

# Compressive Sensing for Efficiently Collecting Wildlife Sounds with Wireless Sensor Networks

Javier J. M. Diaz\*, Juan G. Colonna<sup>†</sup>, Rodrigo B. Soares\*  
Carlos M.S. Figueiredo<sup>‡</sup>, Eduardo F. Nakamura<sup>‡</sup>

\*Federal University of Minas Gerais (UFMG)  
{jvrdiaz, rodrigoborges}@dcc.ufmg.br

<sup>†</sup>Federal University of Amazonas (UFAM)  
{juancolonna}@gmail.com

<sup>‡</sup>Analysis, Research and Technological Innovation Center (FUCAPI)  
{mauricio.figueiredo, eduardo.nakamura}@fucapi.br

**Abstract**—Wildlife sounds provide relevant information for non-intrusive environmental monitoring when Wireless Sensor Networks (WSNs) are used. Thus, collecting such audio data, while maximizing the network lifetime, is a key challenge for WSNs. In this work, we propose a methodology that applies Compressive Sensing (CS) aiming at collecting as little data as possible to allow the signal reconstruction, so that the reconstructed signal is still representative. The key issue is to determine a sparse base that best represents the audio information used for identifying the target species. As a proof-of-concept, we focus on anuran (frogs and toads) calls, but the methodology can be applied for other animal families and species. The reason for that choice is that long-term anuran monitoring has been used by biologists as an early indicator for ecological stress. By using real wild anuran calls, we show that 98% classification rate can be achieved by using as little as 10% of the original data. We also use simulation to evaluate the impact of our solution on the network performance (energy consumption, delivery rate, and network delay).

**Keywords**—compressive sensing; sensor network; anuran classification;

## I. INTRODUCTION

Wireless Sensor Networks (WSN) are increasingly being used for environmental monitoring [1]. Through the monitoring of species, we can move towards understanding and preserving a target ecosystem [2]. However, due to the limited capacity of the batteries, there are still lifetime limits in this type of network [3], [4].

Wildlife sounds usually provide enough data to classify and to monitor the fauna. In this context, anurans (frogs and toads) have been used by biologists as early indicators of ecological stress. The reason is that these animals are closely connected to the environment, providing information about aquatic and terrestrial ecosystems [5]. In addition, anuran calls can be used to classify the species within a given site. However, any audio data collected from the environment and wildlife may represent a heavy communication load

for WSNs. Thus, Compressive Sensing (CS) techniques can be used for reducing the amount of data being gathered by means of sampling techniques and a posteriori signal reconstruction. For this, the signal must be representable in a sparse base [6], [7]. In WSNs, the number of samples is directly related to the network energy consumption, as well as the amount of packets that are sent over the network [8], [9].

Previous efforts for automatic animal classification, based on audio streams, have pointed out some features and classifiers for a successful task [10]–[12]. The list of features, in such efforts, includes *Spectral Centroid* (S), *Signal Bandwidth* (B), *Zero-crossing Rate* (ZC), *Hybrid Spectral-Entropy* (H) and *Mel Fourier Cepstral Coefficient* (MFCCs). Classifiers often adopted are: *k-Nearest Neighbor* (kNN) and *Support Vector Machine* (SVM). These strategies do achieve significant results for different animals, especially when using noise-free audio samples. However, there is neither a standard method, nor a definitive approach for anuran monitoring and sensor networks.

In this work, we collect the anuran calls and send the audio streams towards the sink node, where the audio streams are processed to identify the anuran specie of each call. In this context, our major goal is to determine the set of features that can be used to correctly identify the anuran species finding a good tradeoff between computational cost and the success rate. The proposed methodology can be applied for different wildlife sound-based monitoring. Anurans are used as a proof-of-concept, due to their importance in current ecosystems' monitoring.

