenced by environmental noises such as blowing wind [20]. To overcome this issue, many works have been proposed a combination of time and frequency features, because frequency-based features provide a robust solution to environmental noises. The spectral centroid is considered as one of the essential frequency features that have been widely used in many frog identification systems [11], [13], [16], [19], [21], [22].

Huang et al. [13], [16] [19] have been proposed different methods on automatic online frog sound identification system that provides the public with easily online consultation. Huang et al. in [16] introduced an intelligent frog call identification agent, by proposing a method with pre-noise and pre-emphasis filtering techniques that applied to the sound samples. An adaptive endpoint detection segmentation algorithm is used to detect the sound syllables. Spectral centroid, signal bandwidth, spectral roll-off, and threshold- crossing rate features are extracted and entered as input parameters for the three well-known classifiers kNN, SVM, and GMM classifiers. Experimental results demonstrated that the GMM classifier achieves up to the best accuracy.

Another published paper by Huang et al. [19] extended the investigation on the automatic online frog sound identification system. Their proposed methods depend first on segmented the frog sound into syllables. Second, they extracted three features, which are spectral centroid, signal bandwidth, and threshold crossing rate, and used them as parameters for the frog sound classification process. Third, they used kNN and SVM classifiers. However, the experimental results showed that the proposed methods in both [19] and [16] are appropriate for the identification of the same frog species dataset, which includes five different classes of frogs, but some species such as Microhyla butleri and Microhyla ornate are needing more analysis.

In a subsequent paper by Huang et al. [13], the authors proposed a system that segmented the sound samples using an adaptive endpoint detection segmentation algorithm that has been introduced in [16]. Then they classify the nine species of frogs by using neural networks classifiers. A combination of the spectral centroid, signal bandwidth, spectral rolls-off, threshold-crossing rate, spectral flatness, and average energy features are used as input for the classification process. Results exhibited that the identification accuracy rate of the proposed system can achieve up to 93.4%. However, The NN classifier required a very high computational cost, and the nature of the frog sound dataset used in [13], [16], [19] is unclear since the datasets were proprietary.

Dayou et al. [21] developed on automatic frog sounds identification. They aimed to study the differences between frogs' sounds, by proposing two features Shannon and Rényi entropies. In their system, they first segmented the frog sound signal by using the Raven Lite tool into syllables. Then second, they used spectral centroid, Shannon entropy, and Rényi entropy as features for frog species classification in the kNN classifier. They demonstrated a total classification rate of 98.0% for the nine species of Microhylidae species frogs, on the case of using spectral centroid with the entropy features. However, the accuracy of the frog species classification is reduced significantly when the noise levels are higher than 20 dB.

Xie et al. [11], [22] have been proposed different methods on the frog sound recognition system . Xie et al. [22] proposed a multi-level classification for family, genus, and species of frogs calls. In their study, the calls were segmented with the Raven tool, which has been used in [21]. For the features extraction step, ten features were extracted from each syllable. They are including spectral centroid, spectral flatness, spectral roll-off, ZCR, Shannon entropy, spread, skewness, kurtosis, root mean square value, and averaged energy. After feature extraction, a DT classifier is used to evaluate the most critical features for classifying the family, genus, and species of frogs. Finally, the SVM classifier is used for the multi-level classification with the three most essential features. The average accuracy for selected acoustic features was up to 85.68%, 75.58%, and 64.07% for family, genus, and species, respectively. However, the accuracy of classification is the lowest among other studies.

Xie et al., in another work [11], introduced a recognition method based on an enhanced feature representation for frog call classification, which includes a fusion of time, energy, spectral, perceptually driven, and cepstral features, as an extension of their previous work [23]. In their system, they segmented the sound using an adaptive endpoint detection algorithm that has been proposed in [16]. Then a combination of syllable duration, Shannon entropy, Rényi entropy, ZCR, averaged energy, oscillation rate, spectral centroid, spectral flatness, spectral roll-off, signal bandwidth, spectral flux, fundamental frequency, MFCCs, and LPC features have been extracted and used in the classification process. LDA, kNN, SVM, Random forest, and ANN algorithms are used to classify the sounds. The method showed the best average accuracy achieved until now, 99.1 %, and had good anti-noise ability. In contrast, the technique consumed high computational resources due to the number of features that have been used. However, The dataset used in [11], [22] had the same source and quality.

