Proceedings of the Technical Sessions, 28 (2012) 22-30
Institute of Physics – Sri Lanka

Artificial Neural Network based Method for Identification of Indian Palm Squirrel Calls

W.A.M. Madhavi[1], D.D.N.B. Daya[1] and W.B. Yapa[2]
[1]*Department of Physics, University of Colombo*
[2]*Department of Zoology, University of Colombo*

ABSTRACT

Use of vocal parameters of the Indian Palm squirrels, (*Funambulus palmarum*) based on Mel-Frequency Cepstral Coefficients and Artificial Neural Networks, to identify different call types is discussed in this paper. Squirrel Calls were recorded from free ranging individuals in a home garden in Homagama, Sri Lanka, (Geographical coordinates of 6° 50' 27" North, 80° 0' 50" East)and of an captive infant squirrel with a digital sound recorder at 32 kHz sampling rate for 7 months were used for analysis. Based on the visual observations of the squirrel behavior made during vocalization, four types of calls (Alarm Calls, Mating Calls, infant squirrel food calls and infant squirrel sleepy calls) were identified. Mel-Frequency Cepstral Coefficients (MFCC) extracted for each type of call was used to train an Artificial Neural Network (ANN) with a view to individually identify the four types of vocalizations. The results show that the trained ANN can be used to identify the variation in vocalizations within an individual and the same group of animals. Alarm, mating and food calls were identified with a percentage accuracy of 100% and sleepy calls with 80% accuracy with the trained artificial neural network.

1. INTRODUCTION

Acoustic signals play an important role in the social life of many groups of animal species, be it to warn a threat of a predator to a group member (alarm calls), to attract or maintain the attention of an animal of opposite sex (mating calls), to claim the ownership of a territory (territorial signals) and to indicate sources of food (food calls) [1, 2].

Perusals of the literature show that the squirrels are known to emit vocalization in many different behavioral situations. Predator associated vocalization of Arctic Ground Squirrels [5], juvenile Richardson's ground Squirrels [4], North American Red Squirrels [6] and mating calls of Columbian ground squirrels [7], Eurasian red squirrels [3] have been studied. Furthermore calls of Gray Squirrel, Malaysian *Callosciurus* Social Desert Rodent, California Ground Squirrel and Thirteen Lined Ground Squirrel are also studied in depth [9-11].

The Indian Palm Squirrel (*Funambulus palmarum*) belongs to the kingdom Animalia, phylum Chordata, class Mammalia, order Rodentia and family Sciuridae.Three species of squirrels are found in Sri Lanka, the small palm squirrels that are seen in all rural and urban areas, the larger or giant squirrel that inhabits the forests, and the flying squirrels. The Palm Squirrel is the most common species found in the island. Sri Lanka has three species of palm squirrels. They are the Flame-striped Jungle Squirrel (*Funambulus layardi*), the Palm

Proceedings of the Technical Sessions, 28 (2012) 22-30
Institute of Physics – Sri Lanka

Squirrel (*Funambulus palmarum*), and the Dusky-striped Jungle Squirrel (*Funambulus sublineatus*) [2]. Only *Funambulus palmarum* is found in urban areas of the country.

The objective of this study is to construct an Artificial Neural Network which can identify different call types of Indian Palm Squirrels. Existing methods of noise reduction, feature extraction and neural network training were amalgamated to achieve the required target. Two call types of an infant squirrel held in captivity and two call types of squirrels found in natural habitats were identified.

2. METHODOLOGY

2.1 Recording squirrel vocalization

Vocalizations were recorded for squirrels found in a home garden in Homagama (an urban area close to the capital city)(Geographical coordinates of 6° 50' 27" North, 80° 0' 50" East), Sri Lanka from September 2009 to March 2010. A digital sound recorder CENIX W990 was used at sampling rate of 32 kHz and a bit rate of 192 kbps.

The recorder was mounted on a stand and kept in the backyard, approximately about 10 m to 20 m away from squirrels. In order to capture the complete behavior repertoires of the animals, the recorder kept continuously on and the behavior of the squirrels were observed and recorded regardless of the fact that they vocalize or not. This method enabled to record the corresponding behavior before, after and the vocalizations. Vocalizing squirrel was identified by its rhythmic movement of mouth and in case the vocalizing squirrel is not visible, the behavior of other squirrels was observed and recorded.

