Feature Subset Selection for Automatically Classifying Anuran Calls Using Sensor Networks

Juan Gabriel Colonna*, Afonso D. Ribas*†, Eulanda M. dos Santos†, Eduardo F. Nakamura*†
†Institute of Computing, Federal University of Amazonas
Rodrigo Octavio Jordao, 3000, Setor Norte, 69077-000, Manaus, AM, Brazil
{juancolonna, emsantos, nakamura}@icomp.ufam.edu.br
*Computer Science Lab, Research and Technological Innovation Center (FUCAPI)
Danilo Areosa de Matos, 381, 69075-351, Manaus, AM, Brazil
{afonso.degmar, eduardo.nakamura}@fucapi.br

***Abstract*—Anurans (frogs or toads) are commonly used by biologists as early indicators of ecological stress. The reason is that anurans are closely related to the ecosystem. Although several sources of data may be used for monitoring these animals, anuran calls lead to a non-intrusive data acquisition strategy. Moreover, wireless sensor networks (WSNs) may be used for such a task, resulting in more accurate and autonomous system. However, it is essential save resources to extend the network lifetime. In this paper, we evaluate the impact of reducing data dimension for automatic classification of bioacoustic signals when a WSN is involved. Such a reduction is achieved through a wrapper-based feature subset selection strategy that uses genetic algorithm (GA). We use GA to find the subset of features that maximizes the cost-benefit ratio. In addition, we evaluate the impact of reducing the original feature space, when sampling frequencies are also reduced. Experimental results indicate that we can reduce the number of features, while increasing classification rates (even when smaller sampling frequencies of transmission are used).**

I. Introduction

Since 1980 researchers point out dramatic declines in amphibian populations around the world. There are several reasons for this phenomenon, such as habitat modification or destruction, exploitation of natural resources, pesticides, water pollution and increased radiation. Although the global biodiversity may be critically affected by these factors, there is neither consensus about how they affect amphibian populations nor which factor is the most determinant to their declines.

However, the state-of-the-art shows that amphibians are directly affected by environmental changes [1]. According to Carey et al. [2] there is a clear relationship between climate change and mortality in amphibian populations, since these animals are closely related to the ecosystem. Therefore, amphibians may be used as early indicators of ecological stress. Hence, amphibians monitoring systems, employing automatic classification methods, may help to estimate long-term changes in amphibian populations and, consequently, to determine their causes.

Anurans live on terrestrial and aquatic ecosystems and produce sounds (calls) with enough information to identify each specie [3]. In addition, the use of calls allows us to develop non-intrusive methods at lower cost, when compared to other types of data like images, for example.

Anuran classification, based on calls, can be done by using a Spectrum Analyzer and a computer [4]. However, the success of this strategy depends on human experts. Therefore, it is usually a slow and error-prone task. An alternative is the combination of automatic classification methods with Wireless Sensors Networks (WSNs). This second strategy is more resilient and less intrusive [5]–[7].

WSNs are usually composed of low cost nodes. Thus, a large amount of nodes may be spread all over the desired areas [8]. Nonetheless, the low cost hardware imposes restrictions, such as lower processing power and batteries with reduced lifetime [9]. Thus, an automatic classification system must cope with these restrictions. There are two general strategies to accomplish these requirements: (1) decreasing the sampling frequency (*fs*) [10] and (2) extracting a set of features [6], [7]. In the first, each audio is represented by a smaller number of samples. In the second, the objective is to select the most discriminant features in order to reduce the number of features and increase the recognition rate.

In this paper, we focus on the second strategy, which is traditionally divided into two categories: filter-based and wrapper-based approaches. The former searches for the optimal subset of features by using probabilistic dependence measurements. The latter takes into account classification rates of concrete classifiers for feature selection. Usually, wrapper-based methods provide the best results.

Although wrapper-based feature subset selection can be exhaustive, search algorithms might be used when a large initial feature space is involved, due to the exponential complexity of an exhaustive search. This complexity is 2^n, being n the number of original features. Several search algorithms may be applied to search for the best subset of features. Genetic algorithms are attractive since they allow the fairly easy implementation of feature selection tasks as optimization processes. Moreover, they frequently achieve high performance on feature subset selection problems.

Our objective is to find the smallest subset of features that represents anuran calls and, consequently, reduce costs while keeping high accurate classification rates. We apply genetic algorithm (GA) on two sets of features: the 12 Mel-Cepstral

Fourier Coefficients (MFCCs), frequently used in bioacoustic signal classification [6], [7], and 9 coefficients based on the Wavelet transform. In addition, we analyze the impact of feature subset selection when the sampling frequency *fs* is reduced. As a consequence, we focus on reducing sampling frequencies and number of features.

Based on these standpoints, our contributions are:

1) We show that it is possible to decrease the costs of processing and transmission, while keeping high classification rates through the use of GA to reduce the original feature space.
2) We evaluate whether or not classification rates are affected, when decreasing quantization and sampling frequencies, by performing simulation with three subsets of features found by GA.
3) We identify the best set of features for anuran classification by comparing MFCCs and Wavelet coefficients.
4) We conduct a comparative analysis among all features investigated in terms of computational cost.

