www.sensorsportal.com

Sensors & Transducers
Published by IFSA Publishing, S. L., 2018
http://www.sensorsportal.com

Design of an Energy Efficient Scheme for Target 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, P. O. Box 51178, Riyadh 11543, Saudi Arabia
* Tel.: +966114676578
* E-mail: asoudani@ksu.edu.sa

Received: 30 April 2018 /Accepted: 31 July 2018 /Published: 31 August 2018

Abstract: The conservation of ecosystem diversity has become a major concern worldwide. To optimize the management of ecosystems, animal recognition has become an important topic of research. Animal-monitoring systems can produce valuable information that assists in the protection of animals, planets and the natural environment in general. The majority of monitoring applications require a lightweight recognition system that can effectively achieve an effective trade-off between cost and performance. In this context, designing a solution for target recognition using Wireless Acoustic Sensor Networks (WASNs) has become a viable low-cost approach for monitoring animals. However, acoustic sensors generate large volumes of data, which represents a major restriction on the network deployment. As sensors are massive energy-constrained devices, energy consumption becomes the most critical consideration for data-intensive computation and communication. In this context, this paper proposes a low-energy scheme that focuses on developing efficient processing algorithms and communication techniques that can optimize energy consumption and prolong the network's lifetime, while guaranteeing the required level of application performance. This scheme, based on temporal feature extraction methods, aims to recognize the target object locally at the sensor node, and then report the event with small-size packets. The results showed that the proposed approach was capable of reducing the amount of consumed energy in the network, while maintaining competitive recognition accuracy.

Keywords: Bio-acoustic, Feature's extraction, Wireless acoustic sensor networks, Recognition, Low-energy processing.

1. Introduction

Recently, animal biodiversity has been threatened with extinction because of the effects of human activities and climate changes. Animal monitoring forms a fundamental solution to various problems in the real world, such as tracking animals' migration paths, protecting plants from harmful organisms, and detecting environmental pollution. Based on such studies, the development of an automatic system for animal monitoring is certain to have significant scientific and economic value. Wireless Sensor Networks (WSNs) have been widely applied to real-world environmental monitoring and offer great potential for long-term monitoring at spatial and temporal scales that are difficult to achieve through traditional systems. Among the broad palette of monitoring applications of WSNs, there has been significant interest in the use of acoustic sensor nodes in various animal-monitoring scenarios. Enabling WSNs support for audio data has expanded the horizon of traditional monitoring applications by providing audio verification, enlarging sensing coverage, and many more. Such acoustic-oriented solutions typically used for detection, classification, identification and tracking of targets. This paper tackles one of the crucial applications of animal monitoring; namely target recognition.

Like most real-time applications, target monitoring requires persistent network connectivity and extensive data transmission. Hence, target recognition via WSNs is a very challenging task as sensor network poses many constraints related to communication bandwidth, time-synchronization, and power supply. The acoustic sensing streaming capabilities of WASNs can further increase the design complexity, especially when dealing with a full acoustic signal. Typically, acoustic signals rely on high sampling rates, which result in expensive computation and communication cost, and may eventually drain the power supply of battery life. Since sensor nodes run mainly on tiny batteries, energy efficiency becomes the most critical consideration in wireless sensor networks and a system lifetime. This has arisen the need to design energy aware sensing strategies during the system operation in order to cope with WASNs restrictions and to achieve low energy depletion rates for sensor nodes.

In this paper, an energy-conscious acoustic sensing solution is proposed, and has been applied in target recognition for animal monitoring application scenarios. Through a smart edge sensing strategy, we enabled individual sensor to locally detect the presence of the target object in its sensing range before sending notification packets to the end user. Hence, rather than transmitting notification packets to report on every object that appears in the monitored area, a sensor will only send notifications once it detects a matching target. This approach makes use of feature extraction techniques to obtain a set of characteristic features from the acoustic signal. These can then be used as a unique signature to identify targeted objects. Such method also work as a dimensionality reduction technique as it aims to send a reduced feature set to describe the target object instead of sending the entire acoustic signal. This smart edge sensing strategy can minimize the size and the number of sent data packets, leading to a reduction in the energy spent on transmitting these notifications and thus in the energy of the overall network during its lifetime. Nevertheless, achieving the intended outcomes is directly dependent on the appropriate selection of feature extraction and classification methods that enable target recognition with a high accuracy rate and low energy consumption.

Moreover, forwarding these data directly to the base station, that is typically located far away from the network field, would require a high level of transmission power especially when it is frequently performed. This has emerged the need to use in-network data processing mechanisms in order to reduce the communication burden between sensor nodes. Towards this goal, cluster-based architectures have been widely adopted to improve resource allocation and power control in WASNs. In this context, we have proposed a cluster-based framework for target recognition as an associated solution for reducing the transmission cost and consequently maximizing network longevity.

The remainder of this paper is organized as follows. Section 2 presents an overview of the related works. Section 3 describes the general approach for low-power acoustic sensing scheme for target recognition. The specification of this scheme, including all design and architectural aspects are described in Section 4. Section 5 presents the experimental evaluation of the scheme performance. The paper closes with a discussion of the scheme capability in achieving energy gains in the WASN and presents the general conclusion.

2. Related Work

Acoustic-based target recognition applications have been attracting a great deal of scientists' attention. These applications usually focus on the use of feature extraction methods to obtain relevant characteristics and then use classification algorithms to make the recognition decision. In the last decades, researchers have experimented various feature extraction techniques for animals' recognition. These techniques can be classified based on their domain of computation into three general categories: temporal, frequency, and cepstrum [1]. Most of the presented schemes in reviewed studies combine various different features in an augmented vector in order to obtain high recognition accuracy levels.

