Chapter 15
Algorithm for Gunshot Detection Using Mel-Frequency Cepstrum Coefficients (MFCC)

Preetam Suman, Subhdeep Karan, Vrijendra Singh and R. Maringanti

Abstract Protection of forests and wildlife needs efficient, reliable, and real-time detection of events such as gunshots, wood cutting, distress call of animals, etc. In this paper, we propose a gunshot detection technique through acoustic signal pattern recognition utilizing Mel-Frequency Cepstrum Coefficients (MFCC). In this work, MFCC is used to extract the features of gunshots from prerecorded analog sound files. Training of the system for gunshot detection has been done using a three layer Artificial Neural Networks (ANN) using extracted parameters of acoustic signals. For the creation of the database, 150 prerecorded gunshots have been used. From the database, 80 gunshot sound samples have been used for the training of the system. Testing has been done with the remaining 70 samples in the presence of noise. The algorithm has also been tested successfully using actual gunshot in noisy environment. Efficiency of algorithm is 95 % without noise and that decreases to 85 % in the presence of noise.

Keywords Acoustics • MFCC • ANN • Gunshot detection • Association rules • Decision tree

---

P. Suman (✉) · S. Karan · V. Singh · R. Maringanti
Indian Institute of Information Technology, Allahabad, India
e-mail: preetam@iiita.ac.in

S. Karan
e-mail: subhadeepkaran@gmail.com

V. Singh
e-mail: vrij@iiita.ac.in

R. Maringanti
e-mail: mkrishna@iiita.ac.in

R. Maringanti et al. (eds.), *Proceedings of Ninth International Conference on Wireless Communication and Sensor Networks*, Lecture Notes in Electrical Engineering 299,
DOI: 10.1007/978-81-322-1823-4_15, © Springer India 2014

15.1 Introduction

Continuous surveillance of forests is a very difficult task for human beings. But the increased hunting, poaching, and deforestation activities make a strong case for real-time and automated surveillance system for the protection of wildlife and forests, and to identify illegal activities.

This paper presents an algorithm that was developed to detect the gunshot sound amidst forest noises and clutter. For this, 150 prerecorded gunshot sound samples [1] have been analyzed to extract the significant parameters required for the detection of gunshot (Analysis is described in Sect. 15.3). Gunshot consists of muzzle blast, shock wave, hammer mechanism during fire, and surface vibration due to muzzle blast. This paper focuses on detection of gunshot by analyzing muzzle blast sound. Feature extraction of gunshot sound has been done by Mel-Frequency Cepstrum Coefficients (MFCC) [2]. Of these 150 samples, 80 sound samples have been utilized to train a three-layer artificial neural network. The three-layers comprise of an input layer that takes sample signals, hidden layer to process input signal, compare with decision rules to identify the signal, and an output layer that produces the result [3]. After training of system, decision tree-based association rules were developed for testing of algorithm.

The Sect. 15.2 of this paper presents a short review of the literature with regard to gunshot sound characteristics, feature extraction of acoustic signal, identification, and localization of event, Sect. 15.3 presents analysis of signal using spectrogram, Sect. 15.4 presents features extraction using MFCC, Sect. 15.5 presents training of the system, Sect. 15.6 presents the association rule learning, Sect. 15.7 presents testing of the algorithm in the forest, and Sect. 15.8 concludes the paper.

15.2 Related Work

The literature is rich with a number of acoustic classification techniques.

Maher et al. [4] discussed some basic characteristics of gunshot like muzzle blast, mechanical action during firing (trigger and hammer mechanism, ejection of cartridge, etc.), supersonic projectile, and surface vibration due to muzzle blast. Each of these characteristics is useful to detect a gunshot. Author also discussed the effect of wind, humidity, temperature, and other ground obstacle on gunshot.

Chu et al. [5] presented the feature analysis for general environmental sound characterization using matching pursuit (MP) algorithm to establish time-frequency features. Authors of this paper combined the features of MP and MFCC for high accuracy in the recognition of environmental sound classification like: inside restaurants, playground, train passing, inside casinos, nature-daytime, ocean waves, raining/shower, thundering, etc. Authors of this paper implemented K-nearest neighborhood (KNN) and Gaussian mixture model (GMM) to classify the sound. Authors have collected 14 different environmental sounds to train and test the algorithm. The overall accuracy of recognition is 82.3 %, and it varied between 50 and 100 % for different environmental sounds. The most difficult environment sounds reported were ocean wave 63 %, sounds from movie 73 %, sounds inside casino 70 %, and traffic noise 74 %.