By using a mixture of anuran classification methods [10], [11], we show how we can reduce up to 90% the number of samples collected and sent within the WSN without compromising the results of classification. Through the analysis of fundamental frequencies of each species, we constructed

a sparse base that ensures the reconstruction of the audio using CS techniques. Moreover, the impact of the technique on a WSN is evaluated by simulation.

The rest of this paper is presented as follows. Section 2 discusses the theoretical foundations and related work. Section 3 explains the methodology used in the experiments of CS. Section 4 shows the simulation scenario adopted for representing a WSN. In Section 5, we present our experiments, complemented by network simulations. Finally, Section 6 presents our conclusions and future work.

## II. BACKGROUND AND RELATED WORKS

### A. Compressive Sensing

The Nysquist theorem states that a sampling rate of at least twice the maximum frequency present in the signal is necessary for recovering it without aliasing nor information loss [6]. By using Compressive Sensing (CS) techniques, we can reconstruct signals with a much smaller number of samples than those required by the Nysquist theorem [7], [13], [14].

Compressive Sensing (CS) can be applied whenever the signal to be reconstructed has two basic properties: sparsity and linear independence (or incoherence) [6]. Once granted the sparsity and linear independence of the signal, it is necessary to generate a base  $\Psi$  with independent components of the signal to use it during the reconstruction [15].

The CS application in image processing is very common [7], [15]. In [7], reconstructions of Mondrian works of art are made using Bayesian Compressive Sensing (BCP) methods. The work in [7] shows a 89% improvement in the reconstruction time, keeping an error of up to 16% in relation to normal reconstruction techniques.

Another example of application of this technique is its use in magnetic resonance equipment. These devices have a high cost of sampling (large number of samples) and the images they generate can be represented in sparse bases [16]. The quality of the CS reconstruction is needed, especially in medical applications [17]. Thanks to the application of CS, the sampling time of this kind of applications can be reduced from hours to a few seconds [16], [17].

Candes and Waking [6] and Carin et al. [18] apply CS for reconstructing photographs and for image segmentation – for example, a static part (background) of a movable part (foreground), by capturing of consecutive images. Moreover, [6] shows how to apply the same methodology to audio signals randomly generated using partial sampling, followed by a reconstruction via minimization  $\ell_1$ . The methodology proposed in [6] is very similar to the one used in this work. However, instead of using randomly generated signals, this paper uses anuran calls captured in their habitat, subject to the real noises.

### B. Anuran classification

Due to the intimate relationship with their habitat, the anuran are considered as a metric for evaluation of environmental impacts in a certain region [5]. However, finding and classifying anuran directly in its habitat is not trivial.

The classification of anuran can be done manually [19], by the analysis of the recordings and its spectra by specialists, or automatically, using WSN as support for the capture and data processing [10], [11], [20].

When done manually, the quality of the anuran classification is related to the experience of the specialist who performs it, it can become a slow, prone to failure and intrusive process. Although automatic techniques are faster and more accurate than the manual, the amount of information used for classification generates a performance impact on the WSN, as will be shown later.

The classification results using automatic ways range from 50% in [21], 82.6% in [20] and 99.3% in [11]. The big difference between the results is mainly due to the information used as base for classification. [21] uses the intensity of the pixels from the recordings spectrogram as a reference for the classification.

Once the intensity of each pixel and its neighbors are evaluated, it was used a C4.5 decision tree to do the classification. Moreover, [21] method causes an excessive battery consumption, because it does not use any kind of compression on the audio samples.

To reduce the impact of energy and storage cost, Bulusu and Hu [9] adopted a partial sampling technique. As a result, the amount of data transmitted is less than needed for signal reconstruction. Thus, energy is saved but the classification success rate is affected.

Works such as [20], [11] and [12] divide the captured audio into segments called syllables. These segments are considered the smallest unit of call of a frog, thus containing the most important information for its classification.