Camacho, Rodriguez, and Bolanos [24] used the loudness, timbre, and pitch to detect frogs calls with a multivariate ANOVA classifier. First, they calculated the loudness of the signal, to extract the sections of the frog vocalizations from the background noise. Second, they used timbre and pitch to recognize the calls and got a pure tone. The last step consists of performed a multivariate ANOVA on pitch and timbre to identify the calls. The performance of the method was evaluated based on the precision rate and recall rate measures. The precision rate was 99%, and the recall rate of 92 %. In general, almost all the prior works adopt achieved a good identification accuracy expect [25]. And all these works have dealt with the noisy audio records [11], [13], [16], [19], [24].

Gunasekaran and Revathy [25] have been developed a classification method tailored explicitly to animal vocalizations identification. In this paper, they proposed a technique to define and extract a set of acoustic features from a wild animal. Their work is similar to our proposed technique in the type of dataset, which consists of different classes of animals. The authors first used a fractal dimension segmentation method. Second, a group of temporal, spectral, perceptual, harmonic, and statistical features used to characterize the audio signals. Third, they used different feature selection methods to minimize the processing power required for the classifier and to improve the classification efficiency as well. Fourth, a fusion of two different classifiers was used: kNN and Multi-Layer Perceptron classifiers. However, the proposed method was able to achieve an overall accuracy classification rate of 70.3%.

Table 1. Features extraction categories

Class	Features	References
Frequency	Spectral centroid Signal bandwidth Spectral roll-off Spectral flatness Fundamental frequency Spectral flux Spectral spread Spectral skewness Spectral kurtosis Spread Variation Crest Pitch Timbre	[11], [13], [16], [19], [21], [22], [24], [25]
Time	Threshold- Crossing Rate ZCR Syllable duration Root Mean Square (RMS) Log-attack time Loudness	[11], [13], [15], [16], [19], [22], [24]–[26]
Energy	Average energy Shannon entropy Rényi entropy Signal power Temporal centroid	[11], [13], [15], [21], [22], [25]
Cepstrum	Mel Frequency Cepstral Coefficients (MFCC)	[11], [14], [23], [25]
Biologically/Perceptually driven	Linear Prediction Coefficients (LPC) Perceptual spectral centroid Odd-to-Even ratio Tristimulus Slope	[11], [25]
Frequency –Time	Spectral peak tracks Wavelet coefficients Oscillation rate Dominant frequency The frequency difference between the lowest and dominant frequencies. The frequency difference between the highest and dominant frequencies. Time from the start to the peak volume of the sound. Time from the peak volume to the end of the sound.	[14], [15], [23], [26], [27]

2.3. Classification

The classification process is the last step in any identification system. Some pattern recognition methods have been used to create the classifier, which used to determine each species/animal whose syllables belong to, depending on the extracted features. Examples of these classifiers k-Nearest Neighbors (kNN) [11], [14], [16], [19], [21], [23], [25], [26], [28] .The kNN classifier depends on the nearest neighbor rule, and it's considered as the most commonly used classifier for its simplicity and easy implementation [4], [5], [28]. SVMs [11], [16], [19], [22] . GMM [16]. Neural networks (NN) [13]. Convolutional Neural Network [6]. One-way multivariate ANOVA [24]. And linear discriminant analysis (LDA) [11]. Besides classifiers, other methods for classifying the animal species include those based on similarity measures such as Euclidean distance [27], [29].

2.4. Dataset

One major problem for animal sound identification is the lack of a global data set. The datasets used in prior work often be proprietary or related to specific geographical regions, because researchers from different countries focus on particular animal species in their particular area. Therefore, it is difficult to compare different animal sound classification methods. More data is needed to analyze the performance and to improve the quality of classification frameworks.

3. SYSTEM DESIGN AND METHODOLOGY

Based on the literature review, the selected method has been found to work well for audio processing. The proposed approach starts with the sound acquisition process, in which the system collects and selects a sound for processing using acoustic sensors. Then, the audio pre-processing techniques are applied to simplify and improve sound quality. In the pre-processing process, FFT had been implemented, to convert the recorded audio from the waveform to spectrum representation. Frequency features had been extracted from the spectrum representation. After that, the features of each sound will be passed on to the classifier to build the classification model.

Figure 3 shows the basic operations that are part of the audio recognition system with a WSN, Note that the activities of object detection, feature extraction, and classification are performed on the sensor nodes. The recognition result with a few bits will be transmitted to the base station.