Freire [6] discussed use of correlation of the audio signal with predefined template (database of features of audio signals) to detect gunshot in a noisy environment. Three techniques linear predictive coding (LPC) coefficients, impulsivity parameter from stable distributions [7], and MFCC [2] have been used to extract the features of sound samples. The authors were able to detect the gunshot up to 30 dB SNR, and when noise is further increased to 25 dB SNR the system gives false positive result and when SNR was lowered 20 dB the system gives false negative results.

Ghiurcau et al. [8] discussed classification of different sounds originating from humans, cars, and birds to protect restricted areas like forest, lakes, natural parks, etc. The paper presents time-encoded signal processing and recognition (TESPAR) algorithm for sound classification. The Authors have created a database of 300 recorded sounds including several types of environmental noise like rain, wind, etc. The noises are recorded separately and then added to testing samples. Authors reported 94 % success to classify the sound in case of recorded sound without noise that decreases to 85 % with decreasing SNR.

Smith [9] proposed a solution for gunshot detection and localization. Location of the gun is determined by time of arrival information of muzzle blast. Time of arrival is determined by microphone embedded in JTRS radio, which acts as sensor node. Authors developed an algorithm that uses muzzle blast time of arrival information to determine the location of the shooter. The combination of correlation filters and rake receiver has been used to detect and localize gunshots. The correlation filter was used to remove uncorrelated surrounding noise from signal, and rake receiver use to eliminate multipath signals. The algorithm was tested in field with eight microphones and a GPS unit connected with laptops. The efficiency of algorithm was 90 % for actual gunshot and it has given 10 % false positive result. The testing with database (created by online available gunshots sound) was 96 %.

Aleksi [3] presented implementation of multichannel FPGA-based acoustic localization device (ALD) which can identify the direction of acoustic signal. The single board acoustic signal detector (ASD) used an electret microphone, and a signal conditioner using LM324N and a digitization unit. Eight ASDs were placed in different directions were connected to a single FPGA board through wire. The setup was tested in laboratory. A finite state machine was implemented on FPGA to recognize direction of acoustic signal. Due to simple processing, the acoustic resolution of the direction identification was not good.

15.3 Analysis of Gunshot Using Spectrogram

Spectrogram is a representation of time varying signal that shows variation of spectral density with time. The x-axis of spectrogram represents time, y-axis represents frequency, and different colors in spectrogram shows the intensity of signal. The spectral analysis of the sounds from pistol, rifle, Sniper, and shotgun are presented in the following figures.

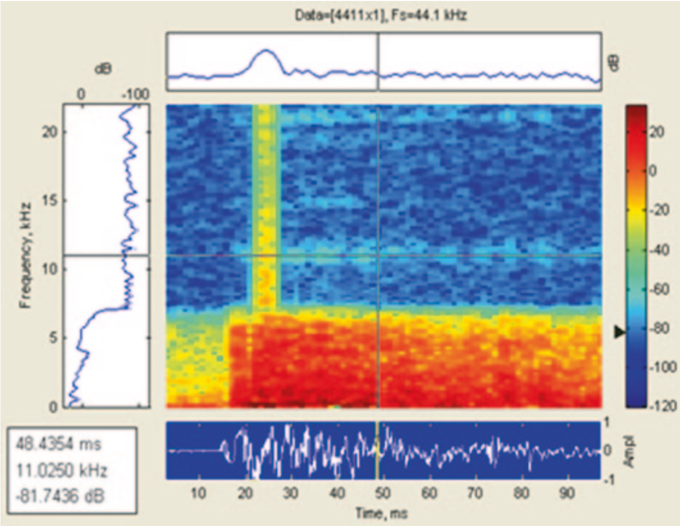

Fig. 15.1 Spectrogram of gunshot by pistol

In Fig. 15.1, it can be observed that the frequency of sound generated by pistol instantaneously increases then constant for short period (5 ms here) and after that it fall to constant amplitude.

In Fig. 15.2, it can be observed that the frequency of gunshot rises suddenly and decreases slowly. The time duration for gunshot is approximately 1 s in this case.

In Fig. 15.3, it can be observed that the frequency of gunshot by sniper decreases slowly after a peak but takes more time to decrease in comparison to the gunshot by rifle.