In [20], the classification is performed using k-Nearest Neighbor (kNN) and Support Vector Machine (SVM) techniques, using the syllables as a reference. Vaca-Castao and Rodriguez [11] used the Mel-Fourier Cepstral Coefficient (MFCCs) to improve the results. MFCCs are known to be widely used in speech recognition [22]. This paper is based on the technique proposed by [20] to perform segmentation and the MFCC to improve the classification with kNN, which has better results in terms of classification [11]. However, instead of using the complete recordings as input for the classifier, the amount of information used is reduced by 90%.

### C. Gradient Projection for Sparse Reconstruction (GPSR)

In this work, we used a method for sparse reconstruction by gradients projection (GPSR) [23] to perform the reconstruction of signals. The method proposed in [23] is based on the optimization problem defined as:

$$\min_x = \frac{1}{2} (\| \mathbf{y} - \mathbf{A}\mathbf{x} \|_2)^2 + \tau \|\mathbf{x}\|_1, \quad (1)$$

where  $\mathbf{x} \in \mathbb{R}^n$ ,  $\mathbf{y} \in \mathbb{R}^k$ ,  $\mathbf{A}$  is a  $k \times n$  matrix,  $\tau$  is a positive parameter.  $\|v\|_2$  is the Euclidean norm of  $v$  and  $\|v\|_1 = \sum_i |v_i|$  is the  $\ell_1$  norm, or the Manhattan distance of  $v$ . This problem class is common in the signal processing area, it is directly related to the estimation of observations in a system of type:

$$\mathbf{y} = \mathbf{A}\mathbf{x} + n. \quad (2)$$

The presence of the  $\ell_1$  term in Equation 1 implies that the smaller components of  $x$  in Equation 2 tend to zero, resulting in sparse solutions, fulfilling the requirements of CS.

The algorithm proposed by [23] implements a solution for minimizing Equation 1 that allows the reconstruction of subsampled signals through an iterative of approximation method. The great advantage of this method is the application of a step for debiasing the approximation, ensuring that any calculation bias is removed. Thereby, it is possible to obtain a solution closer to the ideal solution.

### III. METHODOLOGY

#### A. Audio capture and analysis

In our experiments, we use audio samples taken directly from anuran habitat from four different anuran species. We collected 30 to 50 MB of WAV audio files of each specie.

As described in Section II-A, to reconstruct a signal based on a sparse sampling of the same, we need to define a reconstruction base that preserves the main characteristics the original signal. In the case of anuran classification, the frequencies of the acoustic signals of each specie are enough to make their classification [11], [20].

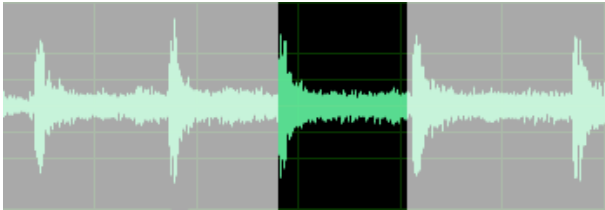


Figure 1. A call of the specie *Adenomera andreae* with a highlighted syllable

For the extraction of the characteristics of each monitored species monitored, we performed a segmentation of the audio into smaller units or syllables. Figure 1 shows that anuran calls are composed of peaks, followed by a moment of rest. However, as Figure 2 depicts, despite the fundamental frequencies of the species have a higher intensity on the peaks, they are still present in moments of rest. In order to maximize the use of fundamental frequencies, the separation