The remainder of this paper is organized as follows. Section II presents an overview of the related work. Section III summarizes features investigated and their computational complexity. The parameters and experimental protocol employed are described in Section IV. The experiments and results are presented in Section V. Finally, conclusions and outlook are discussed in Section VII .

II. Related work

Automatic classification using WSN may be divided into two groups. In the first, audio samples are transmitted to a sink node which is responsible for the classification. By contrast, the second group contains methods that work by using the network nodes to extract and transmit only the features needed for the classification.

In this context, Hu et al. [5] and Taylor et al. [11] present an approach for classifying one anuran specie from the Kakadu National Park in Australia. In their proposals, first, all audio samples are collected and transmitted to a central node. Then, the central node generates a spectrogram. Finally, a C4.5 Decision Tree classifier makes the decision based on pixels power features. However, they did not investigate the possibility of reducing samples to accomplish WSN low cost requirements. In order to overcome such a drawback, Bulusu and Hu [10] present a partial sampling technique, which is based on minimizing the impact of energy cost related to the transmission of the whole audio samples. This technique sends to the central node only the smallest rate of samples needed for signal reconstruction. Thus, the amount of transmitted information is reduced.

Cai et al. [6] use a WSN to monitor birds' habitats in Brisbane. They compute differences among neighborhoods of frames in terms of 12 MFCCs. This strategy allows to model changes between frames. Then, an Artificial Neural Network (ANN) classifies the samples. It is worth noting that all 12 coefficients were used by the authors. Vaca-Castao and Rodriguez [7] also used MFCCs as features to classify birds and anurans through kNN (*k-Nearest Neighbors*) classifier. However, they applied PCA (*Principal Component Analysis*) for dimension reduction in feature space. The authors highlight that the dimension reduction step increased the computational cost of the whole process, because it requires additional processing resources. Besides, PCA does not take into account classification rates as objective to guide feature selection.

Yen and Fu [12] classify four anuran species by using the DWT (*Discrete Wavelet Transform*) to extract features. Their method is divided in five basic steps: (1) apply the band-pass filter; (2) apply a clustering algorithm using values of the spectrogram to identify patterns of repetition; (3) apply the WPT (*Wavelet Packet Transform*) to compute the energy of each node in the tree; (4) apply Fisher's Criterion for dimension reduction in feature space; and (5) classify all samples with MLP (*Multilayer Perceptron*). Despite the fact that this method achieved 71% of recognition rate, it requires two transforms, a dimensionality reduction strategy, a series of filters IIR (*Infinite Impulse Response*) and two different classification techniques, leading to a computational cost impractical for WSNs.

Table I summarizes the related work. Note that methods that do not use classification rates for guiding dimensionality reduction are most often applied. In this paper, we use GA to perform wrapper-based feature subset selection, i.e., the dimension reduction in feature space process is guided by recognition rate.

TABLE I
Summary of related work.

Reference	Animal	Feature	Classifier	Result
Hu et al. [5]	Bufo marinus	Spectrogram	C4.5	60%
Cai et al. [6]	14 birds	MFCCs	ANN	86%
Vaca-Castano and Rodriguez [7]	10 birds 20 anurans	MFCCs PCA	k-NN	86% 90%
Yen and Fu [12]	4 anurans	Wavelet Fisher	MLP	71%

In terms of WSN, GA has been applied to improve specific tasks, such as density control, routing, covering and scheduling [13]–[16]. The main objective of these tasks is to reduce energy consumption and increase the network lifetime.

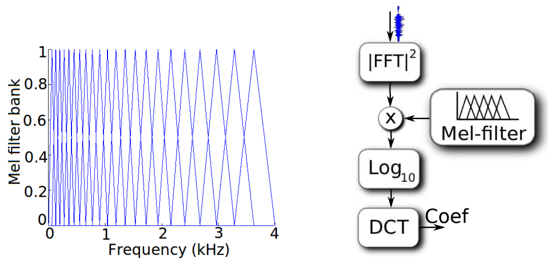
III. Features

A. Mel-Fourier Cepstral Coefficients (MFCCs)

The MFCCs are the most popular features, due to computational efficiency and noise robustness [17]. These coefficients allows us to reduce the amount of information needed to describe a signal, for both periodic and aperiodic signals. This technique performs an spectral analysis based on a bank of triangular filters logarithmically spaced in the frequency (Figure 1(a)). Figure 1(b) summarizes all steps involved in computing MFFCs.

In this paper, spacing between filters is defined in Mel scale [18], which is computed as:

$$f_{mel} = 1127 \ln\left(1 + \frac{f_{\mathrm{Hz}}}{700}\right). \quad (1)$$



(a) (b)

Fig. 1. (a) Banks of Mel filters and (b) Feature extraction with Fast Fourier Transform (FFT) and Discrete Cosine Transform (DCT).

Then, it is produced a M_r value for each filter by applying the filter bank on the signal spectrum. Once the M_r values are obtained, a logarithm is performed and a discrete cosine transform (DCT) is applied to produce the coefficients [19]:

$$mfcc_m = \frac{1}{R}\sum_{r=1}^{R} \log\left(M_r\right)\cos\left(\frac{2\pi}{R}\left(r+\frac{1}{2}\right)m\right), \quad (2)$$

in which m is the number of coefficients, R is the amount of filters. It is important to mention that the positive part of the spectrum is used to compute both equations.