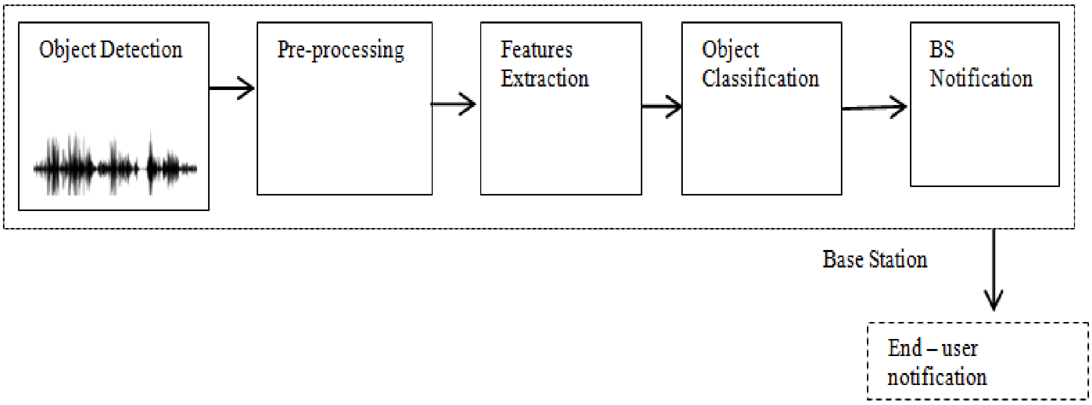

Figure 3. General scheme for the object recognition process

3.1. Object detection

Object detection is based on detecting and receiving the audio signal. The purpose of detection is to determine whether the target object is present or not and to discriminate the object signal from the environmental noise. The sensor will acquire a new acoustic signal. Then, the received signal energy will be measured using the root mean square (RMS) feature. The RMS should be measured; if it is higher than the value of th threshold, then a new object is detected, and the acoustic signal will be moved to the next step for signal pre-processing and feature extraction. If the RMS is less than the threshold value, then the detected signal will be ignored. The detection function can be defined as follows:

$$\text{Detection function} = \begin{cases} 1 & \text{RMS} > \text{threshold} \\ 0 & \text{RMS} \leq \text{threshold} \end{cases} \qquad (2)$$

3.2. Pre-processing

3.2.1 Signal Segmentation

Acoustic signals are generally pre-processed before features are extracted to enhance the feature extraction process and improve the classification accuracy. The significance of segmentation is to select the most representative parts of the signal because these signals usually contain long periods of silence, that consume large amounts of memory. The segmentation step is considered a foremost pre-processing step in any acoustic signal processing system [13], [18].

Once the signal has been appropriately segmented into some frames, a set of features can be assembled to represent each frame. In this work, we divided the signal into frames using the Hamming window. Silent areas in the frame have been ignored, and the nonsilent regions have been processed to extract features (Figure 4). Frames overlap by 50 %, to prevent any loss of information at the end of the frames [30]. The acoustic signal is partitioned into equal-sized frames; each frame has 1024 sequential samples.

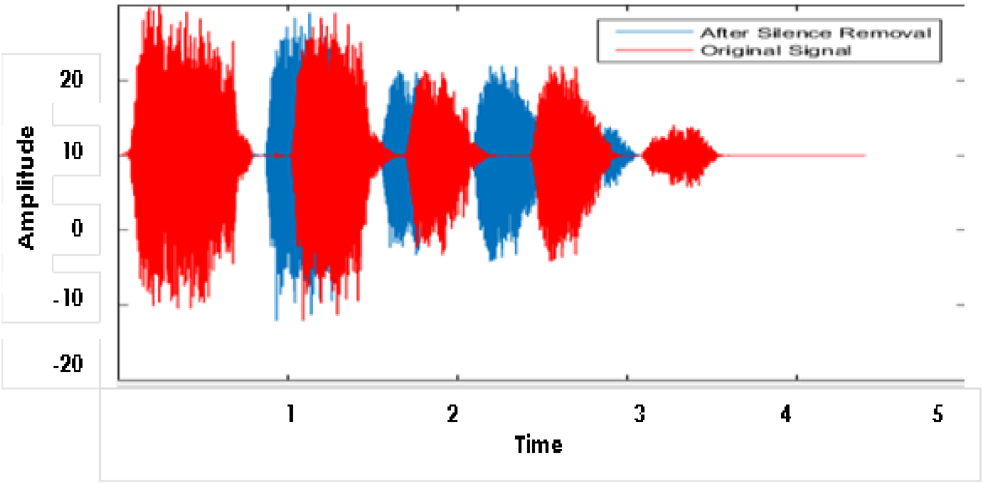

Figure 4. Animal signal before, and after silence removal

3.2.2 Signal transformation

Frequency features are essential and robust. Therefore, to extract these features, we need to transform the audio signal from the time domain (waveform) to the frequency domain (spectrum). Fourier transform is used precisely for this purpose. In our solution, we use the FFT that computes the frequency representation of 1024-sample windows at each frame. The output of FFT will be used to construct the desirable spectral features to represent the audio signal content [31].