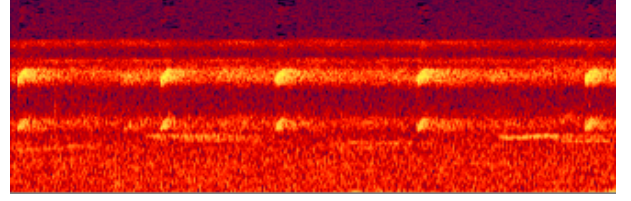


Figure 2. Spectrogram of a call of the specie *Adenomera andreae*

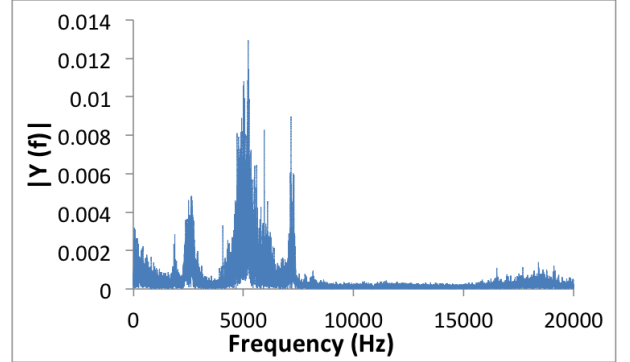


Figure 3. Frequency rate using Fast Fourier Transform (FFT) on a syllable of the specie *Adenomera andreae*

of syllables in this work was done according to Figure 1, where all frequencies are considered to be present between two calls.

Once syllables are segmented, we extract the most representative frequencies and used as a criterion for reconstruction and classification of the anuran. To separate the frequencies we use a Fast Fourier Transform (FFT) to identify and choose the frequencies of greater occurrence in each species. The spectrum of a syllable, as well as the fundamental frequencies, are shown in Figure 3.

#### B. Reconstruction base and experimental parameters

Once the fundamental frequencies of each species are identified, we create an audio reconstruction base. The reconstruction is based on the matrix  $\mathbf{A}$ , which will be used to evaluate and reconstruct the audio according to sampled vector  $x$  (Eq. 2).

For audio data, the reconstruction base must represent the characteristics of a sound wave. Thus, we generated sine waves similar to the ones in the anuran calls. Each frequency was converted into a wave and inserted as a row in matrix  $\mathbf{A}$ , consistent with 10 seconds, which is the size of the samples to be reconstructed.

Once the reconstruction base is built, and by using GPSR in [23], original recordings can be reconstructed by using subsamples from an original audio. To accomplish that, we define a downsampling parameter, called  $\alpha$ , and generate packets where only one out of every  $\alpha$  samples is captured and sent over the network. Complete recordings correspond

to a downsampling of  $\alpha = 1$ , and the compression increases proportionally to  $\alpha$ .

The main objective of the techniques of CS is to minimize information loss. Thus, even for signals generated with undersampling, where  $\alpha > 1$ , the result of the signal reconstruction is equivalent (in number of samples) to a signal where  $\alpha = 1$ .

#### IV. SIMULATION SCENARIO

In order to evaluate the impact of applying CS in WSN, simulations were performed by using the Network Simulator 2 (NS2) for different values of  $\alpha$  (5 to 100), in accordance with Section III-B. The simulated scenario was a grid with 625 nodes with 100  $m^2$  cells where, to simulate a real situation, the  $\langle x, y \rangle$  coordinates were randomly disturbed.

We defined an origin position to the audio generated by the anuran, in which was added a node to send information to all his neighbors. The information was broadcasted and the nodes that received the message forward it to the sink, according to the routing protocol. The positions of the audio generator node (anuran) and the sink were randomly chosen in each scenario.

To evaluate the performance of the WSN we considered: the average deliver delay, package loss and energy consumption of all the nodes. The routing protocol used was the One-Phase Push Diffusion (OPPD) [24]. The OPPD consists in two steps: recognition and transmission. On the recognition phase a routing tree is defined, using the sink node as the tree root. In the transmission phase, each node send information to its parent according to the generated tree, until the information arrives at the sink. The energy consumption was calculated according to the information on the Table I and the initial energy was defined as 22.000 mAh, as suggested on [2].

Table I  
POWER REQUIRED BY VARIOUS MICA OPERATIONS [2].