2.2 Sampling methods

Recording was done for three or four hours at a time and for a total time of 100 hours. Rainy days were excluded from recording as no squirrels come out during rain and recording process is quite unsuccessful in the presence of high noise level.

2.3 Selection of vocalization

Recorded audio clips initially consisted of long durations of clutter (such as areas without any vocalization, human or other animal vocalization and sounds of vehicles and machines other than the squirrel vocalization). From this, segments which mostly contained squirrel vocalization were selected using Adobe Audition 1.5 software by listening to the originally recorded clip and removing the unnecessary segments manually. Then the clips were cut to segments of 5 seconds duration to make all the input clips equal in length.

Audio signals then matched with the visual observations and categorized according to the behavior exhibited while a particular audio segment is recorded. One category of vocalization was given in the presence of predators i.e. when a cat or a rubber snake was placed in the vicinity of the squirrel. At these instances squirrels utter a chain of pulses

Proceedings of the Technical Sessions, 28 (2012) 22-30
Institute of Physics – Sri Lanka

while throwing their tail up and down. Sometimes they stay at the very position for a long time and sometimes they run on a tree bark giving rapid chain of pulses. Once the vocalization started they keep vocalizing for a long time even after the predator is removed from the area. This type of vocalization were observed most often and categorized under the name of alarm calls.

Squirrels give certain calls while exhibiting pre and post mating behavior. Male and female squirrel chase each other and sometimes bite each other. They stay close to each other for a long time grooming each other's body. After a long time they mate and if the process is interrupted they immediately separate. The calls given at any of these instances were categorized as mating calls. Mating calls had a soft texture than the alarm calls.

When the single infant squirrel was kept in captivity, a call was given when food is around but the squirrel cannot reach that. The condition was proved by the observation that when the food is given, high interest was shown to consume. This is not a hungry call indeed because, squirrels seems to be not giving any calls when they are hungry. This conclusion was helped with the observation that even when the captive squirrel kept in starvation, no calls were made. Therefore the call is named a food call.

The special series of chirps given by the infant squirrel every evening (5.00 p.m. – 6.00 p.m.) was named as the sleepy call since it was given about one hour before sleep. No interest was shown at that time when the food is given. Though named as sleepy call, this could be an isolation call emitted when the pup is not with the parent.

2.4 Noise reduction

Recordings done in natural habitats include high levels of noise. The sound of wind and trees lead to a stationary background noise. Other noises occur from vehicles, people speaking or other animals present in neighborhoods. It also arise electronically from thermal noise, tape hiss or distortion products. Often this background noise corrupts across all frequencies and is of substantial amplitude. Sound files obtained in this research were also consisted of high level of noise.

Noise reduction algorithms developed including Wiener filter methods, signal subspace methods, and statistical methods assume that the first few frames of the signal consist of noise only so that the initial noise estimation can be obtained from the first few frames [13]. Adaptive filtering changes the parameters used for the processing of signals according to some criterion to meet a performance requirement. But this requires the existence of a reference signal that is usually hidden in the approximation step of fixed-filter design. The minimum mean square error (MMSE) based algorithm is one of the best algorithms for noise reduction, but it is not designed to estimate noise from frames with signals [14].

Noise profile method determines all frequencies present in a given clip (which is only noise) and filters it from the whole sound clip. This method found to be the best among three noise reduction procedures, Noise Profile, Band Pass, and Noise Estimate. Noise

Proceedings of the Technical Sessions, 28 (2012) 22-30
Institute of Physics – Sri Lanka

Profile filtering was found to be effective in reducing noise and returning fairly high-quality signals from even severe levels of masking noise. Noise Profile filtering has proved to produce good results for cases where noise is approximately constant over the signal duration and the signal intensity exceeds noise intensity over the frequencies of interest [17].

Noise filtering was done in two steps. First the stationary background noise and the unwanted vocalization were removed using a noise profile obtained with Adobe Audition 1.5. Noise profile was selected to be a small segment of the same audio clip which only consisted of noise. This method was capable of removing noise to a very high level, retaining only the interested squirrel vocalization. Then a band pass filter with cut off frequencies 1000 Hz – 4000 Hz was applied to filter the higher order harmonics, retaining only the fundamental frequency. This made the vocalization less confused and easy to analyze. As a third step the filtered audio clips were normalized to make all clips equal in amplitude by removing the effect of distance to the vocalizing animal.

