2014 IEEE International Conference on Advanced Communication Control and Computing Technologies (ICACCCT)

An Automatic Method to Detect the Presence of Elephant

Arpit Sourav Mohapatra[1], Sandeep Singh Solanki[2]
[1,2]Department of Electronics and Communication Engineering,Birla Institute of Technology, Mesra,Ranchi, India
[1]arpitsourav12@gmail.com,[2]sssolanki@bitmesra.ac.in

***Abstract*— In this paper an approach has been discussed to automatically detect the presence of elephant by the detection of a low frequency sound produced by elephant called rumble. A detection system has been developed which uses features of rumbles as input to recognize the rumbles. Feature extraction techniques have been used to extract the features of rumble which includes the greenwood function cepstral coefficients (GFCC) and first three formant frequencies. The GFCC features have been extracted in a procedure similar to mel frequency cepstral coefficients (MFCC) extraction while the formant frequencies have been extracted using linear predictive coding (LPC). A robust feature vector has been created by cascading the formant frequencies to the GFCC features. The detection system has been developed using feed forward neural network which is trained using backpropagation algorithm.**

***Keywords*— Rumbles, LPC, STFT, MFCC, GFCC, formant**

I. INTRODUCTION

As human beings are more indulging in deforestation the wild animals especially elephants are getting more frequent visitors into human habitations. In many rural areas of India the elephant attacks are common news where these mammals come in search of food and destroy lives, properties and crops in order to feed themselves. The villagers try to drive off them by firing crackers and beating drums. However, they start driving off the animals only when the villagers are attacked as they do not have any prior information of the advancing elephants. In such attacks it is not only the humans are victims but the elephants are also affected because in the course of drive off action the mammals sometimes get injured and even get killed, thereby dwindling the human approach towards elephant. Several techniques have been developed to scale down the human and elephant conflict which includes electric fencing, radio collar technique and global positioning system. The use electric fencing can lead to deaths of elephants, other animals and even humans. The radio collars are expensive and there is a chance of its damage by the animals. In such case research has been going on to find alternative ways. One such way has been discussed by [1] which tells about the analysis of rumbles to detect elephants automatically and localize their position. The rumbles are calls which contain infrasonic components [2] made by elephants for communication. These sounds can propagate long distances at least 2 KM [1], thereby indicating that such sounds can be used to detect elephants from a long distance which means the villagers could have ample amount of time to take suitable action.

The spectrogram, which is a plot between time and frequency, of rumble, can be plotted using short time Fourier transform (STFT), can be used to visualize the rumbles. Figure 1 shows the spectrogram of a rumble and Fig. 2 shows a spectrogram of a human spoken phrase, “Birla Institute”. From the figures the differences between the rumble and human voice can be seen clearly. The former shows that the rumbles have a curve like feature which is due to presence of frequency modulated components in the sound [3]. Also the information of rumble has a frequency range of 1000Hz while that of human voice has more. Although the rumble has frequency information span of 1000Hz but most of the energy are concentrated within 500Hz.

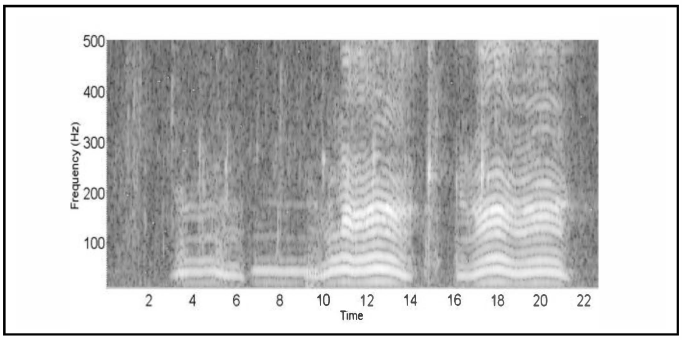

Fig.1. Spectrogram of rumble

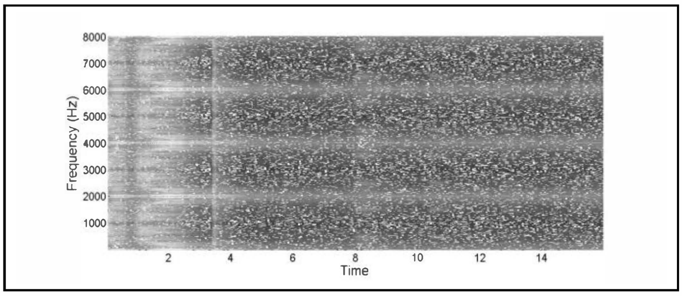

Fig.2. Spectrogram of voice

Research has been going on the automatic detection of rumbles for a long time. Two approaches have been found out in literature which is based on feature extraction. Both the approaches make use of two different kinds of features viz. Greenwood Frequency Cepstral Coefficients (GFCC) and the