3.3. Features extraction

After performing the pre-processing step, feature extraction usually takes place to detect various features that describe the audio content. Selecting proper acoustic features is the key to achieving effective classification performance. Frequency-based features provide a robust solution against environmental noise. The spectral features are one of the essential frequency features that have been widely used in many identification systems [13], [16], [19], [21].

In our approach, we focused on two temporal features: Root Mean Square (RMS), and Zero Crossing Rate (ZCR), and on two frequency features: spectral roll-off (SR) and spectral centroid (SC). To construct the feature vector, a general feature vector will be obtained for the whole acoustic signal by computing the mean for each feature obtained from all the (k) frames {RMSi, ZCRi, SCi, SRi} where i = 1,2, k.

3.3.1. Root Mean Square

It is a time-domain feature that calculates the signal strength average along with the signal. The mathematical equation of ZCR can be defined as follows, where x_i is the ith sample value of the frame, and N is the frame length [4], [11], [20]:

$$\text{RMS} = \sqrt{\frac{1}{N}\sum_{i=1}^{N} {x^2}_i} \tag{3}$$

3.3.2. Zero-crossing rate

It is a time-domain feature that measures the number of times the signal changes its sign per frame. The mathematical expression of ZCR can be defined as follows [7], [32] :

$$\text{ZCR}=\frac{1}{2(N-1)}\sum_{m=1}^{N-1}|signal[x(m+1)] - signal[x(m)]| \tag{4}$$

Where $x(m)$ is the value of the mth sampled signal and signal[ ] is the sign function, which defined as:

$$Signal[x(m)]=\begin{cases}1 & x(n) \geq 0\\ -1 & x(n) < 0\end{cases} \tag{5}$$

3.3.3. Spectral roll-Off

SR points to the spectral shape by estimate the spectral differences between a frame and the next one[11], [13], [25]. Spectral roll-off can be expressed by:

$$\text{S=max}(M\,.\sum_{n=1}^{M}|X_n|^2 \ \leq C.\sum_{n=1}^{M}|X_n|^2) \tag{6}$$

Where C is empirical,   is the DFT of the signal for the n-th sample, and M is half of the DFT [16].

3.3.4. Spectral Centroid

The spectral centroid is the center point of the spectrum. It is often associated with the brightness of the sound [4], [33].

3.4. Object Classification

Several classification techniques can be used for acoustic recognition. These techniques can be categorized into two general categories: First, supervised machine learning, in which all data are labeled, and the algorithms learn to predict the output from the input data, e.g., DT, MMD, and k-NN [34]. Second, is unsupervised machine learning, in which all data are unlabeled, and the algorithms learn to inherit the structure from the input data, e.g., artificial neural network, and clustering algorithms.

Once the features are extracted, the classification of the signal will take place. The feature vector will feed the classifier to build the model of the automated identification system. However, most of these classification methods cannot be directly applied in the WASNs because they significantly consume a large amount of energy and resources. To overcome these limitations, the used algorithm must be computationally low and consume limited storage space [11], [14].

Amongst the classification techniques used in recent studies, the kNN was widely used due to its simplicity; a kNN had been used in the proposed system. The kNN learns quickly from the input training instances because it does not require further processing tasks. In our solution, for the classification, the Euclidean distance function is applied, to determine the most similar k instances to the new unclassified instance. Then the k-NN uses the nearest k instances to determine the class with the majority vote and assigns that class to the new instance. The k value refers to the number of nearest neighbors used to help specify the class value to a given new instance [35].

The Euclidean distance function is a distance (similarity) function. It is commonly used when the ranges of all input attributes are in the same width, and numeric. It can be computed using the below formula where a, and b refer to the two input vectors, and the number of attributes is defined with t. Thus, for a given input attribute, j, aj, and bj are its input values [35].

$$E(a,b)=\sqrt{\sum_{\boldsymbol{j=1}}^{\boldsymbol{t}}(\boldsymbol{aj}-\boldsymbol{bj})^2} \tag{7}$$

3.5. Notification

Transmission of acoustic data over a long distance can significantly increase energy consumption in WSNs. Thus, the notification transmission aims to minimize energy consumption by sending only a few bits that represent the classification result (the object/animal).In each sensor node, if the target object was detected, a classification process will be applied. Then a notification packet with the classification result will be sent to the base station.

