

# Effects of Compression and Window Size on Remote Acoustic Identification using Sensor Networks

*Ruben Gonzalez*

Institute for Integrated and Intelligent Systems,  
Griffith University, PMB 50, Gold Coast Mail Centre, QLD, 4217, Australia

## Abstract

Recently the cost-benefits of automated sensing over traditional field surveys for population management of fauna has been recognised [1,2]. Remote monitoring through automatic identification based on sensor networks has followed one of two approaches [3,7,18]; using the sensor nodes to perform data analysis within the network or alternatively using the sensor network as a means for collecting data to be centrally processed. In either case a key goal is minimising power consumption in sensor nodes which imposes constraints on both processing and communication capabilities. While the first approach aims to minimise communication requirements the other aims to reduce processing requirements. In the context of sensor networks for remote monitoring utilising centralised processing, this paper considers the impact on two different strategies for reducing communication requirements on the overall system performance.

## Introduction

Automatic identification of different fauna based on acoustic information has been an area of study over the last decade. This includes automatic identification of bird and frog calls [4,5,6,7], bat echolocation calls [8, 9] and insect

sounds such as those made by grasshoppers and cicadas [10].

Traditionally this has been performed by deploying data loggers in the field that are later retrieved and the captured audio data analysed. More recently wireless sensor networks have been identified as providing advantages for automatic remote monitoring. These advantages stem from being able to simultaneously deploy a large number of low cost sensing nodes since the potential effectiveness of sensor networks is related to the number of sensing nodes. The power constraints of sensor network nodes means they typically operate at low bandwidths and have limited processing capabilities. A typical sensor node such as one based on the Zigbee or IEEE 802.15.4 protocol has a total bandwidth of 128 kbps [11] and a low power embedded processor.

Early attempts at fauna identification made use of decision tree machine learning approaches [12]. Improved performance can be obtained via more advanced methods including k-Nearest Neighbour, Gaussian Mixture Models, Support Vector Machines [13,14] or Neural Networks [15] which all yield performance within a few percent of each other. These methods however require more processing power than is typically

available in a low power sensor network node and require the use of central processing.

The disadvantage of centralised processing in a sensor network is that many streams of essentially raw data from each node must be sent to the centralised server for analysis. The aggregate bandwidth capacity of any node receiving multiple streams of data can be easily overloaded if not carefully managed. This, as well as the need to minimise transmit power dissipation in each node drives the need to reduce the amount of data being transferred. Assuming a statistical multiplexing model in the time domain, the number of sensing nodes that can be active in any period of time (hence the effectiveness of the network) is inversely proportional to the size of each data stream. Data stream size reduction can be achieved by having sensor nodes select how much data to send and then compressing data prior to being sent through the network.

The question is, what is the impact of these data reduction strategies have on the performance of automatic identification? The answer will depend on the specific approaches used to implement these strategies and in the case of compression, the processing constraints in sensor nodes determines what type methods can be used. Accordingly the analysis in this evaluates the impact of ADPCM [16], being a relatively simple yet effective compression strategy that fits within the capability profile of typical embedded processors used in low-power sensor nodes as well as MP3 compression as an upper limit benchmark reference.

Apart from compression, how much data is transferred from a node through the network depends on the sample rate and length of the capture window. The sample rate is however not an independent variable since identification of certain animals requires specific frequencies to be captured which according to the Nyquist criteria determines the required sample rate. Hence the only means for controlling the amount of data being sent left to us is setting the length of the capture window. This in turn determines the analysis window size of the feature extraction algorithms and the performance of the automatic identification.

### **Acoustic Identification**

Acoustic identification of different animals is typically based on classification of a combination of temporal features of the audio signal together with spectral features derived from the short term Fourier transform or Mel-Frequency Cepstral Coefficients (MFCCs) [17]. Success rates vary in the range from 60% [18] to about 90% [15] depending mainly on whether the identification is performed in the field or under laboratory conditions. As the more advanced classification methods [13-15] perform within a few percent of each other we have selected the k-Nearest Neighbour approach for this evaluation.