Operao	Consumption in nAh
Transmitting a packet	20.000
Receiving a packet	8.000
Operating sensor for 1 sample (analog)	1.080
Operating sensor for 1 sample (digital)	0.347
Reading a sample from the ADC	0.011
EEPROM Read Data	1.111
EEPROM Program/Erase Data	83.333

To test the variability of the environment, we considered nine scenarios in accordance with the range of the anuran calls (20 m, 40 m and 60 m) and network nodes (15 m, 30 m and 45 m). Each scenario consists in sending five minutes of audio (Section III-A), where node, sink and anuran positions are random. In addition, each scenario was repeated 33 times to establish an average of the parameters evaluated.

#### V. RESULTS

Once the  $\alpha$  parameter is applied on the recorded audio, we can reduce the amount of information in 90% according to the Figure 4. After the undersampling of the recordings, the audio files were tested on the classifier kNN and SVM with MFCCs features to verify the quality of the reconstruction. As can be seen in the Figure 4, once the compression increases, the quality on the classification decreases. We verified that for  $\alpha < 15$ , we can maintain the classification rate around 95%, even when using only 10% of the original audio.

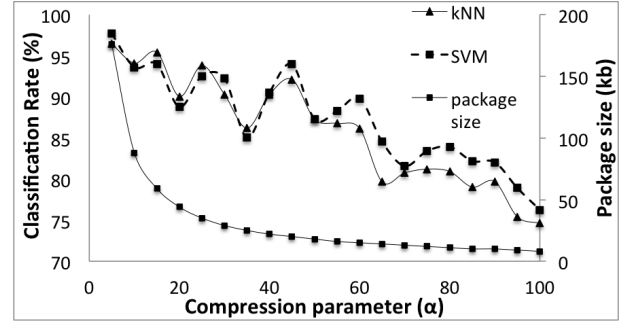


Figure 4. Audio package size according to the compression parameter  $\alpha$  and the classification rate using kNN and SVM methods.

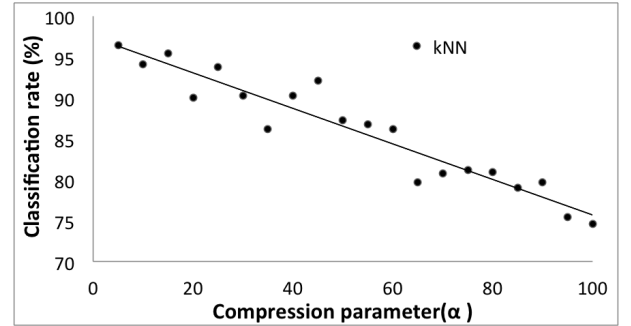


Figure 5. Linear regression of the classification rate using kNN according to the compression parameter  $\alpha$ .

To better understand the behavior of each classification method, we computed a regression using different values of  $\alpha$  and the associated Mean Squared Error (MSE) for each case. On the Figure 5 we can see the regression done for the kNN method, where we have an  $MSE_{kNN}$  of 1.6%.

In Figure 6, we have the same approach for the SVM method, where the  $MSE_{SVM}$  is of 2.2%. It is important to highlight that even if the  $MSE_{kNN}$  is lower than the  $MSE_{SVM}$ , the kNN method is much more sensitive to the variation on the  $\alpha$  parameter than the SVM. That can be evaluated by considering the angular coefficient of the curves in Figure 5 and Figure 6.

Beside the impact of the compression on the classification process, we also evaluate the performance of the WSN

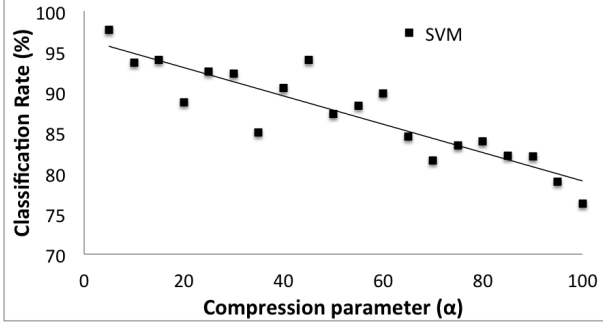


Figure 6. Linear regression of the classification rate using SVM, according to the compression parameter  $\alpha$ .