In Fig. 15.4, it can be observed that the frequency and intensity of gunshot remains constant for certain milliseconds, after that it starts decreasing.

These observations would help us to identify the event (gunshot) by looking for the signature of event. The identified signature and it values are presented in Table 15.2.

15.4 Features Extraction of Sound

Generally, acoustic signals are analyzed in terms of a set of features or parameters of interest and based on the parameters, the acoustic signal is classified. In this paper, mel-frequency cepstrum coefficients (MFCC) are used to extract 22 parameters of interest to gunshot, which are listed in Table 15.1 [5, 10]. The process steps of MFCC to extract the features of sound signals are as follows:

1. Estimation of amplitude and intensity of sound.
2. Grouping of frequency into bands of equal bandwidth.

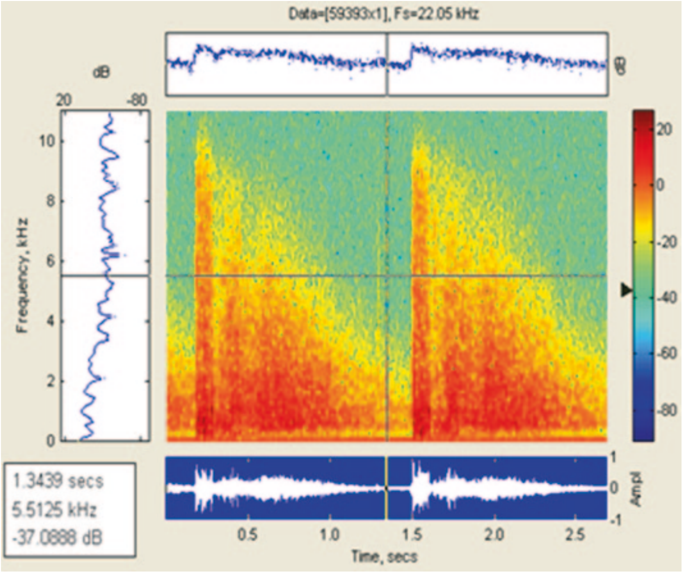

Fig. 15.2 Spectrogram of gunshot by rifle

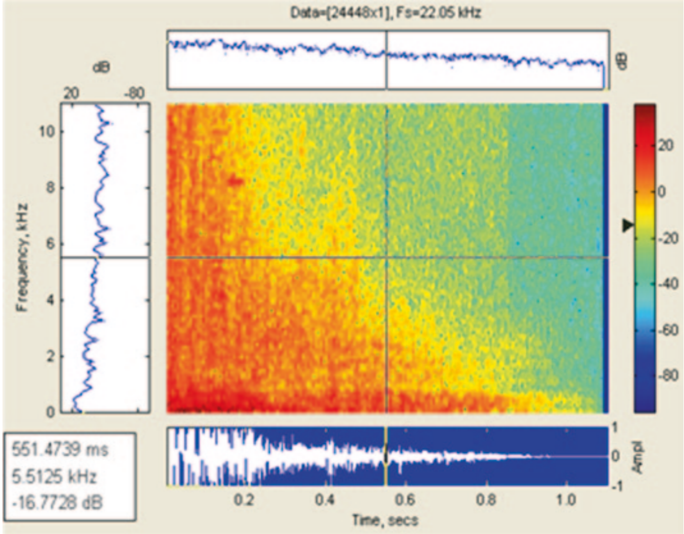

Fig. 15.3 Spectrogram of gunshot by sniper

3. Fast Fourier Transform (FFT) of bands.
4. Computation of the logarithm of bands after FFT.
5. Discrete cosine transforms (DCT) Computation of bands after computation of logarithm.