In [2], Zero Crossing Rate (ZCR) and energy entropy were used with k-Nearest Neighbors (k-NN) classifier to classify seven different species of anurans, achieving an accuracy of 90 %. Whereas in [3], a wavelet transform with k-NN classifier is selected to identify ten frog species with an average classification accuracy of 97.95 %. While in [4], different species of anurans were classified with an identification rate of 98 % by employing k-NN classifier on a combination of extracted Mel Frequency Cepstral Coefficients (MFCCs) features and Linear Prediction Coefficients (LPC) features. Among all adopted features, MFCCs features have been widely and commonly used for animal recognition tasks beside machine learning classification algorithms.

In [5], the authors developed an intelligent system that capable of recognizing the frog species using a combination of six acoustic features, including cepstral coefficients and acoustic indicies. While in [6] Linear Frequency Cepstral Coefficients (LFCC) and MFCC are extracted from the acoustic signal of different reptiles and classified using two classification algorithms, which are kNN and Support Vector Machine (SVM). In [7-9], introduced a method to recognize anurans species using syllable feature, in conjunction with MFCC, and other features such as energy and LFCC. The results showed that the proposed set outperforms the approaches that use only

MFCC, LFCC, or both of them for the classification task. In [10], Xie, *et al.* used a set of syllable features, Linear Predictive Coding (LPC), and MFCC features to classify twenty-four frog species using a five different machine learning algorithms Alternatively, In [11], the syllable extraction approach is adopted to classify anuran species using various temporal features such as the classical signal energy and the zero crossing rate.

Most of the presented approaches involve performing some kind of transformation and extracting a large set of features. Specifically, these techniques transform the signal from temporal domain to frequency domain before extracting acoustic features. Although the presented approaches have outstanding performance compared to time and frequency domain features, they also demanding huge computational power, processing time, and memory space. Moreover, the machine learning classification algorithms have been widely adopted in many of these studies to recognize a single target. However, the proposed object recognition solutions in the literature didn't address the energy consumption and transmission cost to prove the feasibility of the proposed schemes for a low-power sensor node. We think these approaches are not suitable for real-time resource-constrained applications. Thus, it becomes essential to develop an energy efficient recognition scheme that provides good balance between algorithm performance, complexity, and resource limitation.

3. General Approach for Acoustic-based Low-power Sensing

The network consists of a finite set of acoustic sensor nodes and a sink node that is distributed over an area of interest to monitor various acoustic targets (Fig. 1). In this paper, however, we only considered detecting a single target at a time. All sensor nodes are assumed stationary and their positions are known in advance. The network is partitioned into several disjointed clusters, each consisting of Cluster Head (CH) and various member nodes. The formation of the clusters depends upon the adopted clustering algorithm, which is beyond the scope of this paper. Basically, our target recognition solution is based on a four-step process: advertising target signature, detecting and recognizing a target object, selecting a vector of features, and transmitting notification messages. Member sensor nodes are responsible for detecting objects that enter theirs sensing range, recognizing targeted objects, and transmitting the targets detection information to their CH. While the cluster head takes the responsibility of broadcasting the target signature, selecting the vector of features, and transmitting notifications to the sink node.

During the network set-up phase, we must ensure that every member node knows the target object reference vector. Hence, the sink node will broadcast the target signature *Ref* (an identification vector) to the entire CHs in the network, which in turn will advertise it to their member nodes. Member sensors will load the received signature into their memory to be used later on during the target recognition task. During the system operation, these member nodes will obtain periodic observation samples to detect any even of interest. A threshold detection procedure is adopted in which sensors will monitor the change in signal intensity and make the detection decision once a certain pre-determined threshold T_{thre} is crossed. In the proposed solution, we focused on decreasing the notification packet generation rate of sensor nodes. Therefore, instead of triggering a notification to the CH upon the detection of any acoustic object, which may not be the targeted object, sensor nodes will first identify the type of detected object using a low-cost recognition strategy.

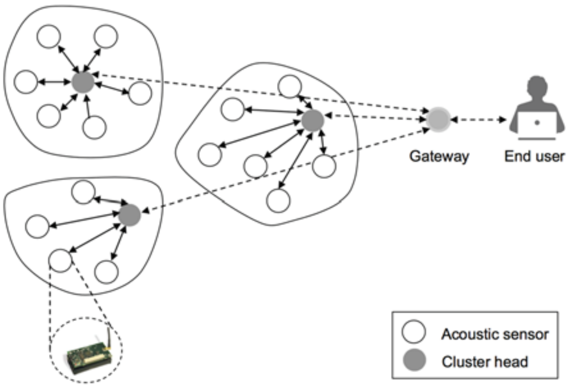


Fig. 1. A Framework for object recognition in WASN.

When an event of interest takes place, sensor nodes will extract a set of characteristic features from the acoustic sample and combine it into a single vector that represents the newly detected object $f_i=\{feature_1, ..., feature_N\}$. Then, with the use of these features, each sensor will identify the type of the object using a similarity-based classification algorithm that matches the newly extracted vector (f_i) and the stored reference signature (*Ref*). Once a sensor recognizes the target object, it will send a notification packet to CH to report the target presence in the area. To further minimize the transmission energy, packet size optimization approach has been proposed to reduce the size of the notification packets and avoid sending the entire acoustic signal to the CH. In the proposed solution, we adopted two different types of short notification messages that can be implemented according to the end user specification, which are binary (yes/no) notification or a vector of features notification. The former message is intended to report the target object presence in the area while the latter will be used to further verify the recognition results with more powerful classification algorithms at the end user side.