We need to perform N multiplications to produce the output of M_r filters. Thus, there are R multiplications for each $mfcc_m$, while the cost of the Fast Fourier Transform is $N\log(N)$ [20]. Finally, the final cost of MFCCs is $N\log(N)+N+mR$ with complexity $O(N\log(N))$, where N is the number of spectrum frequency samples required, i.e., twelve coefficients are traditionally used [6], [7].

B. Discrete Wavelet Transform (DWT)

In real applications, signals are frequently obtained through sampling, which generates a set of samples in discrete time. Hence, a discrete transform [21], for instance DWT, must be used in this context. When dealing with DWT, a discrete-time signal $f(n)$ can be represented by two sets of coefficients, namely approximations (c_k, or average) and details ($d_{j,k}$, or difference). Following this definition, a signal is computed as:

$$f(t) = \sum_{k=-\infty}^{\infty} c_k\phi(t-k) + \sum_{k=-\infty}^{\infty}\sum_{j=0}^{\infty} d_{j,k}\psi(2^j t-k), \quad (3)$$

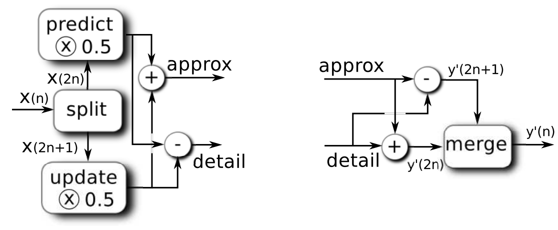
where $\phi(t)$ and $\psi(t)$ functions are known as father and mother respectively. We can extract a sequence of coefficients h and g by applying the follows equations:

$$\psi(t) = \sqrt{2}\sum_{n\epsilon\mathbb{Z}} h_n\phi(2t-n), \quad (4)$$

$$\phi(t) = \sqrt{2}\sum_{n\epsilon\mathbb{Z}} g_n\phi(2t-n). \quad (5)$$

Values h_n and g_n are used for a Z transform in order to become discrete filter coefficients. Thus, the signal convolution may be performed with two filters: a low-pass filter g_n and a high-pass filter h_n for approximation and details coefficients.

Actually, two operations are required for DWT: the signal convolution with two filters and a downsample by two. This is produced by a low cost scheme which is known as *Lifting Scheme* [21], as illustrated in Figure 2(a). It can be observed in this figure that the *predict* and *update* blocks multiply the input signal samples by the filter coefficients, while the block *split* divides samples into odd and even. Finally, Figure 2(b) shows the steps used to retrieve information from the coefficients. The block *merge* recomposes the signal at the end of the whole process.



(a) (b)

Fig. 2. (a) Wavelet Transform with Lifting Scheme. (b) Information retrieval by inverse transform.

DWT presents complexity as $O(L)$, in which L is the amount of samples of the incoming signal. Once the coefficients values produced by DWT are obtained, we consider the features described in the following.

The **Energy (E)** and **Power (Pw)** of the coefficients are defined as:

$$E = \sum_{n}^{L} [x_{(n)}w_{(n)}]^2, \quad (6)$$

$$Pw = \frac{1}{L}\sum_{n}^{L} |[x_{(n)}]|^2, \quad (7)$$

where $x_{(n)}$ is the signal and $w_{(n)}$ is a Hamming window. Energy was used by Vaca-Castao and Rodriguez [7] for bioacoustics signals segmentation process. These features present L multiplications as approximate cost, leading to complexity equals to $O(L)$ [22].

The **Zero-Crossing Rate (ZC)** indicates the frequency of transitions of coefficients values between positive and negative values. This feature provides a signal-length estimation:

$$ZC = \frac{1}{2}\sum_{n=0}^{L-1} |\text{tsgn}(x_n) - \text{tsgn}(x_{n+1})|, \quad (8)$$

$$tsgn(x_n) = \begin{cases} +1 & x_n \geq \eta \\ -1 & x_n \leq -\eta \end{cases}, \quad (9)$$

where x_n denotes the amplitude and η the threshold amplitude. If the length of the signal is L, then L comparisons are necessary, resulting in a $O(L)$ cost.

The **Pitch (P)** is the difference between two peaks of autocorrelation [23], [24]. This feature is assumed to be a similarity measure in the waveform of the coefficients and is related to the signal fundamental frequency:

$$R_{xx} = \frac{1}{L}\sum_{r=1}^{L}(x(n)x(n+r)), \tag{10}$$

$$P = max_1(R_{xx}) - max_2(R_{xx}), \tag{11}$$

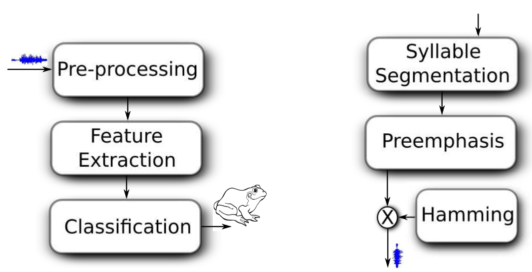
where L is the amount of samples of calls. In terms of complexity, $2L-1$ operations are required to obtain R_{xx} and to find the position of two max values: max_1 and max_2. Thus, the cost is $2L-1+L=3L-1$,i.e., $O(L)$.