For audio classification, a typical feature vector is composed of the signal bandwidth (BW), spectral centroid (SC), short-term energy (E), energy flux (F), Zero Crossing Rate (ZCR), and pitch (P) by means of sub harmonic summation [19]. These are defined as follows:

$$BW = \sqrt{\left(\sum_{k=1}^K (k - SC)^2 |X(k)|^2\right) / \left(\sum_{k=1}^K |X(k)|^2\right)} \quad (1)$$

$$SC = \left(\sum_{k=1}^K k \times |X(k)|^2\right) / \left(\sum_{k=1}^K |X(k)|^2\right) \quad (2)$$

$$E = \frac{1}{K} \left(\sum_{k=1}^K |X(k)|^2\right) \quad (3)$$

$$F = E(n) - E(n-1) \quad (4)$$

$$ZCR = \sum_{k=2}^K |\text{sgn}(x(k)) - \text{sgn}(x(k-1))|$$

$$\text{sgn}(n) = \begin{cases} 1, n > 0 \\ 0, n = 0 \\ -1, n < 0 \end{cases} \quad (5)$$

$$P = f : f \geq 0 \wedge \forall g \geq 0, H(f) \geq H(g);$$

$$H(f) = \sum_{k=1}^K h_k X(k \cdot f) \quad (6)$$

The first six MFCCs are also used as part of the feature vector. These are calculated as the cepstrum of the mel-warped spectrum:

$$\text{Mel}(f) = 2595 \log_{10}(1 + f/100) \quad (7)$$

$$c(k) = \text{IDFT}\{\log|DFT\{x(n)\}|\} \quad (8)$$

## Experimental Results

For evaluation we have used a database of 2590 instances of fifteen different species of native Australian frog calls from the Myobatrachidae family (Southern Frogs). These calls were each sampled at 22kHz and 16 bit resolution. A 12 dimensional feature vector consisting of the features (1) – (8) was extracted from the raw data using an analysis window of 1024 samples.

Using tenfold cross validation and k-Nearest Neighbour classification where k=1 achieved an 83.8% correct classification as the base performance. The F-measure defined as two times the product of precision and recall divided by the their sum is shown in Table 2. The confusion matrix is shown in Table 3.

At 22 kHz using 16 bit samples requires a 352 kbps bandwidth, which is not viable for continuous streaming data, but works as long as the nodes are not continuously capturing and the data transfer load is properly managed. Reducing the amount of data sent by any one node improves the network utilisation. Accordingly we reduced the capture window length and then extracted the feature vector and classified it again using the same method as before. The new window sizes were 128, 256 and 512 samples. The results using 2500 training vectors presented in Table 1 confirms that classification accuracy falls proportionately as a logarithmic function of the window size.

We next tested the effects of compression. This is required for continuous capture and reduces the bandwidth requirements to 88 kbps, which is within the network capacity. All the instances

of frog calls were then compressed offline using 4-bit ADPCM and then decoded again into raw audio for analysis. The ADPCM introduces quantisation noise into the audio data. The same features as before were then extracted from this less accurate data for all four different window sizes and classified using the same k-Nearest Neighbour method.

As is evident from Figure 1 ADPCM compression using 4 bits per sample relates to about a 2.5% drop on average in classification accuracy irrespective of analysis window size. This results because decreasing window size reduces the frequency resolution that the classification algorithm has to work with.

Window Size	ADPCM encoded	Raw Accuracy
128	67.5 %	69.8%
256	73.2 %	75.6%
512	76.4%	78.8%
1024	81.2 %	83.1%
2048	83.4 %	86.0%

**Table 1. Compression and Window Size Effects on Classification Accuracy**

In rare cases sensor nodes may have sufficient processing capability to utilise more advanced compression methods such as MP3. The higher level of compression provided by an MP3 encoder operating at 24 or 32kbps provides about a three times bandwidth saving over ADPCM and a commensurate drop in identification accuracy as is evident in table 2.