After receiving a sufficient number of detections from the neighboring member sensors, the CH will make the final detection decision. This strategy will conserve the CH transmission energy by avoiding sending detection notification to the sink node based on false detections. The CH will deal with received packets in two ways based on the system specification, which can be binary notification or vector of features notification. However, aggregating all the received packets and forwarding them to the end user may represent a source of energy wastage. For instance, these packets may contain redundant data or noisy measurements that are unbeneficial to the end user. To this end, the CH will execute a selection procedure in order to send only the vector of features that belongs to the nearest senor to the target object, which often provides more reliable measurements. The adopted scheme is expected to be able to reduce energy consumption levels and conserve communication bandwidth, leading to a prolonged network lifetime.

4. Specification of the Sensing Scheme for Target Recognition

Target recognition has received much attention in recent years. Several recognition mechanisms were proposed for achieving high recognition accuracy using WASNs. However, most of them require a complicated processing of the input acoustic signal using high computational complexity feature extraction and classification algorithms, which increase the energy consumption levels in the network. Energy efficacy is one of the fundamental issues in sensor networks since battery-driven sensor nodes usually deployed in inaccessible areas, which makes it almost impossible to replace or recharge their batteries. Hence, once a WSN is deployed, sensor nodes should be able to operate with their initial amount of residual energy. Moreover, these sensors may provide unreliable observations when their energy levels drop under a certain threshold [12]. Thus, individual sensor nodes have significant power constraints as the amount of consumed energy has a major impact on the network performance and the sensor lifetime. Consequently, the appropriate selection of the feature extraction and classification algorithms becomes a decisive factor for reaching the intended goals from the proposed scheme.

The selected auditory processing methods should be capable of optimizing energy usage while guaranteeing the required level of performance. The problem of choosing the most suitable algorithms for target recognition via WASNs is guided by two main factors, which are network resource constraints and algorithm computational complexity. These emerging networks carry many constraints and technical challenges that related to the limited butter size, communication bandwidth, and power supply. Hence, we seek to find a low dimensional feature vector that can sufficiently represent a specific object uniquely using minimum number of features. Moreover, the classification algorithm should be capable of correctly classifying objects from a limited number of input features or model parameters and using a considerable high accuracy rate. This approach can contribute in reducing the overhead on the memory space and bandwidth demands. Nevertheless, the energy consumption levels in sensor nodes are directly influenced by the computational complexity (in the number of clock cycles) of adopted algorithms, as it can increase the energy expenditure rates during data processing. Hence, it is necessary to adopt low complexity algorithms that based on simple mathematical operations, which can generally execute fewer instructions to perform the recognition process.

The general scenario considered for target recognition is depicted in Fig. 2. An object is supposed to appear in a specific area covered by WASN. Let us assume that at the time *t*, member sensors will measure the average acoustic signal energy $E_i(t)$ and detect the presence of an event of interest. These sensors will then extract the signal's features and each will construct a vector of features. Afterward, the detected object feature vector will be compared with the target descriptor, which has been previously loaded into the sensor memory. Once the target is recognized, sensors will notify the CH by transmitting an appropriate notification packet. Then, the CH will find the vector that belongs to the closest sensor to the target object and forwarded it to the end-user for further verification.

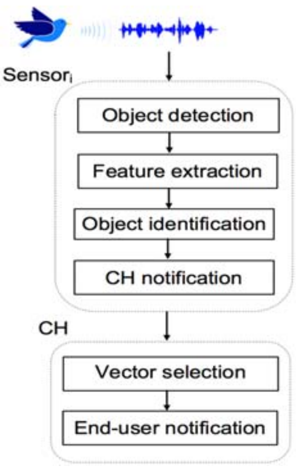


Fig. 2. General scheme for object recognition process.

A target recognition task consists of five major steps: object detection, signal framing, feature extraction, object classification, and detection notification. In the following sections, we detailed the specifications of each stage.

4.1. Object Detection

The object detection algorithm presented in this paper is based on an energy threshold detector. The main idea is that the member sensors in each cluster will be needed to directly monitor changes in the acoustic signal energy during acquiring a new acoustic signal in a constant periodic manner (Δ*t*), over a time interval (*t*). The usage of an energy detector is to enable the differentiation between the acoustic events and the anomalous noise events, without determining whether the detected object is the target or not. We employ the Root Mean Square (RMS) time domain feature in the detection algorithm in order to measure the average signal energy $E_i(t)$ during that time interval. After computing the energy, the detection parameter that represents the average energy is compared to a predefined threshold value (T_{thre}) in order to yield a detection decision. The captured audio stream of (*M*) samples is then passed to the next step for further signal processing and features extraction. The detection function (D) can be defined as follows.

$$D = \begin{cases} 1 & E_i(t) > T_{thre} \\ 0 & E_i(t) \leq T_{thre} \end{cases} \tag{1}$$

The object recognition technique described here is based on the following sequential steps: signal framing, feature's extraction and object classification.

4.2. Signal Framing

Acoustic signals are generally preprocessed before features are extracted to enhance the computational efficiency of the extraction process and hence increase overall classification accuracy. Signal framing is considered as one of the foremost preprocessing steps in any acoustic signal processing system. This framing comes from the necessity of transforming the signal into blocks (time windows) during which it is assumed to be statistically stationary.

To prevent any loss of information at the end of the frames, consecutive frames typically should overlap by 50 %, which is a common choice in signal processing [13]. In our solution, the continuous acoustic signal of (*M*) samples is partitioned into equal sized (*K*) frames of short duration (Fig. 3); each frame has 1024 sequential samples in which 50 % of them are overlapped between two successive frames. Next, each frame is passed through a set of feature extraction algorithms.