IV. Parameter Settings on Experiments

Anuran calls classification is traditionally composed of three main stages: pre-processing, feature extraction and classification (Figure 3(a)). Each stage focus on different tasks.

A. *Pre-processing*

A typical example of anuran call is depicted in Figure 4. This figure shows a sample of the *Adenomera andreae* call. In this work, the signals are first adjusted and normalized during the pre-processing step. Initially, the signals are segmented into smaller units called syllables, the pre-emphasis and windowing (Figure 3(b)), as was proposed by Huang et al. [3].



(a) (b)

Fig. 3. (a) Anuran classification stages. (b) Pre-processing steps.

The syllable segmentation algorithm has the following steps:

1) Search for the largest amplitude $S_{(t)}$ in the signal;
2) If $S_{(t)}$ is smaller than the threshold α, go to `step 5`;
3) Extract ε milliseconds to the right and to the left of the peak, $S_{(t-\varepsilon)} \leq$ syllable $\leq S_{(t+\varepsilon)}$ (this defines a syllable);
4) Extract the feature vector for the syllable and remove the values of the original signal S within the interval $t \pm \varepsilon$, go to step 1;
5) End.

The syllable length (Figure 4), denoted by ε is set to 200 ms in this work. This value guarantees that we acquire complete syllables for all species that we are considering in this work.

After the pre-processing step, each signal enters into the feature extraction block. In this block, features described in Section III-B are extracted to represent each syllable, and

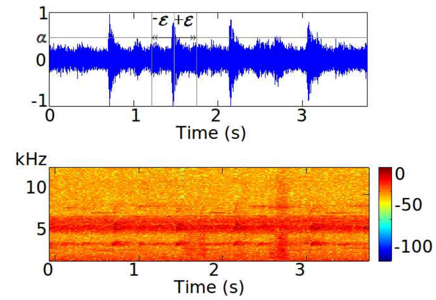
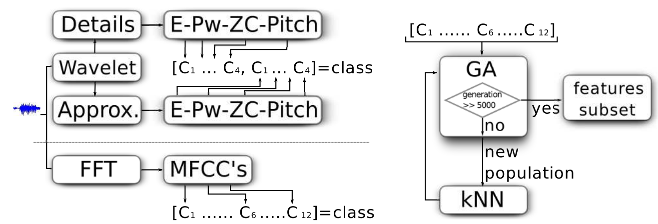


Fig. 4. Audio sample (wave form and spectrogram) for the *Adenomera andreae*.

are arranged sequentially in a feature vector. These vectors sequences are used to compose the database used by the classifier (Figure 5(a)). Then, GA is applied to seek for the best subset of feature. Hence, the optimization process is performed outside of the WSN and the classifier recognition rate is used as fitness function. Once the best subset of features has been selected, the classification is conducted by kNN (Figure 5(b)).



(a) (b)

Fig. 5. (a) Feature extraction. (b) Feature selection and classification.

B. *Dataset*

The dataset used in our experiments comprises nine species of anurans, whose audio streams were collected in the anuran habitat, under real noise conditions. The sound samples of the following species were collected on the campus of the Federal University of Amazonas in Manaus, Brazil: ***(a)*** *Adenomera andreae*, ***(d)*** *Hypsiboas cinerascens*, ***(e)*** *Leptodactylus fuscus*, ***(f)*** *Osteocephalus oophagus* and ***(g)*** *Rhinella granulosa*. The sounds of the species ***(b)*** *Ameerega trivittata*, ***(c)*** *Hyla minuta*, ***(h)*** *Scinax ruber* and ***(i)*** *Hylaedactylus* were extracted from three different regions: Mata Atlntica, Brazil [25], Bolivia [26] and French Guiana [27]. Our objective is to show that the selected features are general enough to be widely used for anuran classification.

Our samples were stored in *wav* format at 44.1 kHz and 32 bit per sample, which allows us to analyze signals up to 22.05 kHz. Moreover, we have consider three values to the α syllable segmentation variable: 40%, 50% and 60% (Section IV-A). Then, we generated three databases with different sets of feature for each α value, namely MFCCs and DWT based on Haar Wavelet and Daubechies. Thus, we carried out experiments with nine databases. Table II shows the relation between α values and the number of samples per species used to compose the datasets investigated in our experiments.

TABLE II
NUMBER OF SYLLABLES IN THE DATABASES USED FOR THE EXPERIMENTS.