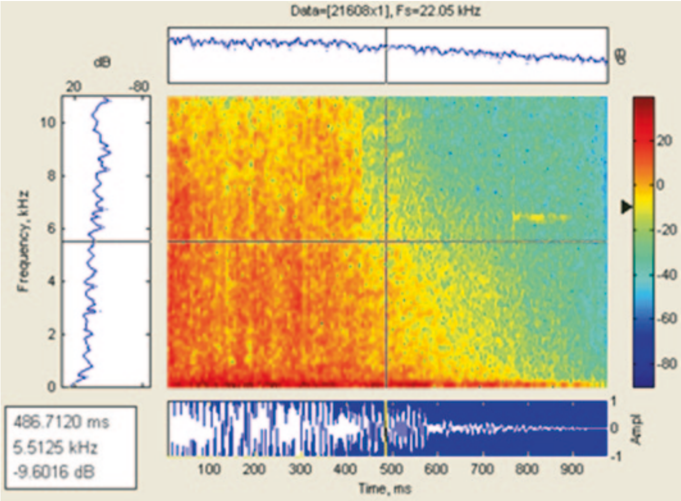

Fig. 15.4 Spectrogram of gunshot by shotgun

Table 15.1 Sound features extracted by MFCC technique

S. No.	Features description
1	Rise time, i.e., the duration of rise
2	Decay time
3	Mean square error of line adjust in 2
4	Slope of line fixed into rms-energy curve after rise
5	Crest factor, i.e., max/rms of amplitude
6	Time between the end of attack and the maximum of rms-energy
7	Mean of normalized spectral centroid
8	Maximum of normalized spectral centroid
9	Standard deviation of spectral centroid
10	Mean of spectral centroid
11	Frequency of amplitude modulation, range 4–8 Hz
12	Standard deviation of normalized spectral centroid
13	Strength of amplitude modulation, range 4–8 Hz
14	Frequency of amplitude modulation, range 10–40 Hz
15	Heuristic strength of the amplitude modulation in range 4–8 Hz
16	Standard deviation of rise times at each Bark band
17	Strength of amplitude modulation, range 10–40 Hz
18	Mean error of fit between each of onset intensities and mean onset intensity
19	Mean error of the fit between each of steady-state intensities and mean steady-state intensity
20	Overall variation of intensities at each band
21	Average cepstral coefficients during onset
22	Standard deviation of fundamental frequency

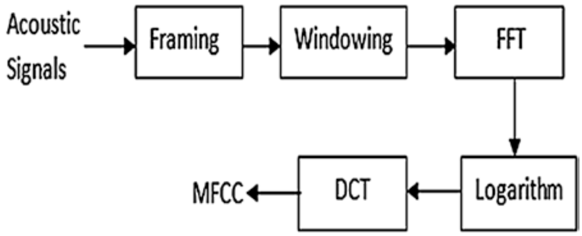

Fig. 15.5 Block diagram of MFCC process [5]

The process of parameter extraction using MFCC is shown in Fig. 15.5 as block diagram.

15.5 Training of System

The values of 22 parameters, that were discussed in the previous section were extracted from each sample signal using MFCC and are used to create a training set. Table 15.2 shows those 22 parameters that were extracted from three samples—two for pistols and one for rifle gunshot.

The extracted parameters were used for training gunshot recognition system using three-layer artificial neural networks (ANN).

The ANN uses back propagation model with momentum (Pattern Mode) learning rule. The input consists of 22 nodes, each node corresponding to one MFCC parameters, three hidden layers and one output node. The hidden layer has 11, 7, and 2 neurons Fig. 15.6. The Normalization of MFCC parameters was done in the range of 0.1–0.9.

15.6 Association Rules Learning

Association rules learning [11, 12] is a method to discover relationship between parameters of a large database. As an example, three parameters like tip in a hotel, quality of food, and quality of service could be linked through a relationship using association rules such as:

Good food + good service = good tip,
Good food + bad service = average tip,
Bad food + bad service = no tip.

After training of parameters, association rules make simple rules for decision making. Based on the association rules a decision tree (Fig. 15.7) is created for gunshot.

According to decision tree, decision rules were decided for gunshot detection.

The selected rules are the following:

Table 15.2 MFCC parameters extraction of gunshots

MFCC parameters	Pistol 1	Pistol 2	Rifle
MFCC 1	0.081929043	0.076333389	0.103508487
MFCC-2	0.345215906	0.317205015	0.450695428
MFCC-3	−0.024371645	−0.020849554	−0.016643999
MFCC-4	−0.005237523	−0.005423513	−0.004682361
MFCC-5	0.017424131	0.004221518	0.007569891
MFCC-6	0.01241868	0.008793181	0.001160881
MFCC-7	0.518305829	0.490094856	0.611755731
MFCC-8	0.032149527	−0.00358984	0.003835878
MFCC-9	−0.028921397	−0.021176352	−0.012461979
MFCC-10	−0.002598315	0.004691264	0.000558011
MFCC-11	−0.006948696	−0.001085565	−0.009483835
MFCC-12	0.521672874	0.494138025	0.615330665
MFCC-13	0.034800541	−0.002446353	0.003520145
MFCC-14	−0.028187497	−0.018725576	−0.011411904
MFCC-15	−0.001422691	0.004074864	0.001267848
MFCC-16	−0.00527637	−0.002008294	−0.01077973
MFCC-17	0.524168656	0.49807753	0.617844461
MFCC-18	0.03571343	−0.001066534	0.001942664
MFCC-19	−0.026175635	−0.014829641	−0.010374597
MFCC-20	0.000199536	0.004610543	0.00075864
MFCC-21	−0.002661296	0.000608799	−0.00731112
MFCC-22	0.040353661	0.040353661	0.039908543

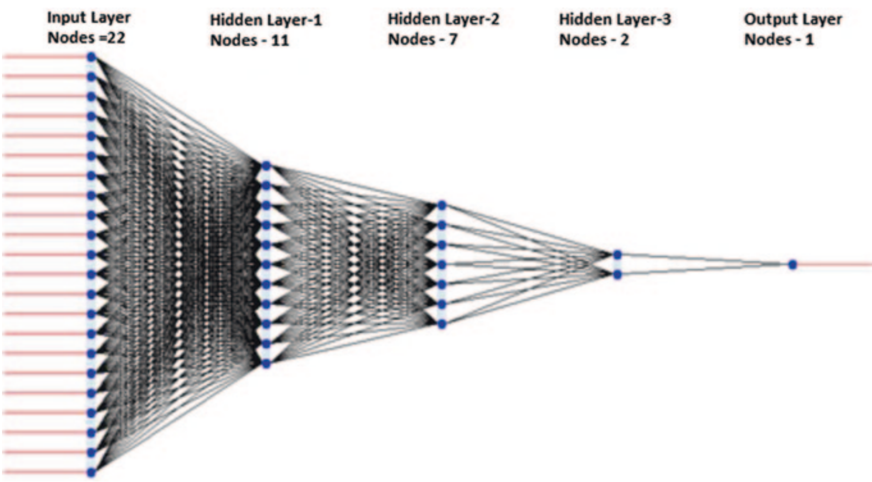

Fig. 15.6 Architecture of training model (Neurons: 22-11-7-2-1)

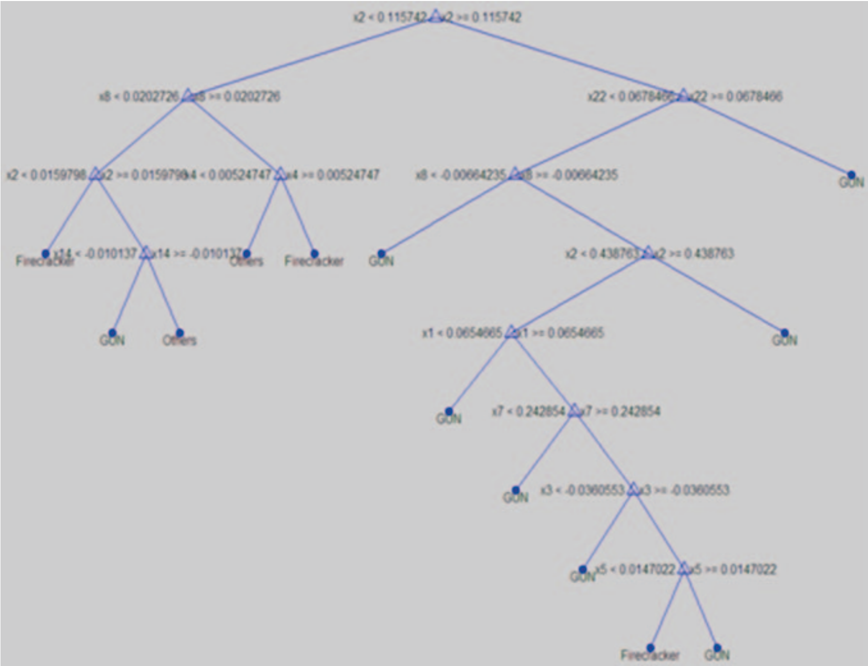

Fig. 15.7 Decision tree for gunshot detection based on MFCC parameters

Case 1

$x2 < 0.115742 \,\&\, x8 < 0.0202726 \,\&\, x2 \geq 0.01597984 \,\&\, x14 < -0.010137$