4.3. Feature Extraction

We are addressing the design of a low complexity recognition method that intends to identify a specific object based on its most discriminate features in order to reduce the power consumption during the notification task. Hence, the proposed scheme should adopt low-complexity features' extraction techniques, in which the goal is to minimize the heavy processing burden and transmission overhead on the network. Nevertheless, the performance of the proposed scheme depends mainly on the ability of the extracted features to discriminate between different objects. The feature selection stage seeks to find a low dimensional feature vector that represents a specific object uniquely and using a considerable low computation cost. In this section, we examined various types of feature extraction methods in order to find an optimal set of low-cost features that can describe the target animal uniquely.

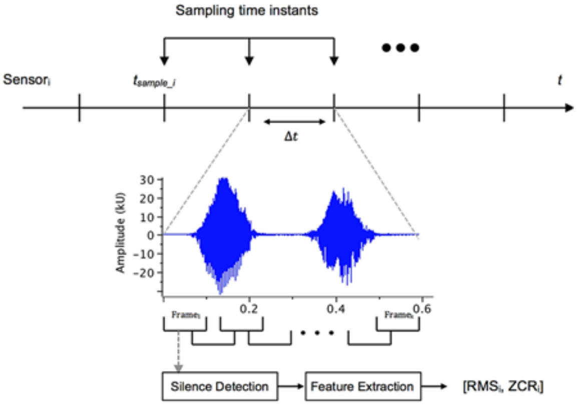
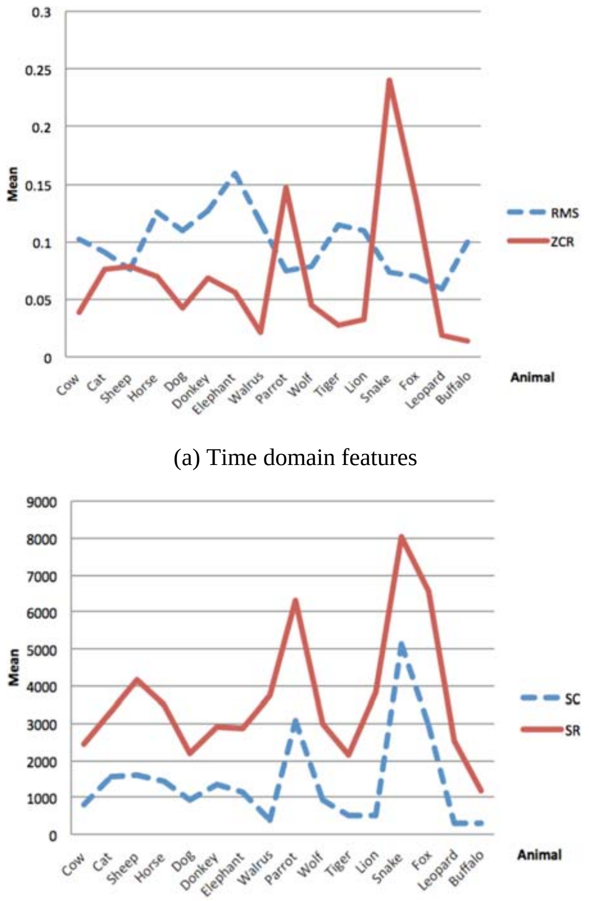


Fig. 3. General scheme for features' extraction process.

In this study, we investigated four different feature extraction methods, in which two of them are a time domain feature: RMS and Zero Crossing Rate (ZCR); and the others are frequency domain features: Spectral Centroid (SC) and Spectral Roll-off (SR) [1]. These features have been widely used in the literature to identify a specific animal from the environmental acoustic signal. In addition, these feature extraction techniques can provide a low complexity solution that is suitable for resource-constrained WASN applications. We tested the performance of these features using a set of sixteen animals: {Cat, Cow, Dog, Donkey, Horse, Parrot, Sheep, Buffalo, Elephant, Fox, Leopard, Lion, Snake, Tiger, Vireo, and Wolf}. We expanded the dataset that used in our previous work [14] to include more animals, in order to demonstrate the efficiency of the proposed selected features. Four features are extracted for each animal, and the mean for each feature was computed, as depicted in Fig. 4.

The results show that, in general, the ZCR, SC, and SR features actually follow the same mean distribution for a different set of animals. However, we noticed that frequency domain features require a high computational complexity compared to time domain features, while providing the same discrimination performance. In particular, the SC feature represents the relative amounts of high and low frequency energy. On the other hand, ZCR is also commonly used to measure the high frequency energy. The two features are closely related as they both can be used to measure the spectral shape of the audio signal [15]. Therefore, in our approach, we adopted the RMS and ZCR features to generate a unique descriptor for the target animal.



(a) Time domain features

(b) Frequency domain features

Fig. 4. The mean value of different features.

- Root Mean Square (RMS) Feature: The RMS is a measure of the power of the signal over time. Hence, it is commonly used for detection of silent segments in the audio signal. The signal amplitude is calculated by squaring every data point in a block of a sample, and then taking the mean square of these values as defined by [1]:

$$RMS = \sqrt{\frac{1}{N}\sum_{i=1}^{N} x_i^2}, \tag{2}$$

where x_i is the i^{th} sample value of the block, and N denoted as the block length.