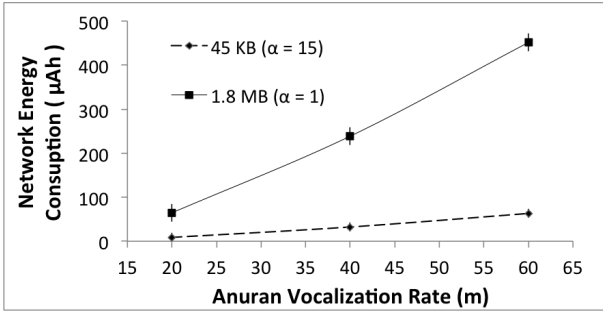


Figure 7. Energy Consumption according to the package size and call range.

according to the parameters defined in Section IV. For all simulations we compute the confidence intervals for 90% of confidence, as can be seen on each graphic.

The first evaluated parameter is the energy consumption, computed for the nine proposed scenarios. Considering that the main advantage of using CS in WSNs is reducing the amount of information that travels over the network, we evaluate two values of  $\alpha$  in each of the proposed scenarios:  $\alpha = 1$ , that represents the original audio (1.8 MB packages) and  $\alpha = 15$  (45 KB packages), that represents the best compression-classification rate (over 95%).

In Figure 7, we show the difference on the energy consumption comparing the use of the original audio ( $\alpha = 1$ ) and the compressed audio ( $\alpha = 15$ ). With a 90% compression rate the energy consumption decreases up to eight times in the proposed scenario (Figure 7) – the transmission range is fixed in 15 meters. The scenarios where the transmission range are 30 m and 45 m had similar results to the ones observed on the Figure 7, therefore, they were omitted for the sake of conciseness.

The second parameter being evaluated is the average delay of the packages sent to the sink from the recorder nodes. Figure 8 shows that, when the anuran call range increases, the delay also increases. That happens because a higher number of nodes perceives and send the anuran calls, overloading the network. In the case of the original

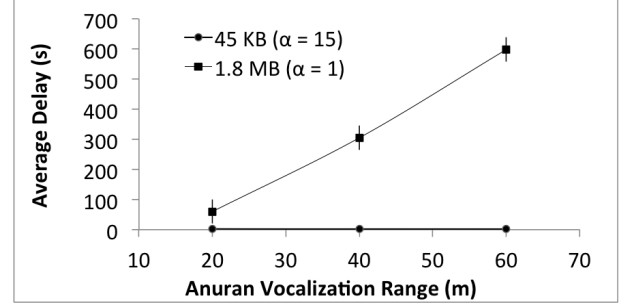


Figure 8. Average package delay according to the call range.

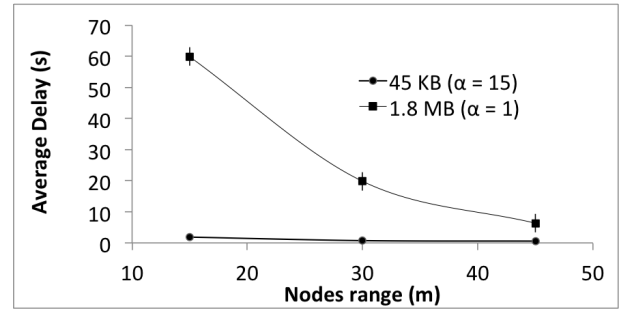


Figure 9. Average package delay according to the node range.

audio ( $\alpha = 1$ ), relies between 59.9 s and 597.5 s. In the case of the compressed audio ( $\alpha = 15$ ), the delay can be as low as 1.8 s and 2.2 s. The experiment observed on the Figure 8 had the range of the nodes fixed in 15 m. However, the other scenarios (30 m and 45 m) had a similar behavior on the delay rate.