Window Size	24 kbps Accuracy	32 kbps Accuracy
128	10.8 %	12.4%
256	39.4 %	40.3%
512	61.0 %	62.3%
1024	77.1 %	77.1%

**Table 2. MP3 Compression Effects on Classification Accuracy with 2500 vectors**

The very low identification accuracy achieved for small window sizes with MP3 is not completely unexpected since the MP3 encoder reduces frequency resolution by discarding frequency information. Some compensation for this problem can be obtained by increasing the number of training vectors used in training the classifier as shown in Table 3, although for the purposes of this evaluation the number of training vectors has been kept constant (2500) for all tests.

Keeping the number of training vectors constant ensures that the same number of data points are used for training each instance of the classifier, however it could be argued that this causes those tests using smaller window sizes to train on less total amount of information as each vector summarises less original data. For example, generating 2,500 vectors using a window size of 1024 samples means that a total of 2.5 million samples are used (or 116 seconds of training data) whereas using a window size of 128 equates to only 320,000 samples used (14 seconds of data). Rather than keep the number of training vectors constant we can

keep the total amount of data samples (time) analysed constant. We note from the results in Table 3 that improved performance can be achieved in all cases with an increased number of vectors used for training the classifier.

Window Size	Training Vectors	ADPCM	MP3 32 kbps
128	24,000	73.5%	65.1%
256	14,000	78.4%	70.4%
512	7,000	81.4%	75.2%
1024	3,500	83.8%	79.0%
2048	1,750	86.8%	-

**Table 3. Accuracy as a function of number of training vectors**

In summary while MP3 provides three times the bandwidth savings over ADPCM the reduction in identification accuracy, using the same number of training vectors is in the range of 9% - 57%. In the best case this reduction is commensurate with the expected bandwidth savings but is far worse with the smaller window sizes. This is presented in Table 4 and graphically in Figure 4 that shows the average reduction in identification accuracy against the compression ratio when using a constant number of training vectors for various window sizes in percent reduction per unit of compression ratio.

Window	ADPCM	MP3
128	0.57	5.22
256	0.60	3.21
512	0.60	1.50
1024	0.47	0.55

**Table 4. Accuracy Bandwidth Ratio for constant number of training vectors**

In contrast the results obtained when using a constant total amount of information (about 160 seconds of data) for the training of the classifier are notably better. Table 5 shows that even in this case the reduction of accuracy against the compression ratio is still lower for the ADPCM compressed data than for the MP3 compressed data.

Window	ADPCM	MP3
128	0.25	0.76
256	0.43	0.73
512	0.50	0.56
1024	0.28	0.44

**Table 5. Accuracy Bandwidth Ratio for constant amount of information**

## Conclusion and Further Work

This study has evaluated the impacts of using data compression within sensor networks for automatic acoustic identification. While compression reduces the identification accuracy in all cases the benefits in bandwidth reduction may outweigh this reduction. It appears that for

this purpose the higher compression methods such as MP3 under perform compared to more lightweight methods such as ADPCM as demonstrated by the accuracy-bandwidth ratio. Using ADPCM a fourfold reduction in bandwidth equates to roughly a 3% reduction in identification accuracy.

The use of smaller capture windows relating to smaller packet sizes increases the effectiveness of sensor networks through the properties of statistical multiplexing but also reduces

identification accuracy by a greater margin. Hence compression is preferred as a means for data reduction.

Further work will consider to what extent the type of sounds being identified affects these results. Consideration will also be given to whether different classifiers respond in the same way to the effects of compression and window size reduction.

Scientific Name	Common Name	F-Measure
<i>Adelotus brevis</i>	Tusked Frog	0.824
<i>Asa darlingtoni</i>	Pouched Frog	0.913
<i>Crinia deserticola</i>	Torrid Froglet	0.755
<i>Crinia parinsignifera</i>	Beeping Froglet	0.881
<i>Crinia signifera</i>	Common Eastern Froglet	0.806
<i>Crinia tinnula</i>	Wallum Froglet	0.753
<i>Heleioporus australiacus</i>	Giant Burrowing Frog	0.91
<i>Lechriodus fletcheri</i>	Black-soled Frog	0.739
<i>Limnodynastes convexiusculus</i>	Marbled Marsh Frog	0.94
<i>Limnodynastes dumerilii dumerilii</i>	Eastern Pobblebonk	0.679
<i>Limnodynastes dumerilii grayi</i>	Eastern Pobblebonk	0.944
<i>Limnodynastes fletcheri</i>	Barking Marsh Frog	0.681
<i>Limnodynastes ornatus</i>	Ornate Burrowing Frog	0.95
<i>Limnodynastes peronii</i>	Striped Marsh Frog	0.727
<i>Limnodynastes salmini</i>	Salmon-striped Frog	0.694