Species	**Individuals**	**Syllables**		
		$\alpha = 0.4$	$\alpha = 0.5$	$\alpha = 0.6$
a	8	673	525	443
b	5	614	545	330
c	11	283	244	213
d	2	352	312	220
e	4	304	249	229
f	4	151	120	82
g	3	1147	1103	964
h	4	221	183	161
i	8	3354	3172	2898
Total	49	7099	6453	5540

The first feature set is composed by 12 MFCCs. The feature vector using the MFCCs has the following form:

$$[coef_1, coef_2, coef_3, \cdots, coef_{12}] \rightarrow \text{class}. \qquad (12)$$

The second and the third feature sets, which are based on Haar Wavelet and Daubechies (Db), are extracted as follows:

1) get the Pitch (P) of a syllable;
2) apply DWT to obtain the approximation and the details coefficients;
   a) use the coefficients to compute:
      - energy (W^d_E);
      - power (W^d_{Pw});
      - difference between the two autocorrelation max values (W^d_P); and
      - Zero-Crossing Rate (W^d_{ZC}).

The DWT-based feature vectors provided to the classifier have the following form:

$$[P, W^d_E, W^d_{Pw}, W^d_P, W^d_{ZC}, W^a_E, W^a_{Pw}, W^a_P, W^a_{ZC}] \rightarrow \text{class} \qquad (13)$$

where W is used to indicate that these characteristics were extracted using DWT instead of FFT, d and a denote whether they have been calculated on the approximation and details coefficients respectively, P represents the Pitch and class is the specie.

V. RESULTS

Our experiments are divided in three groups. In the first, all features are applied and the three feature sets are compared to each other. In the second, GA is used for feature selection. Finally, the third is performed by reducing the sampling frequencies.

The recognition rate, used as evaluation metric, is defined as the reason between the number of correct classifications (true positives) and the total number of samples. In addition, leave-one-out cross-validation (LOOCV) was employed for performance measurement.

A. Feature Sets Comparison

We begin our analysis by taking into account all features available in each feature set. First, however, kNN k parameter must be experimentally defined. We have tested different values to fine-tuning this parameter. Figure 6 shows the error rates achieved by kNN with different k values, over samples contained in datasets generated by using $\alpha = 0.6$ and the three feature sets.

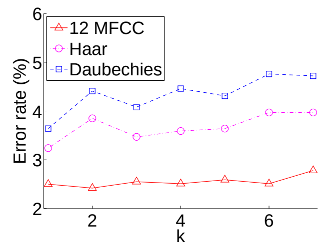


Fig. 6. Classification error rates achieved by kNN by varying k value using LOOCV on databases generated with $\alpha = 0.6$.

Table III shows the relationship between kNN recognition rate and the amplitude segmentation value α for all three feature sets. The recognition rate values reported in this table were obtained as the mean of the replications performed through LOOCV. The best recognition rate achieved for each database and α value is shown, as well as the classifier k value, between “()”. The feature sets (first column) are shown in descending order, according to the approximate computational cost calculated in Section III.

TABLE III
RECOGNITION RATES VS. α VALUES AND FEATURE SETS USING LOOCV.

Features	**kNN**		
	$\alpha = 0.4$	$\alpha = 0.5$	$\alpha = 0.6$
Wavelet features Daubechies transform	93.94%(3)	95.54%(1)	96.36%(1)
Wavelet features Haar transform	94.63%(1)	96.21%(1)	96.76%(1)
MFCCs	97.39%(1)	97.60%(1)	97.58%(2)

These experiments show that the most successful feature set was composed by the MFCCs. The solutions were tested using the *Wilcoxon* test, by comparing LOOCV values. The confidence level was 95%. This statistical test indicates that the recognition rates achieved by kNN were significantly higher using the 12 MFCCs than using DWT-based features. Another important point to mention is that features based on DWT can capture information similar to the MFCCs, with slightly higher cost. Finally, MFCCs are less prone to noise, when compared to the DWT-based feature sets, since kNN recognition rates are highly stable when varying threshold α. This indicates low correlation with noise levels.

B. Feature Subset Selection

We continue our experimental study by using GA to select the best subset of features for each feature set. To determine the best combination of features, it is essential to maximize the recognition rate and minimize the processing cost. This can be accomplished by testing all combinations of features and by evaluating the impact on classification rate. The combinatorial approach has exponential cost that increases with the amount of features. Therefore, features subset selection is not exponentially addressed, due to the high cost involved. In addition, similar feature may lead to wrong classification, being unnecessary to sort them for some combinations [28].

In our experiments, we use a meta-heuristics based on GA to conduct three wrapper-based feature selection processes [28], [29] to find the best combination for each feature set (MFCCs, Daubechies DWT, and Haar DWT). GA is guided by kNN recognition rates as objective function during a fixed maximum number of generations. The selection of features is applied in the context of GA based on binary vectors. Each individual, called chromosome, is represented by a binary vector with a size n, since the initial sets of features are composed by n features. Each bit value determines whether a feature is active (1) or not (0). For each selection process, a population with a fixed number of chromosomes is randomly created initially, i.e. a random population of candidate feature subsets. Thus, at each generation step, the algorithm computes fitness, based on kNN recognition rates, of each candidate in the current population. The population is evolved through the operations of crossover, mutation and elitism.

Experiments were carried out to define genetic parameters. The same parameters were used for all three feature sets and were defined as: probability of elitism $= 0.1$; probability of crossover $= 0.5$; probability of mutation $= 0.4$; the number of individuals at any generation was 20 and the maximum number of generations was 5.000. Even though, we observed in all three selection processes that the algorithm reached its convergence around 200 generations. It also important do mention that 10-fold cross-validation was used to conduct the feature subset selection processes.