2.5 Feature extraction

In speech or sound recognition systems, it is essential to determine which features are to be used as features for recognition of signal properties. Several algorithms and signal processing methods have developed and tested for procession of speech [14], bird species recognition, synthesis of bird vocalization [13-16], studies on animal vocalization and recognition of animal psychological conditions [12] based on LPC, ZCR, RASTA PLP, FFT, DCT, DWT, CWT, MFCC and BFCC.

MFCC (Mel-Frequency Cepstral Coefficients) are robust to noise and proved to be good features in bird song analysis and human speech recognition [18]. Therefore MFCC's are used as the features for call recognition. Extraction of MFCC coefficients was done using Matlab 7.0. First the signal was imported as a wav file and the data were windowed with a hamming window of size 400. Windowed data were shifted to FFT order then and were put in linear filter banks and log filter banks. A weighting function is applied to each block to control the shape of the frequency responses of the filters. 13 linear filters and 27 log filters were used and FFT bins were combined so that each filter has unit weight, assuming a triangular weighting function. The lowest frequency was set to 1000 Hz. After shifting the data to log filter banks, the inverse cosine transform was taken to reduce dimensionality.

2.6 Training and testing of artificial neural network

Obtained matrix of MFCC contained 5 coefficients, each consisting 500 points (i.e. value of the coefficient was calculated for each frame for 500 frames throughout the signal). These values were averaged to obtain 5 values for each call and 20 calls were selected from each of the four categories.

A feed-forward backpropagation neural network (Network A) of three layers was constructed with the first and second layers having two neurons while third layer having

Proceedings of the Technical Sessions, 28 (2012) 22-30
Institute of Physics – Sri Lanka

one neuron. Number of neurons in the input and output layer are defined by the problem since the system needed to differentiate one call type from two types of calls. Second layer was also selected to be two to maintain the simplicity of network. Training data set was 40 vectors (20 sleepy calls and 20 food calls), each containing average values of five MFC coefficients of respective vocalization. Log sigmoid transfer function was applied for first and second layers while pure linear transfer function was applied to third layer of neurons. Supervised learning batch training was used with Levenberg-Marquardt (trainlm) training algorithm. Secondly, a similar network (network B) was fed with an input of 40 vectors, consisting of 20 alarm calls and 20 mating calls. Finally, the process extended to identify all four vocalization types at once by forming a feed-forward network with three layers (network C), each layer having four neurons and the training set was consisted of 80 vectors (20 from each category) constructed in the same manner.

Testing data set was constructed with 10 vectors (5 from each category) and was given as the input of neural network. Testing was also done in two steps. First, networks A and B were simulated separately to identify between two vocalization types and then network C was simulated with a training set of 20 vectors (5 from each category).In order to determine to which type the input vocalization belongs, each output of the network obtained after simulation was rounded off and compared with expected outputs.

First, the testing set was introduced in a particular order, i.e. the first five vectors belonged to one category and next five to another category. When testing with all 20 testing vectors, the set was randomly introduced. After the network has trained, testing vectors were introduced one by one to see whether it can correctly identify the call type.

3. RESULTS AND DISCUSSION

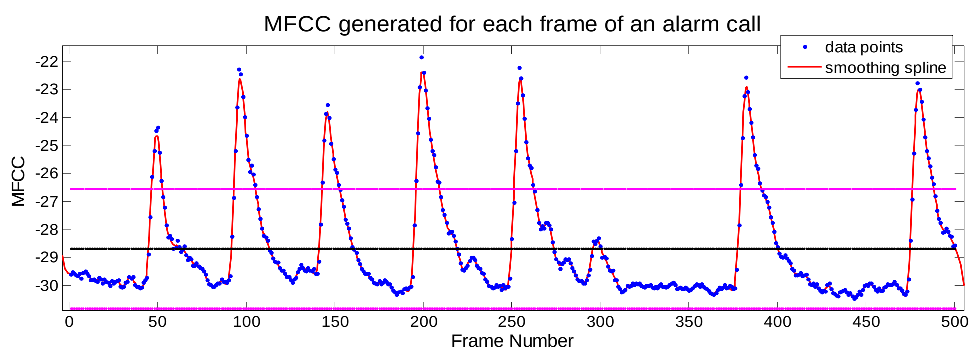

Fig. 1: Generated MFCC for each frame of an alarm call

Proceedings of the Technical Sessions, 28 (2012) 22-30
Institute of Physics – Sri Lanka