ISBN No. 978-1-4799-3914-5/14/$31.00 ©2014 IEEE

formant frequencies to automatically detect rumbles. The first method [4] uses Asian elephant (Elephas Maximus) rumbles for analysis. It has used 12 recorded sounds to train a Hidden Markov Model (HMM) which has approximately 2000 samples of features and 5 recorded signals which have approximately 800 samples for testing. It has used ratio of formants as features. The second method [5] uses African elephant (Loxodonta Africana) rumbles for analysis. It uses GFCC features along with mean and variances of these features over time. These features are used to train a Support Vector Machine (SVM) for automatic detection of rumbles. It has used four hour of recorded rumble in their data set. The features that have been used are described in the next section

II. FEATURES USED

A. Greenwood frequency cepstral coefficients

It has been found that many mammals perceived frequency on logarithmic scale along cochlea and the relationship can be modeled as [6]

$$A(10^{ax} - k) \tag{1}$$

Where a is a constant and x is the cochlea position. If f is the actual frequency in Hz and F_P is the perceived frequency on cochlea then F_P is given by

$$F_P(f) = \frac{1}{a}\log_{10}(\frac{f}{a} + k) \tag{2}$$

If the hearing range of species under study is from f_{min} to f_{max} then the values of species specific coefficients a and A are given by

$$A = \frac{f_{min}}{1-k} \tag{3}$$

$$a = \log_{10}(\frac{f_{max}}{A} + k) \tag{4}$$

Experimental data shows that the value of k can be taken as 0.88 [6]

B. Formants

Formants [7] are resonating frequencies of vocal tract and are different from fundamental frequency. Formants frequencies can be identified from the spectrogram where the intensity of the plot is maximum. To extract formants [2] the linear prediction codes can be used to model the signal as an all pole filter:

The LPC model of signal as all poles filter is given by

$$H(z) = \frac{1}{1 - \sum_{1}^{N} a_k z^{-k}} \tag{5}$$

Where N is the number of poles and a_1, $a_2 \ldots a_N$ are coefficients. We assume the poles of this function be P_1, $P_2 \ldots P_N$. Then

$$H(z) = \frac{1}{1 - \prod_{k=1}^{N}(1 - p_k z^{-k})} \tag{6}$$

Each pair of conjugate pole in (6) is associated with a formant and is given by

$$F_k = \frac{f_s}{2\pi}\arg(p_k) \tag{7}$$

where f_s is the sampling frequency. However, there are non resonating pairs of poles, which we need to remove. Such poles are to be removed by higher bandwidths which can be calculated as

$$B_k = \frac{f_s}{\pi}\ln|p_k| \tag{8}$$

The formants corresponding to poles B_K greater than 300 Hz are removed while computing the formants.

III. PROPOSED METHOD AND METHODOLOGY

The data base for this work was created by downloading rumbles of different kinds having different call durations, of African elephants (Loxodonta Africana) from website www.elephantvoices.org. The downloaded rumble files have a sampling frequency of 44100 Hz. Since most of the information lie within 500 Hz the sound files were re sampled with a sampling frequency of 2000Hz using MATLAB 2013.

A feature based method has been proposed similar to [5] which uses both Greenwood Function Cepstral Coefficients (GFCC) and formant frequencies as features. The GFCC extraction process is similar to Mel Frequency Cepstral Coefficients (MFCC) [8] but the Mel scale is replaced with Greenwood scale [6]. The block diagram of GFCC extraction process is given in Fig 3. As discussed in earlier section about the hearing range f_{min} to f_{max}, for elephant it was taken to be 10 Hz (f_{min}) to 10000 Hz (f_{max}). Thirty filters have been used in the range of 10 Hz to 500 Hz which corresponds to the frequency range of rumble to extract GFCC. The GFCC extraction process would give 30 coefficients corresponding to 30 filters of which 18 coefficients have been used as cepstral features to represent the sound.