- Zero Crossing Rate (ZCR) Feature: It counts the number of times a signal cross the zero axes within a frame. Since the RMS doesn't provide any information regarding the spectral proprieties of a signal, the ZCR is adopted for estimating the fundamental frequency of the audio waveform [1]:

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

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

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

The feature extraction process is based on the following sequential steps:

I. Silence detection: The recorded signal may contain periods of silence that need to be filtered out in order to enhance the extracted features' quality and decrease the processing time. In our approach, a silence frame is the frame whose RMS value is less than 10 % of the average energy of the entire acoustic target recognition. This step is based on the following three stages:

a) The silence threshold value S_{thre} is determined based on the calculated signal energy $E_i(t)$ during the object detection phase, as follows

$$S_{thre} = E_i(t)x\ 0.10 \tag{5}$$

b) The RMS_i value for each frame *frame_i* is calculated to measure the frame's energy level using (2).

c) Check the RMS_i value to determine the silence frame. If the value is less than the S_{thre}, it will be considered as silent and the next frame will be processed. Otherwise, the sensor will proceed forward to extract the second feature. The silence detection function (*S*) can be defined as follows

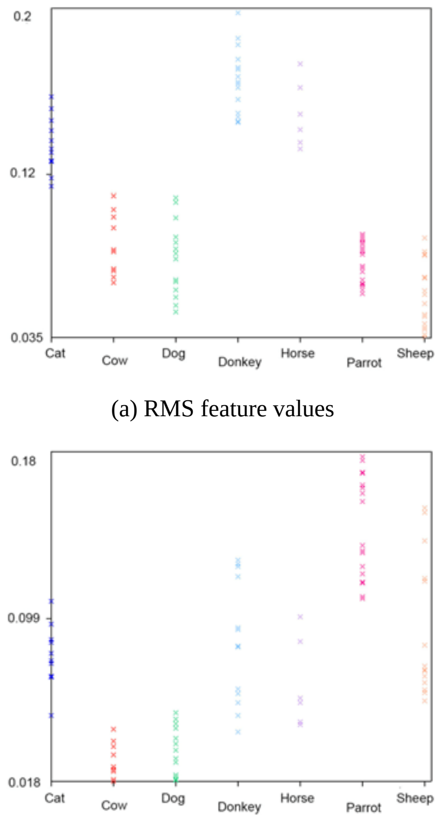
$$S = \begin{cases} 1 & RMS_i > S_{thre} \\ 0 & RMS_i \leq S_{thre} \end{cases} \tag{6}$$

II. Feature extraction per frame: For each non-silent frame, the RMS_i value is stored and the ZCR_i feature for that frame is extracted.

III. Feature vector construction: Obtain the global feature vector for the whole acoustic signal by computing the mean for each feature obtained from all the (*K*) frames $\{RMS_i, ZCR_i\}$, where $i = 1,2,..,K$.

4.4. Object Classification

The second step in the monitoring task involves using the obtained signal features to take the identification decision using suitable classification procedures. In the proposed solution, we adopted the Minimum Mean Distance (MMD) classifier [16], which is used to classify unknown object vector to classes that minimize the distance between the newly extracted vector and the target class mean vector. In spite of its simplicity, the low computation complexity and fast processing speed of the MMD classifier makes it an attractive solution in comparison with other algorithms. It was proven in previous work that this classifier could be an efficient tool for acoustic event recognition in various monitoring applications, where it was capable of providing high accuracy results [17]. Nevertheless, the huge overlapping in the feature values of different animals makes almost impossible to accurately identify a specific animal. This problem is illustrated in Fig. 5, where the feature values for different animals are showing high overlapping intervals.



(a) RMS feature values

(b) ZCR feature values

Fig. 5. The mean value of time domain features.

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 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. Later, the extracted feature vector can be used at the end user to do further classification in order to assign the target object to only one class.

The classification process is composed of two phases:

- Target reference vector construction: This step is performed offline at the end-user end during the system configuration process. In this phase, at first, we extract the feature vector of each training records, and then, we compute the mean vector for each object class. Hence, the target object will be represented by one reference vector that contains two mean values of RMS and ZCR features, as $Ref_i = \{\mu_{RMS}, \mu_{ZCR}\}$. The constructed vectors will be broadcasted to the sensor nodes using the configuration packet to load the sensors memory with the target signature.

- Object vector similarity matching: This step is performed locally at the sensor node level during the object recognition process. Each member sensor will measure the distance between the unknown object vector f_i and each target class mean vector f_i', which can be represented in the following distance vector $D_i = \{d_i, \ldots, d_N\}$. Then, it will find the two shortest distances in D_i vector, which represent the object matching target class labels. The similarity between objects is typically determined by the Euclidean distance metric [9] as follows

$$d = \sum_{i=1}^{D} |f_i - f_i'|^2 \qquad i = 1, \ldots, D, \qquad (7)$$

where f_i is the detected object vector, and f_i' is the mean vector of classes.

4.4. Object Notification

Data transmission may potentially represent a source of significant energy demand in the WSNs. Thus, the notification transmission phase is the most important key to minimize energy consumption and maximize system performance. In our solution, we were interested in decreasing the number of bits to be transmitted over the network even at the cost of increasing the processing overhead. We presented two notification opportunities to allow member nodes as well as cluster heads to reduce the amount of data to be received, processed, stored, and eventually transmitted. In this step, an important energy and time gain can be achieved for the acoustic sensor, as well as the whole network, which can be summarized as follows:

- Application level: A sensor node is composed of different hardware components such as a micro-controller, a transceiver, and an external memory. Typically, the transceiver is usually consuming a significant share of the total energy compared to other hardware components on a sensor node [18]. Hence, the overall energy consumption at the sensor mote is directly depends on the duration over which the radio is transmitting or receiving packets. Thus, reducing the power cost per transmitted bit can reduce the transmission time of the transceiver and conserve node energy.

- Network level: Controlling the amount of traffic in the network can reduce the cost of overall transmission power. In our solution, the source node restricts its packet flow; thereby balancing packets loads and reducing network congestion. In addition, it contributes in reducing transmission delay and minimizing the energy consumption required for retransmissions. Thus, the transmission scheme can achieve reduced energy.