**Table 5. F-Measures for fifteen species of Southern Frogs**

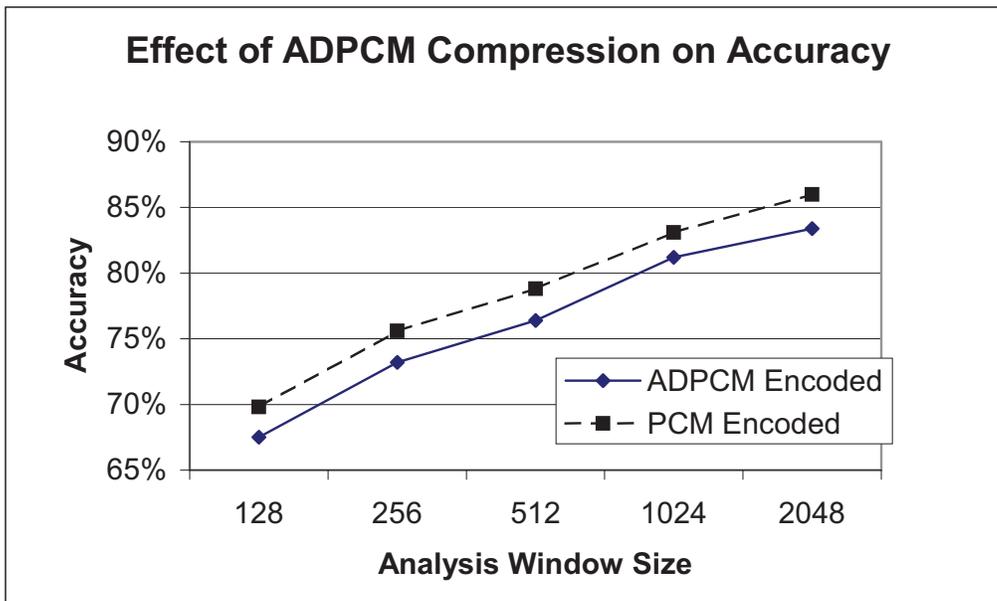


Figure 1. ADPCM Compressed Audio Classification Performance

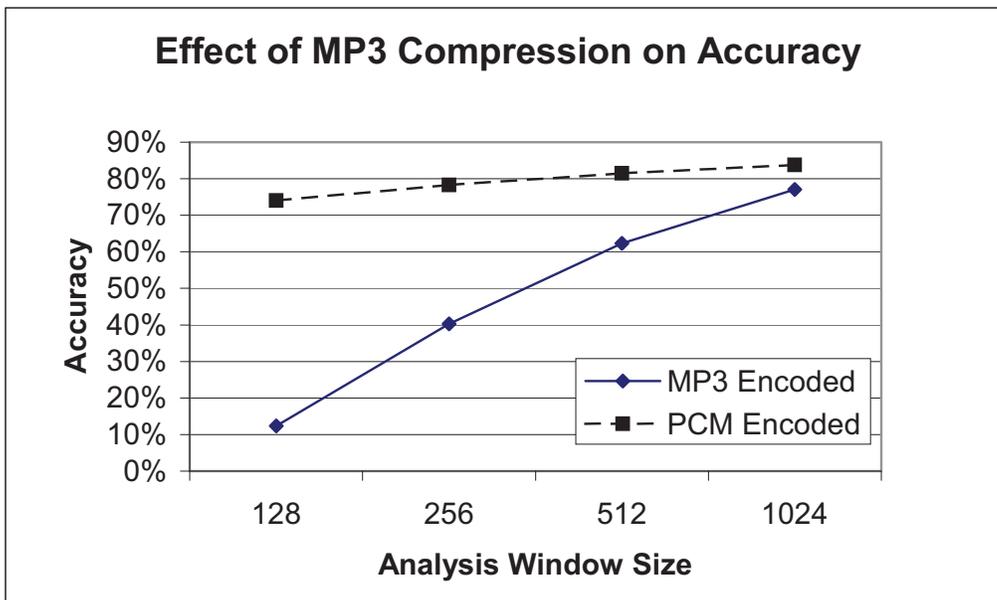


Figure 2. MP3 Compressed Audio Classification Performance

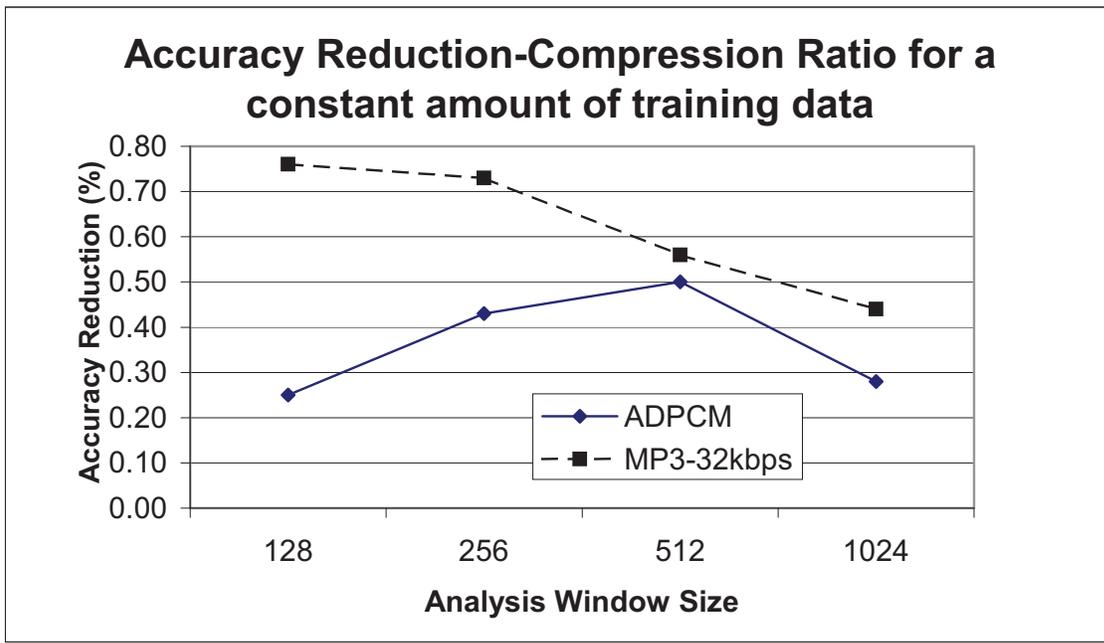


Figure 3. Accuracy Reduction-Compression Ratio

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
A	7											1			1
B		257		1	4	2					3				10
C			123	41								4		3	
D		1	22	353	1						1	3		3	
E		7			56	5							1	2	3
F		5		1		84		1		1	5		3		9
G							233	2		11					5
H						1	4	123	1	11				4	25
I		2		1				4	289		8	1	6	1	1
J						2	15	10	0	125					28
K		2		2		4			5		345		2	1	
L	1		3	2					2		1	64	1	13	13
M						2			3		6		229		
N		2	7	5	3			11	1		1	8		112	9
O		9		7	1	14	9	13	1	40		7		8	245