4. EXPERIMENTAL RESULTS AND DISCUSSION

4.1. Dataset description

The proposed system and all experiments were tested on the publicly available dataset from ESC-50 [36], [37]. The used dataset had 11 classes. The audio records were stored in WAV format with a sample rate of 44.1 kHz, mono channel, and encoding at 16 bits.

Collected audio data have to pre-process before being used. Hence, the pre-processing data stage should be considered to enhance the quality and the format of data that is directly affecting the quality of the accuracy. For the animal sound dataset, the pre-processing involves data filtering, audio segmentation, and audio transformation. The data filtering aims to correct the records that have outliers, or extreme values, by normalizing their values into normal records range. Then, each audio record had been segmented into equal-sized of frames using a Hamming window; each frame has 1024 sequential samples. Frames are overlapped by 50 %, to prevent any loss of information at the end of the frames. Finally, in the audio transformation process, we used FFT; to extract the frequency features spectral centroid and spectral roll-off.

4.2. Feature extraction in the proposed schema

In this research, we investigated four different feature extraction methods, in which two of them are a time-domain feature RMS and ZCR, and the others are frequency domain features SC and SR. We tested the performance of these features using different subgroups of our dataset and the whole dataset. Four features are extracted for each animal, and the mean for each feature was computed, as shown in Figure 5. The results show that, in general, the SC follows the same mean distribution for a different set of animals so, it can't distinguish different classes accurately. Therefore, we concluded that RMS, ZCR, and SR features are the most suitable set of features to describe the target animal while adding SC will not contribute to enhancing the recognition accuracy and will increase the complexity of the application. Thus, in this research, we had been adopted the RMS, ZCR, and SR features to generate a unique descriptor.

So, in the features extraction process, we constructed the classification model depend on the RMS, ZCR, and SR. Different features' vectors, had been extracted from all training sets. Where these three features and the class label were assigned to each record and saved them in CSV file as a feature reference vectors, this CSV file will be stored in the database to be used later in the similarity matching stage. These feature reference vectors are supposed to be loaded into the sensor memory during the set-up process. During the classification process, these vectors will be compared with the detected object feature vector (original test record) to identify the type of the detected animal.

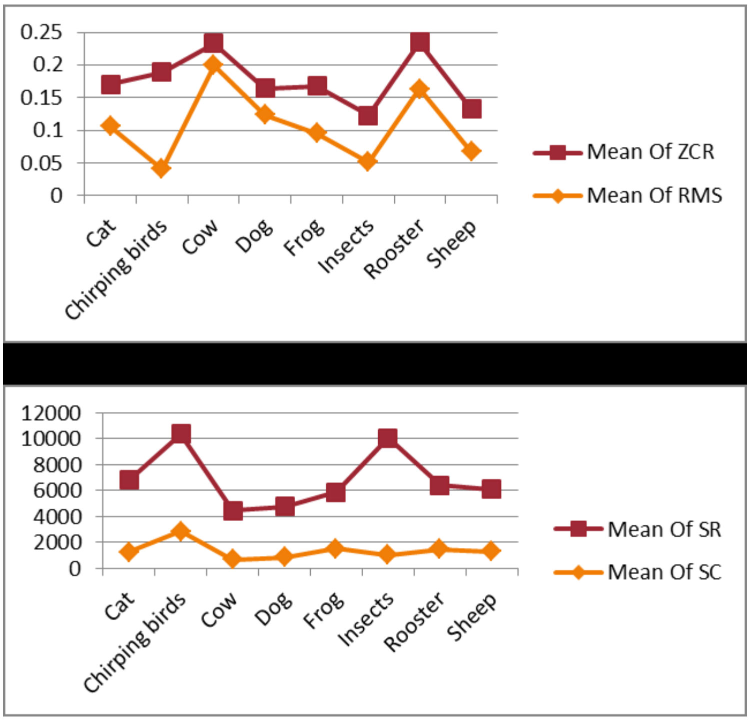

Figure 5. The mean distribution for four features using the full datasets

4.3. Performance Evaluation

The main criterion used to evaluate the performance of the proposed scheme is the acceptable recognition rate. To measure the performance of the classification system, we used the accuracy measure, which can be calculated using the following formulas:

Accuracy = (Number of corrected predication X 100) / (Total of all cases to be predicated) (8)
We studied the capability of three features (RMS, ZCR, and SR) to discriminate between different animals using k-NN, Decision Tree (DT), Support vector machine (SVM), and Naive Bayes (Figure 6). We first used different subgroups, and combinations from our dataset classes depend on the classification confusion matrix. Then we split the dataset into 85% as the training set and the remaining as a testing set.