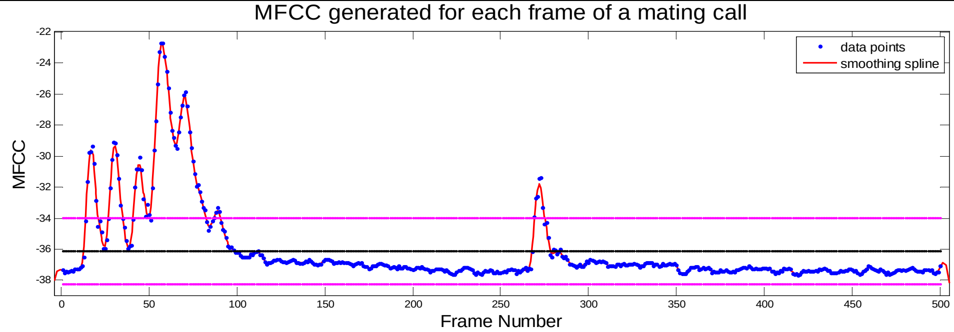

Fig. 2: Generated MFCC for each frame of a mating call

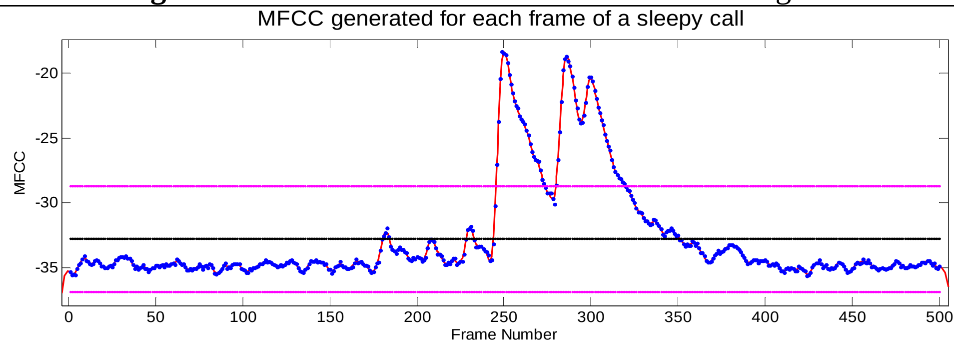

Fig. 3: Generated MFCC for each frame of a sleepy call

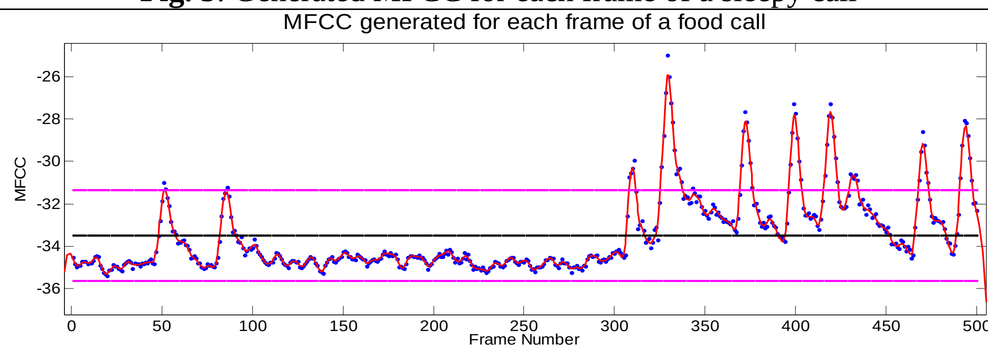

Fig. 4: Generated MFCC for each frame of a food call

Figs 1-4 show the generated MFCC for different calls and it is evident that each pattern differs significantly. Figure 5 was generated by plotting the average value of first MFC coefficient for forty squirrel calls; ten calls from each category. X axis shows the call number (1 to 10). The four call types can be separated into four regions. In fact the identification constraints were automatically determined by the artificial neural network using five coefficients while it is trained to identify different calls. Fig. 5 shows only the possible separation using only one coefficient since five dimension picture cannot be generated. The obtained output of Network A and Network B are shown in Fig. 6 and illustrated in Table 1. Cross marks are the expected output (i.e. if the network properly identifies the vocalization, network output should be similar to the expected output). Network output overlaps with the expected output for all five alarm, mating and food calls.