Case 2

$x2 > 0.115742 \,\&\, x22 > 0.0678466$

Case 3

$x2 > 0.115742 \,\&\, x22 < 0.0678466 \,\&\, x8 < -0.00664235$

Case 4

$x2 > 0.115742 \,\&\, x22 < 0.0678466 \,\&\, x8 > -0.00664235 \,\&\, x2 > 0.438763$

Case 5

$x2 > 0.115742 \,\&\, x22 < 0.0678466 \,\&\, x8 > -0.00664235 \,\&\, x2 > 0.438763$

Case 6

$x2 > 0.115742$ & $x22 < 0.0678466$ & $x8 > -0.00664235$ & $x2 < 0.438763$ & $x1 < 0.0654665$

Case 7

$x2 > 0.115742$ & $x22 < 0.0678466$ & $x8 > -0.00664235$ & $x2 < 0.438763$ & $x1 > 0.0654665$ & $x7 < 0.242854$

Case 8

$x2 > 0.115742$ & $x22 < 0.0678466$ & $x8 > -0.00664235$ &$x2 < 0.438763$ & $x1 > 0.0654665$ & $x7 > 0.242854$ & $x3 < -0.0360553$

Case 9

$x2 > 0.115742$ & $x22 < 0.0678466$ & $x8 > -0.00664235$ & $x2 < 0.438763$ & $x1 > 0.0654665$ & $x7 > 0.242854$ & $x3 > -0.0360553$ & $x5 > 0.0147022$

where $x1$, $x2$,…, $x22$ are the parameters extracted by MFCC. These rules were implemented on MATLAB, and then results were checked using real gunshot. The results are described in next section.

15.7 Testing in the Forest Area

The gunshot recognition algorithm was tested in a forest area. A Microphone connected to a laptop was used to capture the acoustic signals. Forest clutter and considerable noise as clapping was present when gunshot was fired and was detected successfully. The Graphical user interface (GUI) of the algorithm is shown in Fig. 15.8. The start button is to enable the system for gunshot detection. It can be put in continuous monitoring mode also, in which it will record and process the recorded sound simultaneously. The spectrogram of the acoustic signal is plotted as a graph. The area below the graph shows the result of detection. It displays either “Gunshot Detected” or “No Gunshot.”

Figure 15.9 shows the gunshot detection in heavy noise. The axis in GUI shows spectrogram of both gunshot and noise together. And it displays result as “Gunshot Detected.”

The algorithm is efficient to detect gunshot in forest environment; the actual gunshot testing is 95 % efficient with environmental noise which decreases up to 85 % decreased SNR. It generates 5 % false positive result for firecrackers (bomb). Algorithm has been tested with 70 gunshots in which 66 gunshots were detected. 50 gunshots were tested in the presence of noise in which 42 gunshots were detected.