A sensor node will assemble a notification packet according to the end user requirements and its role in the cluster as shown in Fig. 6.

- Cluster head notification: Each member node will assemble a notification and send it to the CH upon the detection of the target object, which can be either: few bits' notification or vector of features. The size of each notification in bytes is typically one Byte for detection notification. While it costs two Bytes per feature for the vector of features notification.

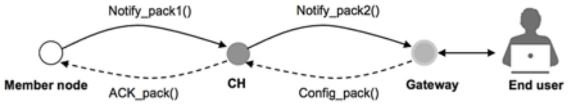


Fig. 6. Notification process in the cluster.

- End user notification: The CH will aggregate and process the received packets according to the notification type. Then, a notification will be send to the base station according to the application requirements, which can be either: few bits' notification or vector of features.

a) Detection notification: The CH will check the number of detected packets, if it is above a certain threshold, then the target object is supposed to be detected and a detection packet will be sent to the base station which will cost one Byte.

b) Vector of features: If the received notifications are above a certain threshold, the CH will find the packet with the highest RMS and forward it to the end user to do further classification, which will cost two Bytes for each feature in the target vector.

5. Implementation

This section presents the implementation details of the proposed scheme and describes the experimental evaluation of its performance at the application and the network level. The evaluation mainly focused on assessing the efficiency of the proposed scheme by analyzing its performance for object recognition. In the first part of this section, the performance's evaluation concerns the algorithm's ability to accurately identify the target object successfully. While the second part focused on evaluating the network performance during object recognition and packet notification tasks. We were mainly interested in measuring clock cycles, power consumption, and processing time of sensor nodes. This study also explores the scheme capability in achieving energy gains in the wireless acoustic sensor network by minimizing the required power consumption to achieve these tasks.

5.1. Performance Analysis at the Application Level

The success of the proposed solution to recognize the target will be the principal metrics to measure the performance of the proposed scheme at the application level. The capability of the proposed scheme to achieve this task is tested and evaluated using MATLAB tool.

5.1.1. Object Recognition

We were mainly interested to evaluate the system's capability to accurately identify and locate a specific object. The accuracy of the recognition algorithm will be measured based on the successful recognition rate. For this purpose, the proposed scheme was implemented with MATLAB tool. We conduct different experiments using different sound waves for the same object that were reordered under different conditions to measure the accuracy of the recognition algorithm. We explored the capabilities of the proposed scheme to classify a specific target object to two classes, in order to be used to detect that presence of the target object in the cluster. We adopted a cluster-based approach in order to reduce the large amount of data transmitted to the sink via local processing and aggregation in CH, and hence, prolong the whole network lifetime. In the proposed solution, we assumed that only one object is expected to appear in the monitored area.

5.1.2. Dataset Pre-processing

The raw acoustic records were collected from Animals & Birds Sound Effects CD [19] to evaluate our proposed two-label classification model. The rest of the records were gathered from various animal sounds libraries. This dataset contain 114 audio records belonging to seven different animals as shown in Fig. 7, namely, Cat, Cow, Dog, Donkey, Horse, Parrot, and Sheep. The selection of these animals was based on their habitat characteristics

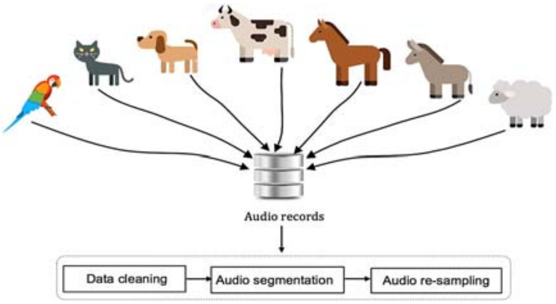
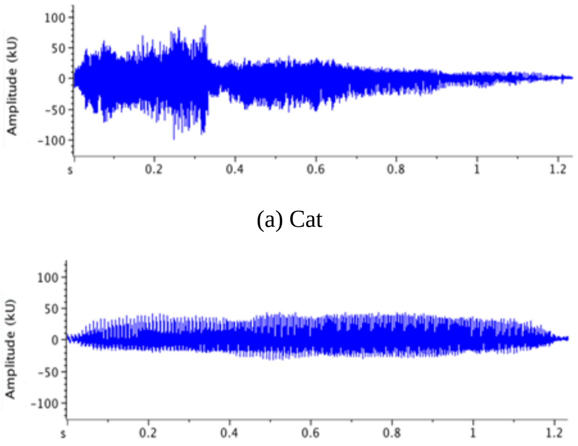


Fig. 7. Audio records pre-processing steps.

Our collection of records contains different sound waves for the same object that were recorded under different conditions. Fig. 8 shows an example of two animal sounds that have been used to evaluate the recognition performance. These records have different sound quality and come in a variety of audio formats. Due to the nature of natural environment, the acoustic records typically contain a large and diverse variety of sounds, which can all occur simultaneously in a single record, including human, other animals, and environmental phenomena such as the sound of waterfall. Moreover, these records can have variable recording length and various sampling frequencies. Hence, preprocessing is an essential procedure in order to take into account these factors and to maintain good recognition accuracy.



(a) Cat

(b) Cow

Fig. 8. Audio samples for two animals: (a) Cat; (b) Cow.

The preprocessing phase can be divided into three stages, including data cleaning, records segmentation, and signal re-sampling. The recorded signals are checked manually to evaluate the quality of the records; bad quality records are removed from the records collection. Then, all records are segmented with one or two seconds' length; in which each record contains only the sound of a single object animal. Finally, all records are resampled at 44.1 kHz frequency and saved as 16-bit wave format to avoid any bias in the classification task.