Figure 9 shows a similar situation, but fixing the calls range in 20 m. As the node range increases the delay decreases accordingly. That is explained by the fact communication paths are smaller, according to the protocol specified on the Section IV. For the original audio ( $\alpha = 1$ ) the delay relies between 6.3 and 59.9 s. In the case of the compressed audio ( $\alpha = 15$ ), the delay relies between 0.6 and 1.9 s.

The last parameter being considered is the average delivery rate. Figure 10 depicts the scenario where the node range is fixed to 15 m. In this scenario, when the call range increases the delivery rate decreases accordingly. A similar behavior is observed on the scenarios where the node range is 30 m and 45 m. The delivery rate decreases because the amount of data captured by the nodes is higher, overloading the network.

Figure 10 shows the delivery rate for a fixed anuran-call range of 20 m. In a similar situation to the one observed on the Figure 8, when the anuran-call range increases the network load is greater, compromising the delivery time and the amount of packages received by the sink.

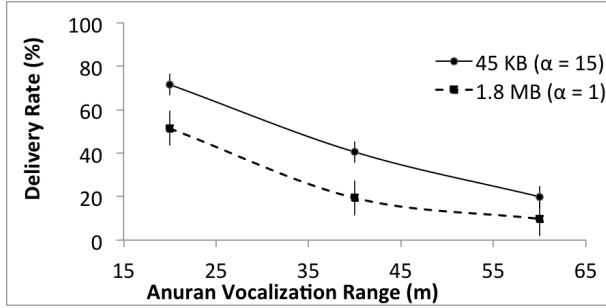


Figure 10. Package delivery rate according to the call range

## VI. CONCLUSION AND FUTURE WORKS

This work proves the viability of the application of Compressive Sensing (CS) techniques to classify anuran species using wireless sensor network and simulation tools. Considering the methods used by [20] and [11] it is possible to create a reconstruction base that considers the most relevant information to the classification software, sparing resources when obtaining the information needed to do a proper classification.

The most critical resource in a WSN is energy. We showed that can compress up to 90% without compromising the classification rate. This compression reduces the energy consumption in up to 88%, compared to a network without any kind of compression (original audio).

In addition, by evaluating the average delay and the delivery rate of the WSN, we verified that CS improves the network performance, achieving a delivery rate of up to 70% higher than a network without CS and decreasing the delay in up to 99% in the best cases.

The main advantage of the application of CS is the definition of a reconstruction base to obtain the signal from the received samples. By choosing the right parameters on this reconstruction base, the signal is good enough using only a small amount of information.

Future work includes implementing distributed algorithms to maximize the amount of information obtained without increasing the costs. In addition, we can work towards proposing methods to automatically generate a reconstruction base, creating a context-sensitive network. In this work, we used four different anuran specie, a future direction would be the applying the methodology for different wildlife sounds, including mammals and birds for several environmental monitoring applications.