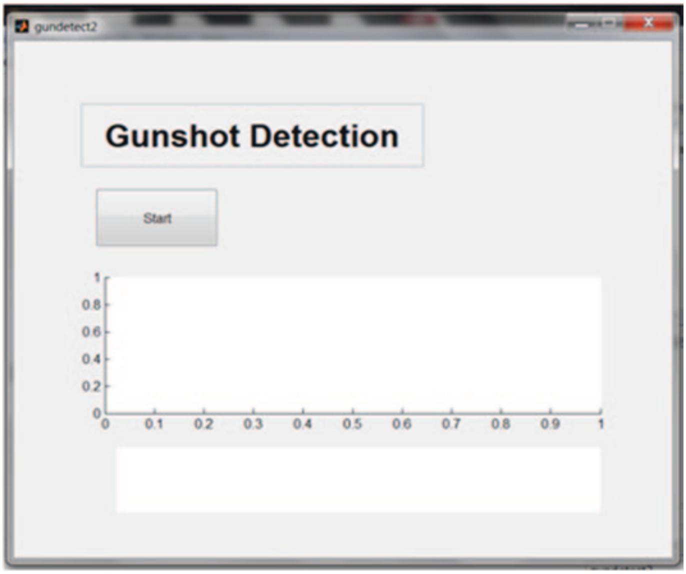

Fig. 15.8 GUI of gunshot detection system

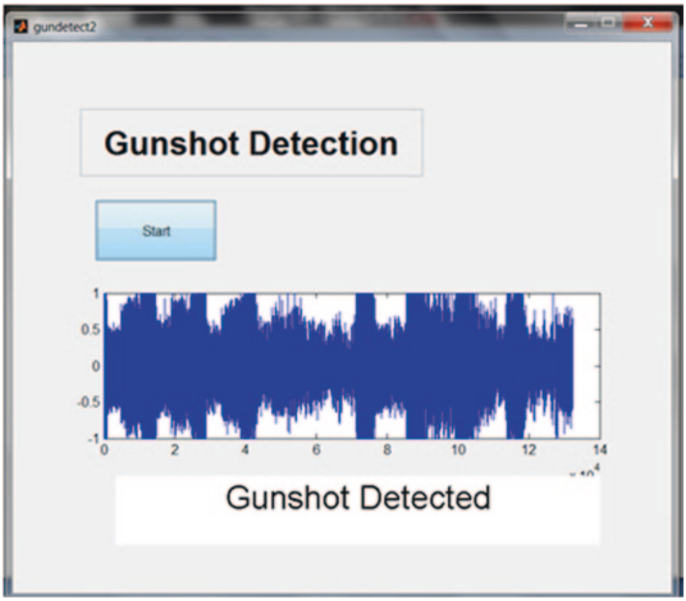

Fig. 15.9 GUI of gunshot detection in heavy surrounding noise

15.8 Conclusion and Future Work

This paper has proposed an algorithm to detect gunshot in noisy environment and forest clutter. MFCC was used to extract parameters of acoustic signals, three-layer ANN was used to train the system to recognize the gunshot implemented by the association rules using decision tree. After implementation, algorithm was tested with recorded gunshot and other sounds, and also with actual gunshot in the presence of noise and forest clutter. The algorithm is able to detect gunshot in heavy noisy environment. It generates false positive results for explosion of fire-crackers (bomb). The system gives false negative results for some shots of shot-gun, and also when SNR is low. The future work has to refine the algorithm to reduce false positives and remove false negatives, implement the algorithm on FPGA and build a robust sensor system, which could help in localization by compensating for wind and temperature effects.

Acknowledgments We thank the officials of Panna Tiger Reserve, Panna (M.P., India) and Wild Life Institute of India to provide environment for testing of gunshot detection algorithm. We also thank to Prof. Mehar Kayal, EPFL, Switzerland for his suggestions and support.

References

1. Graupe, D.: Principles of Artificial Neural Networks. World Scientific Publishing Co. Pte. Ltd, Singapore (2007)
2. Database of Free Sound Samples. Available at http://www.freesound.org. Accessed on 26 July 2012
3. Aleksi, I.: Acoustic Localization based on FPGA. Opatija, Croatia, 24–28 May 2010
4. Maher, R.C., et al.: Acoustical characterization of gunshots. In: IEEE SAFE (2007)
5. Chu, S., et al.: Environmental sound recognition with time frequency audio features. IEEE Trans. Audio Speech Lang. Process. **17**(6), 1142–1158 (2009)
6. Freire, IL.: Gunshot detection in noisy environments. In: IEEE 7th International Telecommunications Symposium (2010)
7. Hasan, M.R., et. al.: Speaker identification using mel frequency Cepstral coefficients. In: Proceedings of 3rd International Conference on Electrical and Computer Engineering (ICECE 2004), 28–30 Dec 2004
8. Ghiurcau, M.V., et al.: Wildlife intruder detection using sounds captured by acoustic sensors. In: IEEE ICASSP (1992)
9. Smith, M.: Gunshot detection system for JTRS radios. In: IEEE Military Communications Conference (2010)
10. Davisand, S.B., Mermelste, P.: Comparison of parametric representations for monosyllabic word recognition in continuously spoken sentences. IEEE Trans. Acoust. Speech Signal Process. **ASSP-28**(4), 357–366 (1980)
11. Huang, X., Acero, A., Hon, H.-W.: Spoken Language Processing: A Guide to Theory, Algorithms, and System Development. Prentice Hall, Upper Saddle River (2001)
12. Tan, P.N., Steinbach, M., Kumar, V.: Chapter 6. Association Analysis: Basic Concepts and Algorithms: Introduction to Data Mining. Addison-Wesley, Boston (2005). ISBN 0-321-32136-7