5.1.3. Target's Signature Construction

This step represents the first phase of the recognition process, which is the construction of the classification (learning) model. For each animal, a unique descriptor will be constructed. The main idea is to extract from each training record one vector of features that consists of two feature values {RMS, ZCR}. Then, for each animal type, the mean of each feature will be computed to construct the animal signature vector as shown in Table 1. These vectors will be used later during the classification process to identify the type of detected animals.

5.1.4. Performance Evaluation

The performance of the proposed scheme is tested using the successful recognition rate metric. This metric examines the number of times the algorithm was able to classify the object correctly, which can be expressed as

$$Accuracy = \frac{N_C}{N_S} \times 100, \tag{8}$$

where N_C is the number of sound samples that were correctly recognized, and N_S is the total number of samples.

Table 1. Animals reference vector.

Animal	Descriptor Vector	
	RMS	ZCR
Cat	0.1348	0.0809
Cow	0.0827	0.0287
Dog	0.0756	0.0338
Donkey	0.1674	0.0871
Horse	0.1496	0.0659
Parrot	0.0744	0.1444
Sheep	0.0568	0.0897

5.1.5. Recognition Results

To evaluate the recognition algorithm performance, four classifiers are considered in this study: MMD, Gaussian Mixtures Model (GMM), KNN, and Decision Tree (DT). The accuracy measures for each animal object using the selected classification algorithms are shown in Fig. 9. The MMD is tested as the local classifier employed by member nodes to classify the target object to two classes. The remaining classifiers are examined to find the most suitable algorithm for classifying the target animal into a single class at the end user. It can be observed that MMD classifier performed better than other classifiers and was capable to predict most of the animal classes correctly, gaining 96.88 % accuracy. We can also find that the GNN classifier outperforms the other two classifiers with 71.88 % recognition accuracy. Nevertheless, the performance of the three classifiers for the Horse and Cow has result in poor recognition accuracy. The relatively small training samples used to represent these two animals are not sufficient for developing a classification model that can accurately predict the animal class correctly, which caused under-fitting of data.

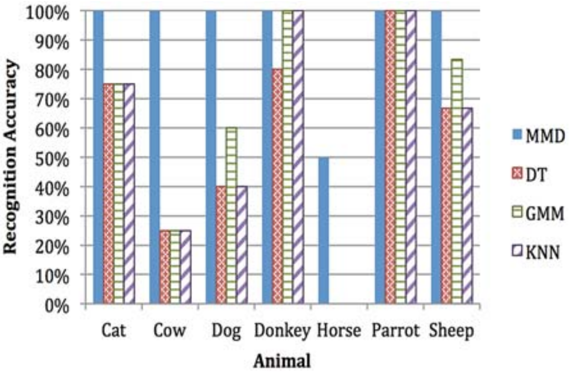


Fig. 9. Recognition accuracy using different classification techniques.

5.2. Energy Efficiency of the Proposed Scheme for Sensor Node

In the experiment we make an attempt to evaluate the efficiency of new object detection, feature extraction, and object classification that performed during local recognition task within each member sensors. We investigated the energy consumption of our scheme using AVRORA, which is an instruction-level sensor network simulator that emulates the real implementation of different wireless sensor nodes, such as (MICA2dot, MICA2, MICAz) [20].

The cost of the whole recognition scheme is depicted in Table 2. We assumed all test signals were sampled at 44,100 Hz samples per second. It is clearly noticeable from the table that the amount of time and energy consumed during object classification process is large compared to other steps. Such task involves performing several expensive computations compared to other steps. Specifically, the classification process involves computing couple multiplication operations that performed during measuring the distance between the target signature and the extracted features' vector. Nevertheless, the local recognition solution is very essential to reduce the extensive transmission of the entire raw audio signal which might question the lifetime of the communication system and the whole network service-availably.

Table 2. Evaluation of the recognition cost on MicaZ [14].

Measured Attribute	Clock cycles	Time (ms)	Energy (mJ)
Object detection	352	0.044	0.0009
Feature extraction	1527	0.19	0.004
Classification	6814	0.85	0.01
Whole scheme	8693	1.084	0.0149

The results of the notification cost to the CH are presented in Table 3. In this test, we simulated the notification task considering three different notification scenarios, which can be one of the following: a simple notification report, vector of features, or the whole acoustic signal. We assumed the adopted MICAz sensor is capable of achieving data transfer rates of 88,200 bytes/second, which computed by multiplying 16 bits/sample by 44100 samples/second. This operation will consume 29106 mJ of energy, which demonstrated that the proposed scheme could reduce the total energy consumption during the communication task in the network.

6. Discussion

Unlike previous works, in our proposed scheme, we guaranteed a good balance between algorithm performance, complexity, and resource limitation. It is shown from the results obtained from the AVRORA simulator that the selected set of feature extraction methods (RMS and ZCR) can easily fit on sensor nodes, which has very limited computational and power resources. Moreover, the local recognition scheme was capable to recognize objects with almost the same level of accuracy and with much less complexity cost compared to their works. We noted that the selected features were able to classify the target object to a single animal type with lower accuracy compared to these works. We think that one of the main parameters influencing the recognition performance is the huge overlapping between the animals and the small dataset that used for validating recognition accuracy. Nevertheless, these features can effectively achieve a good tradeoff between cost and performance, and hence, prolong the network lifetime.

Table 3. Evaluation of the notification cost on MicaZ [14].