Finally, the machine learning classifiers were tested to approve, the first assumption that the k-NN classifier best suits to the animal sound classification in WSN. The results show that the proposed scheme was capable to correctly classify 32 out of 36 targets (accuracy of 88.88%). We also note from Figure 6 that the DT classifier got better accuracy than SVM and Naive Bayes. Nevertheless, the two remaining classifiers performed poorly for both accuracy and time needed to build a classification model.

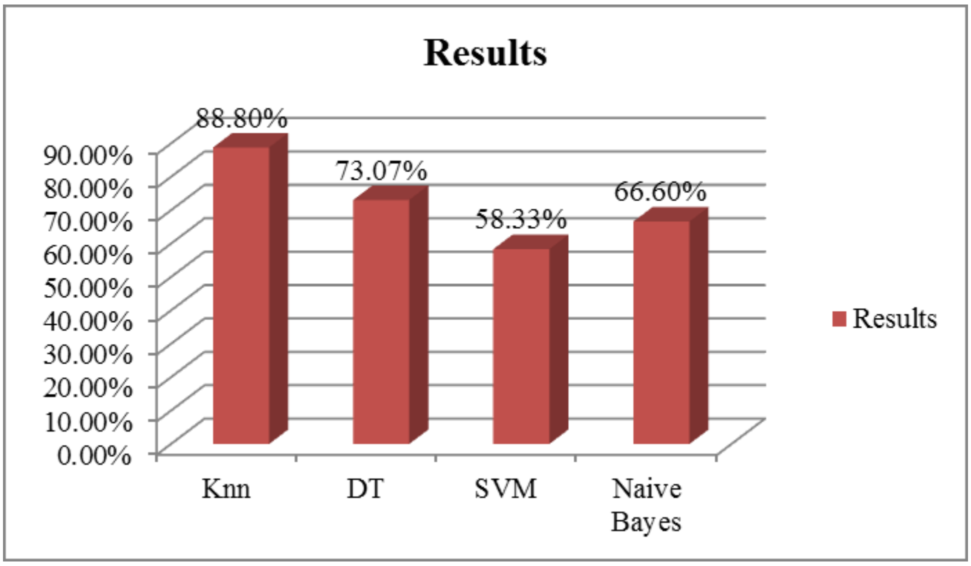

Figure 6. Accuracy of the proposed schema using different classification techniques

4.4. Energy Efficiency of the Proposed Scheme for Sensor Node

In our research, we investigated the energy efficiency performance of our proposed scheme. We measured the energy consumption of features extraction, classification, and notification tasks for one sound file on the sensor node. We compared the results when we send a whole audio file to the BS for classification tasks. The AVRORA [38], [39] simulator is used to evaluate the consumed energy. AVRORA is an instruction-level Sensor network that emulates the real implementation of different wireless sensor nodes. AVRORA simulator supports the most widely used sensor nodes such as (MICA2dot, MICA2, and MICAz). These sensors are equipped with low power (limited battery lifetime). The primary purpose of such a framework is to execute TinyOS applications before deployment in actual WSN nodes. The AVRORA helps to estimate the number of power consumption and processing time.

We focused mainly on measuring the performance of a sensor node (Mica2) during recognition and notification tasks. Table 2 presents the energy consumption of audio recognition and sends notification tasks. As shown in Table 1, the audio recognition process with the submitted notification task is less in energy and time-consuming than sending one sound file, Therefore, performing the classification step is essential to reduce the communication overhead by sending only useful information to the end-user and prolong the network lifetime.

Table 2. Evaluation of the recognition cost and sending sound file on MICA2

Measured Attribute	**Time (s)**	**Energy (Joule)**
Classification one sound file	4	0.09
Classification one sound file	3.7	0.08
Send Notification (I packet)	0.018	0.0177
Send One sound file (441.000 bytes)	7.938	7.8075
Classification + Notification	4.018	0.1077

5. DISCUSSION

The proposed recognition methods in the literature review didn't address the complexity and the power consumption to prove algorithms' efficiency. Most of the adopted features are either too complicated or requiring too much memory. Also, these approaches require extracting a large set of features, which consumes additional storage, processing, and communication resources. In such studies, authors focused on increasing the recognition accuracy and don't take into consideration the energy and power consumption. Hence, these approaches are not suitable for real-time resource-constrained applications.