Table 6. Frog Call Confusion Matrix for window size of 1024 samples

## References

- 
- [1] T.D.Penman, F.L. Lemckert, M.J.Mahony, 2005, "A cost benefit analysis of automated call recorders" *Applied Herpetology*, Vol.2, No.4, 2005, pp.389-400.
- [2] Bridges and Dorcas 2000 "Temporal Variation in Anuran Calling Behavior: Implications for Surveys and Monitoring Programs." *Copeia*. 2:587-592
- [3] H.Wang, J.Elson, L.Girod, D.Estin and K.Yao, "Target Classification and Localization in Habitat Monitoring", In *Proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP 2003)* Hong Kong, China.
- [4] Miguel A. Acevedoa, Carlos J. Corrada-Bravoc, Héctor Corrada-Bravob, Luis J. Villanueva-Riverad and T. Mitchell Aidea, "Automated classification of bird and amphibian calls using machine learning: A comparison of methods" *Ecological Informatics* Volume 4, Issue 4, September 2009, Pages 206-214.
- [5] Mills H. 1995. "Automatic detection and classification of nocturnal migrant bird calls." *J Acoust Soc Amer.* 97: 3370-3371.
- [6] H.Wang, J.Elson, L.Girod, D.Estin and K.Yao, "Target Classification and Localization in Habitat Monitoring", In *Proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP 2003)* Hong Kong, China.
- [7] Darren Moore. "Demonstration of bird species detection using an acoustic wireless sensor network." 33rd IEEE International Conference on Local Computer Networks (LCN 2008) : SenseApp 2008; Montreal, Que. IEEE; 2008: 730-731. ISBN: 9781424424122
- [8] Vaughan N, Jones G and Harris S.1997. "Identification of British bat species by multivariate analysis of echolocation call parameters." *Bioacoustics* 7: 189-207.
- [9] Parsons S.2001. "Identification of New Zealand bats in flight from analysis of echolocation calls by artificial neural networks." *J Zool London* 253: 447-456.
- [10] Ohya Eand Chesmore ED.2003. "Automated identification of grasshoppers by their songs." Iwate University, Morioka, Japan: Annual Meeting of the Japanese Society of Applied Entomology and Zoology.
- [11] Burchfield, T.R., Venkatesan, S., Weiner, D. 2007. "Maximizing Throughput in ZigBee Wireless Networks through Analysis, Simulations and Implementations." UTDCS-24-07 and in proceedings of First International Workshop on Localized Algorithms and Protocols for Wireless Sensor Networks (LOCALGOS 2007)

- 
- [12] A.Taylor, G.Watson, G.Grigg and H.McCallum, "Monitoring Frog Communities: An application of Machine Learning", in the Proceedings of 8th innovative applications of AI conference (IAAI) Portland, Oregon 1996, pp.1564-1569
- [13] Chenn-Jung Huang, Yi-Ju Yang, Dian-Xiu Yang, "Frog classification using machine learning techniques", *Expert Systems with Applications: An International Journal* Volume 36 , Issue 2 (March 2009) Pages 3737-3743 ISSN:0957-4174
- [14] Miguel A. Acevedoa, Carlos J. Corrada-Bravoc, Héctor Corrada-Bravob, Luis J. Villanueva-Riverad and T. Mitchell Aidea , " Automated classification of bird and amphibian calls using machine learning: A comparison of methods," *Ecological Informatics*, Volume 4, Issue 4, September 2009, Pages 206-214
- [15] Yen, Gary G.; Fu, Qiang, "Automatic frog call monitoring system: a machine learning approach" *Proc. SPIE Vol. 4739*, p. 188-199, *Applications and Science of Computational Intelligence V*, Kevin L. Priddy; Paul E. Keller; Peter J. Angeline; Eds, 2002
- [16] ITU Recommendation G.726, "40, 32, 24, 16 kbit/s Adaptive Differential Pulse Code Modulation (ADPCM)"
- [17] Peltonen, V. Tuomi, J. Klapuri, A. Huopaniemi, J. Sorsa, T., "Computational auditory scene recognition", *Proceeding of. International Conference on Acoustics, Speech, and Signal Processing, (ICASSP 2002)*. IEEE, May 13-17, 2002, Orlando, FL, USA, vol.2, pp:1941-1944
- [18] Ning-Han Liu , Chen-An Wu and Shu-Ju Hsieh, Long-Term Animal Observation by Wireless Sensor Networks with Sound Recognition, WASA 2009, LNCS 5682, p.1-11, 5th Aug 2009
- [19] D.J.Hermes, "Measurement of pitch by subharmonic summation" *J. Acoust. Soc. Am.* Volume 83, Issue 1, pp. 257-264 (January 1988)