Proceedings of the Technical Sessions, 28 (2012) 22-30
Institute of Physics – Sri Lanka

Only four sleepy calls were correctly identified and one call was misidentified as a food call.

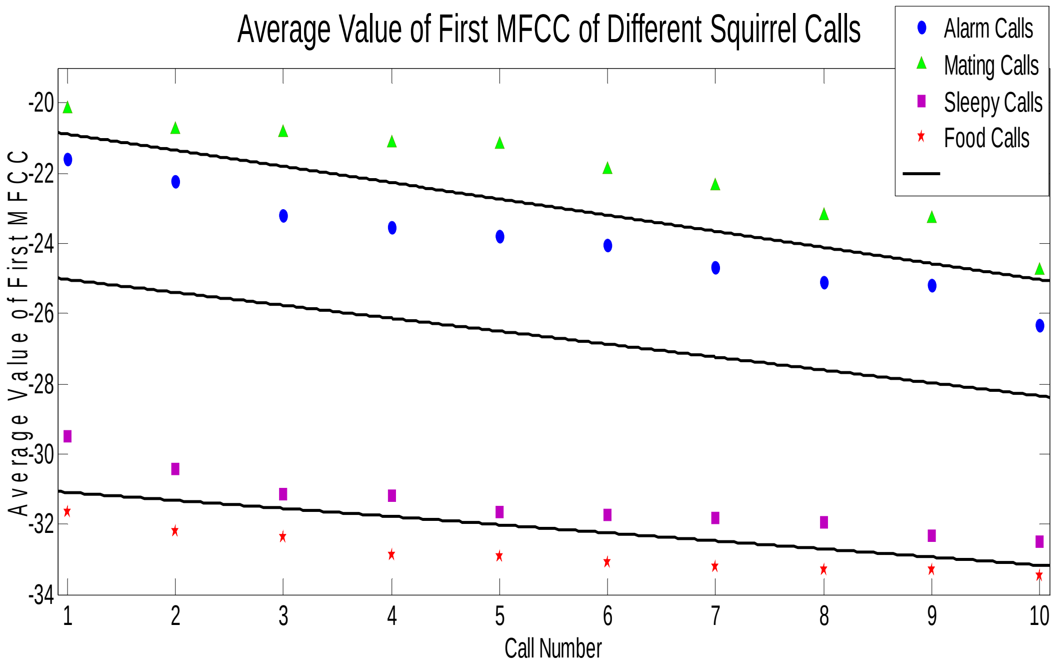

Fig. 5: Average value of first MFCC for different squirrel calls

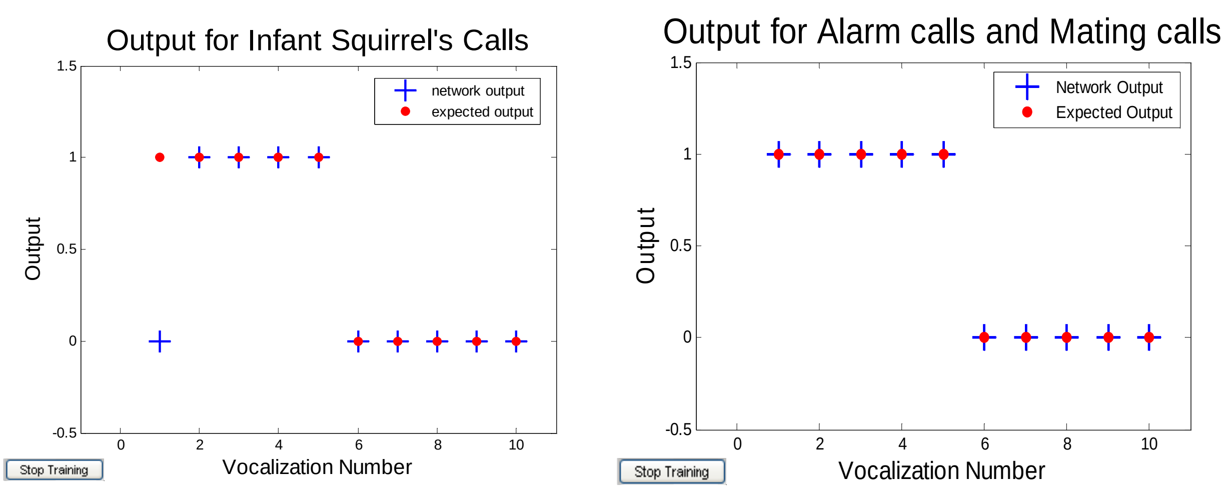

Fig. 6: Neural Network Performance for different types of calls

Proceedings of the Technical Sessions, 28 (2012) 22-30
Institute of Physics – Sri Lanka