The first three formants were calculated using the (7). While calculating the formants the number of LPC coefficients is taken to be eight. The 18 cepstral features were cascaded with three formants making a feature vector of 21 dimensions. The rumble audio files contain two parts viz. the actual rumble which we call signal and absence of rumble which we call noise.

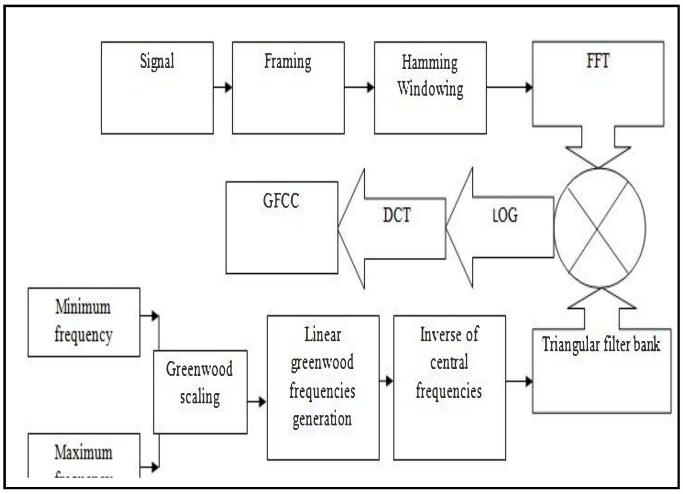

Fig.3. Extraction of GFCC features

Some audio files were selected as training data which corresponds to 150 seconds and which includes both signal and noise. The audio files were divided into frames of length 60ms with an overlapping of 30ms. The frames were windowed using a hamming window. For each windowed frame the 21 dimension feature vector was computed. The rumble audio files were analyzed and frames which correspond to signals were labeled with a numerical value 1 while frames corresponding to noise were labeled with numerical value 2. A feed forward neural network was created using MATLAB 2013 having 21 input neurons and a single neuron as output and three neurons in hidden layer. The labels ‘1’ and ‘2’ are used as reference output for the neural network corresponding to signal and noise respectively. The 21 dimension feature vector of each frame of training data was made input to a feed forward neural network for training the network using back propagation algorithm. The network was trained for 1000 iterations and was ready for testing. The block diagram of neural network showing number of inputs and output is given in Fig. 4.

To test the trained network the rest of the files were made testing data which corresponds to 100 seconds. When the testing data was given input, it was found out that the network output is deviating from the reference outputs 1 and 2; hence it was decided to set a range to the reference output. The range for reference output of signals was computed as

$$R1 = mean(Y_i) + STD(Y_i) \tag{9}$$

$$R2 = mean(Y_i) - STD(Y_i) \tag{10}$$

i=1, 2…N

Where Y_i is the output of neural network for ith input frame of training data corresponding to signals and N is the number of frames in the training data corresponding to signals. We say a rumble is truly detected if and only if a signal input is mapped to an output within the calculated range. If a noise input is mapped into an output within the same range then we say rumble has been falsely detected.

The detection rate of rumble was calculated as the percent of truly detected rumbles over total number of input testing frames corresponding to signals. The false detection rate was calculated as the percentage of falsely detected rumbles over total number of testing frames signal corresponding to signals.

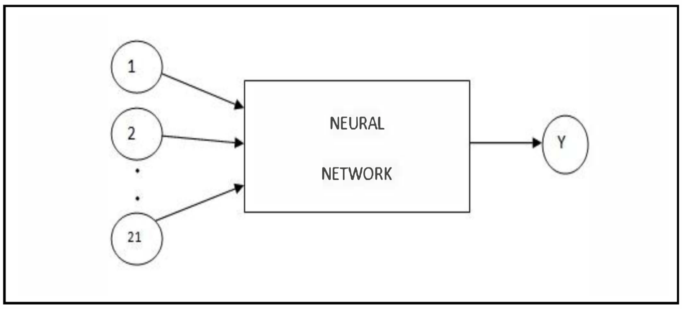

Fig.4. Neural network showing 21 input neurons and a single output neuron

IV. RESULT AND DISCUSSION

Figure 2 shows the spectrogram of an audio file which has rumble as well as noise. The portion labeled as A, B and D are rumbles. The neural network designed is able to detect the portion A, B and D as rumbles. Besides the network also detects portion C as a rumble which is, however, a noise.