Unlike previous works, in our proposed scheme, we guaranteed a good balance between recognition performance and sensor node resource restrictions. It is shown from the results obtained from the Avrora simulator that the selected set of feature extraction methods (RMS, ZCR, and SR) can easily fit on sensor nodes. Moreover, the recognition scheme was capable of identifying different objects with an acceptable level of accuracy.

6. CONCLUSIONS

In this paper, we presented the implementation of an animal sound recognition system in WASN. We can conclude that the results of applied RMS, ZCR, and SR feature using k-NN shown acceptable classification accuracy. Furthermore, we proved that sent notification tasks would reduce the energy consumption and prolong the network lifetime rather than transmitting a whole sound file to the BS, which hardly consumes the sensor energy.

REFERENCES

[1] J. Yick, B. Mukherjee, and D. Ghosal, "Wireless sensor network survey," Comput. Networks, vol. 52, no. 12, pp. 2292–2330, 2008.

[2] A. Bertrand, "Applications and trends in wireless acoustic sensor networks: A signal processing perspective," 2011 18th IEEE Symp. Commun. Veh. Technol. Benelux, SCVT, 2011, 2011.

[3] B. Chen, "Audio Recognition with Distributed Wireless Sensor Networks," pp. 1–48, 2010.

[4] J. Xie, "Acoustic classification of Australian frogs for ecosystem surveys," 2017.

[5] D. Mitrovic, M. Zeppelzauer, and C. Breiteneder, "Discrimination and Retrieval of Animal

Sounds," 2006 12th Int. Multi-Media Model. Conf., pp. 339–343, 2005.

[6] J. Colonna, T. Peet, C. A. Ferreira, A. M. Jorge, E. F. Gomes, and J. Gama, "Automatic Classification of Anuran Sounds Using Convolutional Neural Networks," Proc. Ninth Int. C* Conf. Comput. Sci. Softw. Eng. - C3S2E '16, pp. 73–78, 2016.

[7] A. D. Mane, R. A. Rashmi, and S. L. Tade, "Identification & Detection System for Animals from their Vocalization," no. 3, pp. 1–6, 2013.

[8] J. Xie, M. Towsey, J. Zhang, and P. Roe, "Frog call classification: a survey," Artif. Intell. Rev., pp. 1–17, 2016.

[9] M. P. Malumbres and C. Science, "On the Design of a Bioacoustic Sensor for the Early Detection of the Red Palm Weevil," pp. 1706–1729, 2013.

[10] P. Xie, Jie, Towsey, Michael, Yasumiba, Kiyomi, Zhang, Jinglan, Roe, "Detection of anuran calling activity in long field recordings for bio-acoustic monitoring," IEEE Int. Conf., 2015.

[11] J. Xie, M. Towsey, J. Zhang, and P. Roe, "Acoustic classification of Australian frogs based on enhanced features and machine learning algorithms," Appl. Acoust., vol. 113, pp. 193–201, 2016.

[12] Q. F. Gary G. Yen, "Automatic frog call monitoring system: a machine learning approach," Appl. Sci. Comput. Intell. V.

[13] C. J. Huang et al., "Intelligent feature extraction and classification of anuran vocalizations," Appl. Soft Comput. J., vol. 19, pp. 1–7, 2014.

[14] J. G. Colonna, A. D. Ribas, E. M. Santos, and E. F. Nakamura, "Feature Subset Selection for Automatically Classifying Anuran Calls Using Sensor Networks," pp. 10–15, 2012.

[15] Z. Chen and R. C. Maher, "Semi-automatic classification of bird vocalizations using spectral peak tracks," J. Acoust. Soc. Am., vol. 120, no. 5, pp. 2974–2984, 2006.

[16] C. J. Huang, Y. J. Yang, D. X. Yang, Y. J. Chen, and H. Y. Wei, "Realization of an intelligent frog call identification agent," Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics), vol. 4953 LNAI, pp. 93–102, 2008.

[17] C. Bedoya, J. M. Daza, and J. D. Lopez, "Automatic Recognition of Anuran Species Based on Syllable Identification," Ecol. Inform., vol. 24, no. January 2016, 2014.

[18] J. G. Colonna, M. Cristo, M. Salvatierra, and E. F. Nakamura, "An incremental technique for real-time bioacoustic signal segmentation," Expert Syst. Appl., vol. 42, no. 21, pp. 7367–7374, 2015.