Table 1: Identification efficiency of the trained artificial neural network

Call Type	No of calls tested	No of calls correctly identified	Percentage accuracy of identifying a call %
Alarm	5	5	100
Mating	5	5	100
Sleepy	5	4	80
Food	5	5	100

4. CONCLUSION

This study demonstrated that the features of squirrel vocalization could be successfully described with Mel Frequency Cepstral Coefficients. It can identify different vocalization within same animal and same species (Indian Palm Squirrels). Artificial neural networks can be trained to predict the purpose of the vocalization when MFCC used as the inputs of the network. The trained artificial neural network could identify two types of calls within the same animal (infant squirrel) and two call types of Indian palm squirrels in natural habitats near Homagama. The network could be trained to identify more call types if more audio clips of squirrel vocalization can be obtained. The identification method used in this work can be adopted to distinguish different animal species from their vocalization and also to identify the variation in vocalizations within the same group of animals.

REFERENCES

1. Steve Csurhes**,** *Indian Palm squirrels,*Department of Employment, Economic Development and Innovation Biosecurity Queensland, (2010).
2. Jayantha Jayawardene, Squirrels Daily News, (2006).
3. Lee, T.H., *Mating behavior of the Eurasian red squirrel (Sciurus vulgaris Linnaeus) in Hokkaido, Japan,*Mammalia (2001)131-142.
4. Warkentin, K.J.W., H.Keeley, A.T. and Hare, J.F., *Repetitive calls of juvenile Richardson’s ground squirrels (Spermophilus richardsonii) communicate response urgency*, NRC Research Press, (2001).
5. Melchior, H.R., *Characteristics of Arctic Ground Squirrel Alarm Calls*, Ecologia, (1971)184-190.
6. Digweed S.M. and Rendall, D., *Predator-Associated Vocalizations in North American Red Squirrels (Tamiasciurus hudsonicus): To Whom are Alarm Calls Addressed and How do They Function*, Ethology, (2009)115.
7. Manno, T.G., Nesterova, A.P., Debarbieri, L.M., Kennedy, S.E., Wright K.S. and Dobson, F.S., *Why do male Columbian ground squirrels give a mating call*, ANIMAL BEHAVIOUR, (2007) 1319-1327.
8. Mitrovic, D., Zeppelzauer, M. and Breiteneder, C., *Discrimination and Retrieval of Animal Sounds*.

Proceedings of the Technical Sessions, 28 (2012) 22-30
Institute of Physics – Sri Lanka

9. R.S.Lishak, *Gray squirrel mating calls:A spectroscopic and ontogenic analysis*, Journal of Mammalogy, (1982) 661-663.
10. Harris, M.A., Murie, J.O. and Duncan, J.A., *Response of Columbian Ground Squirrels to Playback or Recorded Calls*, (1982).
11. Randall, J. A.and Rogovin, K.A., *Variation in and Meaning of Alarm Calls in Social Desert Rodent* (2002).
12. Ikedaa, Y. and Ishii, Y., *Recognition of two psychological conditions of a single cow by her voice*, Computers and Electronics in Agriculture, (2008) 67–72.
13. Cai, J., Ee, D., Pham, B. and Roe, P., *Sensor Network for the Monitoring of Ecosystem:Bird Species Recognition*, Proceedings The Third International Conference on Intelligent Sensors, Sensor Networks and Information Processing, (2007) 293-298.
14. Vijaykumar, V.R., Vanathi, P.T., Kanagasapabathy, P., *Modified Adaptive Filtering Algorithm for Noise Cancellation in SpeechSignals*, Electronics and Electrical Engineering, (2007) 1392 – 1215.
15. Jeremy Hyman, *Countersinging as a signal of aggression in a territorial songbird*, Animal Behaviour, (2003), **65,** 1179–1185.
16. Brumm, H. and Naguib, M., *Environmental Acoustics and the Evolution of Bird Song*, Advances in the Study of Behaviour, 40 (2009).
17. Baker, M.C. and Logue, D.M., *A comparison of three noise reduction procedures applied to bird vocal signals*, (2007).
18. Agranat, I., *Automatically Identifying Animal Species from their Vocalization,* (2009).