Table IV summarizes the mean kNN recognition rates and feature subset sizes achieved by GA. We also report the original feature sets sizes and the recognition rates obtained by kNN, when employing all original features. These results were computed over samples contained in the database generated with $\alpha = 0.5$.

TABLE IV
GENETIC SEARCH RESULTS

Initial Number of Features	**Recognition Rate before GA**	**Number of Features After GA**	**Recognition Rate After GA**
9 features with Db	95.54%	4 features	95.73%
9 features with Haar	96.21%	7 features	96.25%
12 MFCCs	97.60%	8 MFCCs	97.95%

These experiments show that:

- GA successfully reduced the number of features for all three feature sets investigated. In terms of MFCCs, the initial set of 12 features was decreased to 8 features, leading to a 33% of information reduction, which must be transmitted by the sensor nodes. The selected coefficients were: 1, 2, 3, 4, 5, 6, 7 and 11. In addition to the lower cost of information transmission, the recognition error rate was 0.35% higher than by using all 12 initial features.
- In Haar DWT-based feature set, GA discarded 2 features from the original feature set composed by 9 features, namely W_P^a and W_{Pw}^d (Section IV-B). This means an information reduction of 22%. Moreover, recognition rate increased 0.04%, indicating that the removed features were not relevant for the classification task.
- The highest feature set reduction was observed on Daubechies DWT-based feature set. An information reduction of 55% was obtained (only 4 features were selected from the 9 original ones). The selected features are: P, W_P^d, W_P^a, W_E^a. Again, a recognition rate improvement is obtained after feature selection. There is a 0.19 % of recognition rate improvement.
- In terms of feature sets comparison. MFCCs lead to the highest recognition rates, while the lowest recognition rates were achieved by using Daubechies DWT-based features.

VI. Case Study

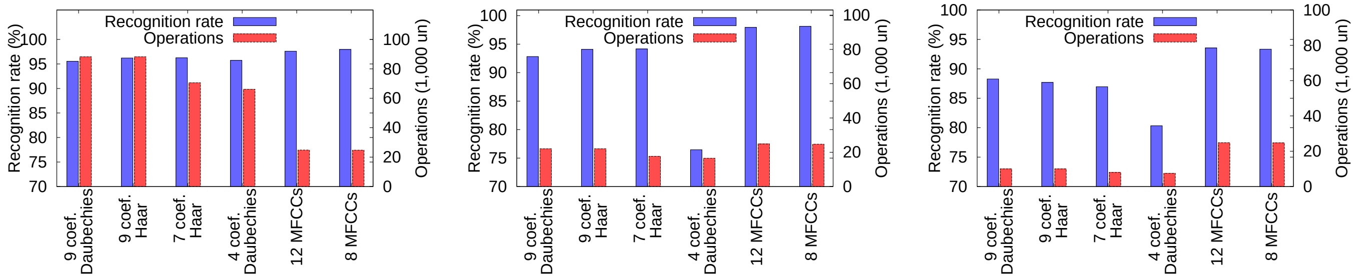
The simplest way to reduce information is by decreasing the signals sample rates, as discussed in the introduction. However, our preliminary experiments used audio samples represented by 32 bit at 44.1 kHz. These rates are not usual in the context of a WSN. In most of low-cost hardware, samples are represented by 8 bit and sampling rates close to 8 kHz. Actually, sensors have an analog-to-digital converter module.

Motivated by both requirements of WSN, i.e. to reduce the number of features and sampling frequencies, we simulated a real situation of using a WSN in this last series of experiments. Audio samples were uniformly quantized into 256 levels (8 bit), decreasing the sampling rate four times (up to 11.0 kHz) and eight times (up to 5.5 kHz) and leading to 75% and 87% of information reduction, respectively. We have converted the samples into integers in a range between [-128, 127], thus allowing quantization noise (N_q). The Signal-to-Quantization-Noise Ratio (SQNR) is 49.92 dB [30].

Table V shows the impact on recognition rate by using 8 bit with 5.5 kHz and 11.0 kHz to represent each sample in database with $\alpha = 0.5$.

TABLE V
COMPARING THE RESULTS OF CLASSIFICATION USING 32 BIT AND 8 BIT PER SAMPLE WITH 44.1 KHZ, 11.0 KHZ AND 5.5 KHZ PER SAMPLE.

Features	**Classification, kNN (LOOCV)**		
	32 bits 44.1 kHz	8 bits 11 kHz	8 bits 5.5 kHz
9 original Db features	95.54%	92.83%	88.27%
4 GA selected Db features	95.73%	76.46%	80.33%
9 original Haar features	96.21%	94.09%	87.69%
7 GA selected Haar features	96.25%	94.16%	86.96%
12 original MFCCs	97.60%	97.58%	93.54%
8 GA selected MFCCs	97.95%	98.13%	93.34%



(a) (b) (c)

Fig. 7. Trade-off between recognition rate and approximate number of operations with: (a) 44.1 kHz , (b) 11.0 kHz and (c) 5.5 kHz.