[19] C. J. Huang, Y. J. Yang, D. X. Yang, and Y. J. Chen, "Frog classification using machine learning techniques," Expert Syst. Appl., vol. 36, no. 2, PART 2, pp. 3737–3743, 2009.

[20] S. Chu, S. Narayanan, and C.-C. J. Kuo, "Environmental Sound Recognition With Time-Frequency Audio Features," IEEE Trans. Audio. Speech. Lang. Processing, vol. 17, no. 6, pp. 1142–1158, 2009.

[21] N. C. Han, S. V. Muniandy, and J. Dayou, "Acoustic classification of Australian anurans based on hybrid spectral-entropy approach," Appl. Acoust., vol. 72, no. 9, pp. 639–645, 2011.

[22] J. Xie, Z. Jinglan, and P. Roe, "Acoustic features for multi-level classification of Australian frogs : Family, Genus and Species," IEEE, pp. 1–5, 2015.

[23] J. Xie, M. Towsey, A. Truskinger, P. Eichinski, J. Zhang, and P. Roe, "Acoustic classification of Australian anurans using syllable features," Information, Commun. Signal Process. (ICICS), 10th Int. Conf. on. IEEE, 2015.

[24] A. Camacho, A. Garcia Rodriguez, and F. Bolanos, "Automatic detection of vocalizations of the frog Diasporus hylaeformis in audio recordings," J. Acoust. Soc. Am., vol. 130, no. 4, p. 2500, 2011.

[25] S. Gunasekaran and K. Revathy, "Automatic Recognition and Retrieval of Wild Animal Vocalizations," IJCTE Int. J. Comput. Theory Eng., vol. 3, no. 1, p. 136�140, 2011.

[26] J. Xie, M. Towsey, P. Eichinski, J. Zhang, and P. Roe, "Acoustic feature extraction using perceptual wavelet packet decomposition for frog call classification," Proc. - 11th IEEE Int. Conf. eScience, eScience 2015, pp. 237–242, 2015.

[27] B. Croker and N. Kottege, "Using feature vectors to detect frog calls in wireless sensor networks," J. Acoust. Soc. Am., vol. 131, no. 5, pp. EL400–EL405, 2012.

[28] Vaca-Castano, Gonzalo, and D. Rodriguez, "Using Syllabic Mel Cepstrumfeatures and K-Nearest Neighbors to Identify Anurans and Birds Species," Signal Process. Syst. (SIPS), 2010 IEEE Work. On. IEEE, pp. 466–471, 2010.

[29] X. Dong, M. Towsey, J. Zhang, and P. Roe, "Compact Features for Birdcall Retrieval from Environmental Acoustic Recordings," Proc. - 15th IEEE Int. Conf. Data Min. Work. ICDMW, 2015, pp. 762–767, 2016.

[30] M. Vasile, C. Rusu, R. Ciprian, and J. Astola, “Audio based solutions for detecting intruders in wild areas,” Signal Processing, vol. 92, no. 3, pp. 829–840, 2012.

[31] M. Nikolic and J. Goldstein, “Enabling Signal Processing over Data Streams,” 2017.

[32] F. Alías, J. C. Socoró, and X. Sevillano, “A review of physical and perceptual feature extraction techniques for speech, music and environmental sounds,” Appl. Sci., vol. 6, no. 5, 2016.

[33] S. Fagerlund, “Bird species recognition using support vector machines,” EURASIP J. Adv. Signal Process., vol. 2007, 2007.

[34] D. Bzdok, M. Krzywinski, and N. Altman, “Machine learning : Supervised methods , SVM and kNN,” pp. 1–6, 2018.

[35] W. D Randall and M. Tony R, “Reduction Techniques for Instance-Based Learning Algorithms,” Mach. Learn., vol. 38, pp. 257–286, 2000.

[36] K. J. Piczak, “ESC: Dataset for Environmental Sound Classification,” Proc. 23rd ACM Int. Conf. Multimedia, MM 2015, pp. 1015–1018, 2015.

[37] Y. Aytar, C. Vondrick, and A. Torralba, “SoundNet: Learning Sound Representations from Unlabeled Video,” no. Nips, 2016.

[38] F. Yu, “A Survey of Wireless Sensor Network Simulation Tools,” Washingt. Univ. St. Louis, Dep. …, pp. 1–10, 2011.

[39] K. Erciyes, O. Dagdeviren, O. Yılmaz, and H. Gumus, “Modeling and Simulation Tools for Mobile Ad Hoc Networks,” no. November, pp. 37–70, 2011.