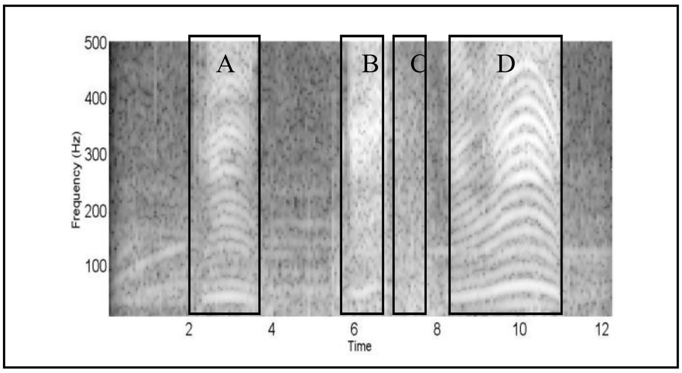

Fig.5. Detection of rumbles

It was found out that the proposed method gives a good accuracy of rumble detection which is equivalent to existing methods. While at the same time there is a trade off with false positive rate which is unwontedly high. Table 1 shows the comparison among existing methods and proposed method.

The accuracy of rumble detection was found to be 90.0% while the false positive rate was found to be 30.0%. .As the training and testing data were only of 250 seconds commutatively and the noise content is less as compared to signals in the database; it could have led to development of less optimized network. This could be the plausible reason of such a high false positive detection. Another challenge in this work was to manually label the rumbles and noises. As this is the work of an expert and the people involved in this work do not have any expertise knowledge regarding the rumbles, this may have led to inaccurate labeling.

TABLEI. COMPARISON OF METHODS

methods	***Detection rate***	***False positive***	***Features uesd***
1	97.0 %	1.93%	Ratio of formants
2	91.0%	26%	GFCC along with their mean and variances
proposed	90.00%	30.00%	GFCC and formants

V. CONCLUSION

The proposed method gives a detection rate of 90.0% but can also be improved by using noise cancellation techniques. Also if this method is implemented with a more comprehensive data set and with the help of experts who could distinguish between rumble calls and noises, the false positive rates may go down. Such an approach could provide better results because in addition to cepstral information the information of formants are also used for analysis

REFERENCES

[1] Dissanayake, C.M.; Kotagiri, R.; Halgamuge, M.N.; Moran, B.; Farrell, P., "Propagation constraints in elephant localization using an acoustic sensor network," Information and Automation for Sustainability (ICIAfS), 2012 IEEE 6th International Conference on , vol., no., pp.101,105, 27-29 Sept. 2012.

[2] Smita Nair and Rohini Balakrishnan, Chandra Sekhar Seelamantula and R. Sukumar,, “Vocalizations of wild Asian elephants (Elephas maximus): Structural classification and social context,”. Acoustical Society of America 2009, vol 126, no.5, pp. 2768-2778, Nov 2009.

[3] Bardeli, R., "Similarity Search in Animal Sound Databases," Multimedia, IEEE Transactions on , vol.11, no.1, pp.68,76, Jan. 2009.

[4] Wijayakulasooriya, J.V., "Automatic recognition of elephant infrasound calls using formant analysis and Hidden Markov Model," Industrial and Information Systems (ICIIS), 2011 6th IEEE International Conference on , vol., no., pp.244,248, 16-19 Aug. 2011.

[5] Matthias Zeppelzauer, Angela S Stoger, Christian Breiteneder, “Acoustic Detection of Elephant Presence in Noisy Environments” MAED, 2013 2nd ACM international workshop on Multimedia analysis for ecological data, pp. 3-8, 21-25 Oct 2013.

[6] Clemins, P.J.; Trawicki, M.B.; Adi, K.; Jidong Tao; Johnson, M.T., "Generalized Perceptual Features for Vocalization Analysis Across Multiple Species," Acoustics, Speech and Signal Processing, 2006. ICASSP 2006 Proceedings. 2006 IEEE International Conference on , vol.1, no., pp.I,I, 14-19 May 2006.

[7] Snell, R.C.; Milinazzo, F., "Formant location from LPC analysis data," Speech and Audio Processing, IEEE Transactions on, vol.1, no.2, pp.129,134, Apr 1993.

[8] Kopparapu, S.K.; Laxminarayana, M., "Choice of Mel filter bank in computing MFCC of a resampled speech," Information Sciences Signal Processing and their Applications (ISSPA), 2010 10th International Conference on , vol., no., pp.121,124, 10-13 May 2010.