It is not surprising that our results in Table V indicate that sampling frequency reduction lead to a decrease in the recognition rate achieved with all three feature sets for both original and optimized feature sets. Noteworthy, the information reduction by four and eight times per sample, do not affect the performance provided by kNN using the original MFCCs-based feature set significantly. Still more interesting is that the same behavior was observed for its optimized feature subset. This comparison allow us to conclude that we may decrease the amount of information, which must be transmitted by the WSN using both strategies: by reducing the number of features and the sampling frequencies. Besides, our results indicate that MFCCs are more immune to quantization noise than the other investigated features.

Finally, we have shown in Section III that $N\log(N)+N+mR$ is the final cost of MFCCs, where N is the number of FFT points, m is the number of Mel-coefficients and R is the number of filters. Thus, given $N = 2048$ spectrum points, 12 Mel-coefficients, and by applying 22 filters, the approximate cost is 24,840 operations. To support our mathematical analysis of cost, we did some tests to measure computational time needed to compute each group of features using MATLAB and an AMD Athlon 64x2 computer.

Figure 7 illustrates the relationship between cost, amount of bits variation and sampling frequency. We show frequency of 44.1 kHz and 32 bit in Figure 7(a), frequency of 11.0 kHz and 8 bit in Figure 7(b), and frequency of 5.5 kHz and 8 bit in Figure 7(c). The horizontal axis, the vertical left axis and the vertical right axis, represent the feature set, the recognition rate for each feature combination and the cost, respectively. These values were measured as the approximate number of operations required for each combination of features.

The best trade-off occurs when the difference between the curves is as largest as possible, maximizing the Cost-Recognition ratio. We observe that in addition to keep the classifier with high recognition rate, even with sampling frequency reduction, MFCCs present constant cost. This means that these features do not depend on the sampling rate (*fs*), regardless of hardware.

The cost related to features based on wavelet transform depends on the amount of signal samples. In Figure 7, we observe such a dependence of *fs* since the lower the sampling rate, the lower the cost. The recognition rates are also affected by the *fs*. These rates were less than 90% in the worst scenario.

VII. Conclusion and Outlook

In this paper, we have employed GA to reduce the amount of information for classifying anuran calls, through wrapper-based subset selection. The objective was to decrease processing and transmission costs in WSN, while maintaining high recognition rates.

The reduction of feature sets, obtained with GA, indicates which features are most important to the classifier. In addition, this reduction allows us to save energy in the pre-processing step and the data communication, increasing the network lifetime. By using the sets of features returned by GA, information required in reduced in 33%, with 8 MFCCs; 55%, with 4 Db Wavelet transform coefficients; and 22%, with 7 Haar transform coefficients. The resulting coefficients MFCCs were: 1 to 7 and 11.

By applying the Wilcoxon test, we conclude that MFCCs are better. With 98.13% of accuracy, in the subset returned by the GA (eight MFCCs), we can say that these have a great trade-off between cost and benefit. As described earlier, we identified several advantages of MFFCs compared to other features:

- A low relationship with the α variable, which indicates greater immunity to environmental noise;
- They are not affected by quantization, indicating immunity to this type of noise;

Future work includes the addition of other species to determine the differentiation capacity of MFCCs. Subsets of features will be implemented in a WSN to assess the impact of these reductions on the energy consumption. We are also characterizing the dependence of MFFCs with the variation of amount of FFT points, as an alternative to reduce processing costs.

Acknowledgments

This work is partially financed by the National Council for Scientific and Technological Development (CNPq) and the Amazon State Research Foundation (FAPEAM), through the grants 575808/2008-0 (Revelar Project - CNPq) and 2210.UNI175.3532.03022011 (Anura Project - PRONEX

023/2009). We also thank professors Marcelo Gordo and Marco Cristo for the kind consulting.

REFERENCES

[1] J. P. Collins and A. Storfer, “Global amphibian declines: sorting the hypotheses,” *Diversity and Distributions*, vol. 9, pp. 89–98, 2003.

[2] C. Carey, W. R. Heyer, J. Wilkinson, R. Alford, J. W. Arntzen, T. Halliday, L. Hungerford, K. R. Lips, E. M. Middleton, S. Orchard, and S. Rand, “Amphibian declines and environmental change: Use of remote-sensing data to identify environmental correlates,” *Conservation Biology*, vol. 15, no. 4, 2001.

[3] C. Huang, Y. Yang, D. Yang, and Y. Chen, “Frog classification using machine learning techniques,” *Expert Systems with Applications*, vol. 36, no. 2, pp. 3737–3743, 2009.

[4] K. Riede, “Monitoring biodiversity: analysis of amazonian rainforest sounds,” *Ambio*, vol. 22, no. 8, pp. 546–548, 1993.

[5] W. Hu, V. N. Tran, N. Bulusu, C. T. Chou, S. Jha, and A. Taylor, “The design and evaluation of a hybrid sensor network for cane-toad monitoring,” in *4th International Symposium on Information Processing in Sensor Networks.* IEEE, Apr. 2005, pp. 503–508.