## REFERENCES

- [1] J. Wang, J. Cho, and S. Lee, "An adaptive power-aware multi-hop routing algorithm for wireless sensor networks," in *Information Technology: New Generations (ITNG)*. IEEE, 2011.
- [2] A. Mainwaring, J. Polastre, R. Szewczyk, and D. Culler, "Wireless sensor networks for habitat monitoring," in *Proceedings of the 1st ACM international workshop on Wireless sensor networks and applications*. ACM, 2002.
- [3] S. Gowrishankar, T. G. Basavaraju, D. H. Manjaiah, and S. K. Sarkar, "Issues in wireless sensor networks," in *Proceedings of the World Congress on Engineering*. WCE, 2008.
- [4] V. Rocha and G. Goncalves, "Sensing the world: Challenges on wsns," in *Automation, Quality and Testing, Robotics. (AQTR)*. IEEE, 2008.
- [5] C. Carey, W. R. Heyer, J. Wilkinson, R. a. Alford, J. W. Arntzen, T. Halliday, L. Hungerford, K. R. Lips, E. M. Middleton, S. a. Orchard, and S. Rand, "Amphibian declines and environmental change: Use of remote-sensing data to identify environmental correlates," in *Conservation Biology*, 2001.
- [6] E. J. Candes and M. B. Waking, "An introduction to compressed sampling," in *Signal Processing Magazine*. IEEE, 2008.
- [7] S. Ji and L. Carin, "Bayesian compressive sensing and projection optimization," in *International Conference on Machine Learning (ICML)*. ACM, 2007.
- [8] Q. W. Hempstead and M. Yang, "A realistic power consumption model for wireless sensor network devices," in *Sensor and Ad Hoc Communications and Networks (SECON)*. IEEE, 2006.
- [9] N. Buluso and W. Hu, "Lightweight acoustic classification for cane toad monitoring," in *Asimolar Conference on Signals, Systems and Computers*. IEEE, 2008.
- [10] 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 *Symposium on Information Processing in Sensor Networks*, 2005.
- [11] 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.
- [12] N. C. Han, S. V. Muniandy, and J. Dayou, "Acoustic classification of australian anurans based on hybrid spectral-entropy approach," *Applied Acoustics*, vol. 72, no. 9, pp. 639–645, Sep. 2011.
- [13] J. Luo, L. Xiang, and C. Rosenberg, "Does compressed sensing improve the throughput of wireless sensor networks?" in *IEEE International Conference on Communications (ICC)*. IEEE, 2010.
- [14] S. Ji, D. Dunson, and L. Carin, "Multi-task compressive sensing," in *IEEE Transactions on Signal Processing*. IEEE, 2009.
- [15] R. Baraniuk, "Compressive sensing," in *IEEE Signal Processing Magazine*. IEEE, 2007.

- [16] J. Aelterman, H. Q. Luong, B. Goossens, A. U. Pi, and W. Philips, "Compass: a joint framework for parallel imaging and compressive sensing in mri," in *Image Proceesing (ICIP)*. IEEE, 2010.
- [17] P. Zamani, M. H. Kayvanrad, and H. Soltanian-Zadeh, "Cardiac cine mri using compressive sensing principles," in *Biomedical Engineering (ICBME)*. IEEE, 2011.
- [18] L. Carin, X. Ding, and L. He, "Bayesian robust principal component analysis," in *IEEE Transactions on Image Processing*. IEEE, 2011.
- [19] K. Riede, "Monitoring biodiversity: analysis of amazonian rainforest sounds." in *Ambio*, 1993.
- [20] J. C. Huang, Y. J. Yang, D. X. Yang, and J. Y. Chen, "Frog classification using machine learning techniques." in *Expert Systems with Applications*, 2009.
- [21] A. Taylor, G. Watson, G. Grigg, and H. a. McCallum, "Monitoring frog communities: An application of machine learning." in *Proceedings of the 8th Innovative Applications of Artificial Intelligence Conference (IAAI)*. AAAI, 1996.
- [22] S. Davis and P. Mermelstein, "Comparison of parametric representations for monosyllabic word recognition in continuously spoken sentences," in *IEEE Transactions on Acoustics, Speech, and Signal Processing*. IEEE, 1980.
- [23] M. A. T. Figueiredo, R. D. Nowak, and S. J. Wright, "Gradient projection for sparse reconstruction: Application to compressed sensing and other inverse problems," in *IEEE Journal of Selected Topics in Signal Processing*. IEEE, 2007.
- [24] X. Cheng, F. Wang, and J. Liu, "Hybrid push-pull for data diffusion in sensor networks without location information," in *International Conference on Communications (ICC)*. IEEE, 2008.