Measured Attribute	Time (ms)	Energy (mJ)
Transmit detection notificition (1 byte)	0.01	0.33
Transmit 2D vector (2 bytes per feature)	0.04	1.32
Transmit raw signal	0.02	0.66

7. Conclusions

This paper presented our approach for efficient low-power acoustic sensing in WASN, focusing on habitat monitoring applications. The proposed scheme is intended to recognize a particular object of interest using the extracted acoustic signature from the received audio signal using lightweight time domain features. The feature selection stage has shown that RMS and ZCR are the most suitable set of features to describe the target animal, while adding some commonly used frequency domain features will not contribute in enhancing the recognition accuracy. The scheme also supports the application of object classification based on the MMD classifier suitable for use in resource-constrained devices. We presented the performance analysis for a low energy acoustic-based object recognition scheme. The investigation focused on measuring the scheme performance during the recognition and notification tasks for a single target. Our results demonstrate this approach is capable to achieve important energy saving that helps to extend the network lifetime. In addition, the scheme was capable to recognize the target with high accuracy. As future work, we are investigating the implementation of a localization technique in this scheme.

Acknowledgements

The research work developed in this paper was funded by King Abdulaziz City for Science and Technology (KACST) under Grant No. (1-17-02-001-0009). We would like to express our deep thanks and gratitude for KACST for facilitating the various requirements during the project.

References

[1]. A. Lerch, An Introduction to Audio Content Analysis: Applications in Signal Processing and Music Informatics, *John Wiley & Sons*, 2012.

[2]. J. Colonna, M. Cristo, M. Salvatierra, E. Nakamura, An incremental technique for real-time bioacoustic signal segmentation, *Expert Systems with Applications*, Vol. 42, No. 21, 2015, pp. 7367-7374.

[3]. J. Xie, M. Towsey, P. Eichinski, J. Zhang, P. Roe, Acoustic Feature Extraction Using Perceptual Wavelet Packet Decomposition for Frog Call Classification, in *Proceedings of the IEEE 11th International Conference on e-Science*, Munich, 2015, pp. 237-242.

[4]. C. L. T. Yuan, D. A Ramli, Frog Sound Identification System for Frog Species Recognition, in *Context-Aware Systems and Applications, Springer*, 2013, pp. 41-50.

[5]. J. Xie, M. Towsey, M. Zhu, J. Zhang, P. Roe, An intelligent system for estimating frog community calling activity and species richness, *Ecological Indicators*, Vol. 82, 2017, pp. 13-22.

[6]. J. Noda, C. Travieso, D. Sánchez-Rodríguez, Fusion of Linear and Mel Frequency Cepstral Coefficients for Automatic Classification of Reptiles, *Applied Sciences*, Vol. 7, No. 178, 2017, pp. 1-10.

[7]. J. Colonna, T. Peet, C. A. Ferreira, A. M. Jorge, E. F. Gomes, J. Gama, Automatic classification of anuran sounds using convolutional neural networks, in *Proceedings of the Ninth International Conference on Computer Science & Software Engineering, ACM,* 2016, pp. 73-78.

[8]. J. Noda, C. Travieso, D. Sánchez-Rodríguez, Methodology for automatic bioacoustic classification of anurans based on feature fusion, *Expert Systems with Applications*, Vol. 50, Issue C, 2016, pp. 100-106.

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

[10]. J. Xie, M. Towsey, L. Zhang, J. Zhang, P. Roe, Feature Extraction Based on Bandpass Filtering for Frog Call Classification, in *Proceedings of the International Conference on Image and Signal Processing (ICISP'16)*, 2016, pp. 231-239.

[11]. J. Colonna, E. Nakamura, O. Rosso, Feature Evaluation for Unsupervised Bioacoustic Signal Segmentation of Anuran Calls, *Expert Systems with Applications*, Vol. 106, 2018.

[12]. G. Ferrari, Sensor Networks: Where Theory Meets Practice, *Springer Science & Business Media*, Berlin, 2010, pp. 1-228.

[13]. H. Karl, A. Willig, Protocols and architectures for wireless sensor networks, *John Wiley & Sons,* 2007, pp. 1-53.

[14]. A. Algobail, A. Soudani, S. Alahmadi, Energy-aware Scheme for Animal Recognition in Wireless Acoustic Sensor Networks, in *Proceedings of the 7th International Conference on Sensor Networks (SENSORNETS'2018)*, 2018, pp. 31-38.

[15]. C. Huang, Y. Yang, D. Yang, Y. Chen, Frog classification using machine learning techniques, *Expert Systems with Applications, Elsevier*, Vol. 36, No. 2, 2009, pp. 3737-3743.

[16]. M. Rudrapatna, A. Sowmya, Feature Weighted Minimum Distance Classifier with Multi-class Confidence Estimation, in *Proceedings of the Australasian Joint Conference on Artificial Intelligence*, Berlin, Heidelberg, 2006, pp. 253-263.

[17]. J. Luque, D. Larios, E. Personal, J. Barbancho, C. León, Evaluation of MPEG-7-Based Audio Descriptors for Animal Voice Recognition over Wireless Acoustic Sensor Networks, *Sensors*, Vol. 16, No. 5, 2016, p. 717.

[18]. I. Akyildiz, M. Can Vuran, Wireless Sensor Networks, *John Wiley & Sons*, New York, NY, 2010.

[19]. Sound-ideas (https://www.sound-ideas.com/ Product/ 380/HD-—-Animals- Birds-Sound-Effects).

[20]. B. L. Titzer, D. K. Lee, J. Palsberg, Avrora: scalable sensor network simulation with precise timing, in *Proceedings of the IEEE Fourth International Symposium on Information Processing in Sensor Networks (IPSN)*, April 2005, pp. 477-482.