[6] J. Cai, D. Ee, B. Pham, P. Roe, and J. Zhang, “Sensor network for the monitoring of ecosystem: Bird species recognition,” in *3rd International Conference on Intelligent Sensors, Sensor Networks and Information*. IEEE, Dec. 2007, pp. 293–298.

[7] G. Vaca-Castaño and D. Rodriguez, “Using syllabic mel cepstrum features and k-nearest neighbors to identify anurans and birds species,” in *Signal Processing Systems (SIPS), 2010 IEEE Workshop on*, 2010, pp. 466–471.

[8] I. F. Akyildiz, W. Su, Y. Sankarasubramaniam, and E. Cayirci, “Wireless sensor networks: a survey,” *Computer Networks*, vol. 38, no. 4, pp. 393–422, 2002.

[9] G. Xing, X. Wang, Y. Zhang, C. Lu, R. Pless, and C. Gill, “Integrated coverage and connectivity configuration for energy conservation in sensor networks,” *ACM Transactions on Sensor Networks*, vol. 1, no. 1, pp. 36–72, Aug. 2005.

[10] N. Bulusu and W. Hu, “Lightweight acoustic classification for cane-toad monitoring,” in *42nd Asilomar Conference on Signals, Systems and Computers*. IEEE, Oct. 2008, pp. 1601–1605.

[11] A. Taylor, G. Watson, G. Grigg, and H. McCallum, “Monitoring frog communities: An application of machine learning,” in *8th Innovative Applications of Artificial Conference*, 1996.

[12] G. Yen and Q. Fu, “Automatic frog call monitoring system: a machine learning approach,” in *Proceedings of SPIE*, vol. 4739, no. 2. SPIE, 2002, pp. 188–199.

[13] I. de Andrade, T. de Oliveira Januario, G. Pappa, and G. Mateus, “An evolutionary algorithm to the density control, coverage and routing multi-period problem in wireless sensor networks,” in *Evolutionary Computation (CEC), 2010 IEEE Congress on*, july 2010, pp. 1–8.

[14] A. Konstantinidis, Z. Qingfu, and Y. Kun, “A subproblem-dependent heuristic in moea/d for the deployment and power assignment problem in wireless sensor networks,” in *Evolutionary Computation, 2009. CEC ’09. IEEE Congress on*, may 2009, pp. 2740 –2747.

[15] L. Ying, H. Xiao-min, Z. Jun, L. Ou, and L. Hai-lin, “Optimal node scheduling for the lifetime maximization of two-tier wireless sensor networks,” in *Evolutionary Computation (CEC), 2010 IEEE Congress on*, july 2010, pp. 1–8.

[16] C. Hui and Y. Shengxiang, “Genetic algorithms with elitism-based immigrants for dynamic shortest path problem in mobile ad hoc networks,” in *Evolutionary Computation, 2009. CEC ’09. IEEE Congress on*, may 2009, pp. 3135–3140.

[17] S. Davis and P. Mermelstein, “Comparison of parametric representations for monosyllabic word recognition in continuously spoken sentences,” *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 28, no. 4, pp. 357–366, Aug. 1980.

[18] M. Cowling, “Comparison of techniques for environmental sound recognition,” *Pattern Recognition Letters*, vol. 24, no. 15, pp. 2895–2907, Nov. 2003.

[19] L. Rabiner and R. Schafer, “Introduction to digital speech processing,” *Foundations and Trends in Signal Processing*, vol. 1, pp. 1–194, 2007.

[20] J. W. Cooley and J. W. Tukey, “An algorithm for the machine calculation of complex Fourier series,” *Mathematics of Computation*, vol. 19, no. 90, pp. 297–297, May 1965.

[21] S. Rein and M. Reisslein, “Low-memory wavelet transforms for wireless sensor networks: A tutorial,” *IEEE Communications Surveys & Tutorials*, vol. 13, no. 2, pp. 291–307, 2011.

[22] J. Deller, J. Hansen, and J. Proakis, *Discrete-time processing of speech signals*. IEEE, 1993, vol. 1.

[23] G. Peeters, “A large set of audio features for sound description (similarity and classification) in the cuidado project,” IRCAM, Tech. Rep., 2004.

[24] C. Plack, A. Oxenham, and R. Fay, *Pitch: Neural Coding and Perception*. Springer, 2005.

[25] C. Haddad, “Guia sonoro dos anfíbios anuros da mata atlântica.” 2005.

[26] R. Márquez, I. Riva, B. Matheu, and E. Matheu, “Sounds of frogs and toads of bolivia.” 2002.

[27] P. Marty, C.and Gaucher, “Sound guide to the tailless amphibians of french guiana.” 1999.

[28] K. Deb, *Multi-objective optimization using evolutionary algorithms*, volume 16 ed. John Wiley & Sons, 2001.

[29] M. Raymer, W. Punch, E. Goodman, L. Kuhn, and A. Jain, “Dimensionality reduction using genetic algorithms,” *Evolutionary Computation, IEEE Transactions on*, vol. 4, no. 2, pp. 164 –171, 2000.

[30] A. V. Oppenheim, R. W. Schafer, and J. R. Buck, *Discrete-time signal processing*, 2nd ed. Prentice-Hall, Inc., 1999.