ISSN (Print) : 2320 – 3765
ISSN (Online): 2278 – 8875

International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Website: www.ijareeie.com

Vol. 8, Issue 3, March 2019

Audio Classification on Passing Vehicles with Feedforward Neural Network

Lwin Nyein Thu[1], Aung Win[2], Htet Ne Oo[3]

PhD-IT Candidate, Faculty of ICT, University of Technology (Yatanarpon Cyber City), Pyin Oo Lwin, Myanmar[1]

Rector, University of Technology (Yatanarpon Cyber City), Pyin Oo Lwin, Myanmar [2]

Lecturer, Dept. of IS, University of Technology (Yatanarpon Cyber City), Pyin Oo Lwin, Myanmar [3]

ABSTRACT: To establish an efficient real-time road traffic control, monitoring and forecasting system with road safety and urban security objectives are necessitated. The secure and proficient highway operations have the need for complete and accurate traffic data. At some stages in the night or foggy conditions and treatment of noise can be a very challenging task in some vehicle classification systems by using the vision sensors such as video camera, satellite, and thermal infrared, etc. The classification on acoustic signal is going to be a smaller amount cost, less computationally intensive and time-consuming than that of image and video signals. Therefore, this system is proposed to overcome such problems. In this research work, the development of the acoustic-based vehicle type classification system is utilized the Mel-Frequency Cepstral Coefficient (MFCC), its Deltas and Delta-Deltas as the main feature extraction method and a Multi-Layer Feedforward Neural Network (MLFFNN) for the classification on passing vehicle sounds. The system is able to objectively recognize the type of vehicles such as bus, car, motorcycle, and truck passing on the highway road with the specific acceleration.

KEYWORDS: MFCCs, Delta Coefficients, Vehicle Sound, Machine Learning, Neural Network.

I. INTRODUCTION

Recording sound is a straightforward process, but analyzing audio signal using a machine on behalf of a human is not as simple, and sometimes encounters unexpected difficulties. Even so, there are a lot of cases rely on the domain of interest. In the automobile vehicles, they generate significant sounds when they are passing or moving on the roadway with a specific speed. These sounds are composed of the tire-roadway interaction sounds, rattles of vehicle body, engine noise, and horns, etc. As per these environmental factors from passing vehicles, some researchers have attempted to carry out the vehicle classification based on audio signals associated with the type of vehicles.

Paulraj et al. [1, 2] developed for the moving vehicle recognition and classification system that is based on time domain approach with probabilistic neural network (PNN) and multi-classifier systems which are developed to classify the vehicle type and its distance. Nooralahiyan and Kirby [3] applied a directional microphone connected to a DAT (Digital Audio Tape) recorder. The digital signal was pre-processed by LPC (Linear Predictive Coding) parameter conversion based on autocorrelation analysis. A Time Delay Neural Network (TDNN) was chosen to classify individual travelling vehicles based on their speed independent acoustic signature to four broad categories: buses or Lorries, small or large saloons, various types of motorcycles, and light goods vehicles or vans. Michael N. J. and Andrew Woodward [4] considered that the two machine learning algorithms namely artificial neural networks (ANN) and naïve Bayesian classifiers (NBC) are compared to apply the audio samples captured from the vehicle engine sounds. The same features extraction using wavelet method has been also constructed by Amir Averbuch [5, 6] for classification and detection of the vehicle types.

Furthermore, several terrorist attacks around the world were appeared from starting at 2016 in Germany and its neighboring countries. Especially on 19th December 2016 in the Berlin Christmas Market Attack [7], more details are emerging about the victims of Monday night's attack at a Christmas market in Berlin when a hijacked truck barreled

ISSN (Print) : 2320 – 3765
ISSN (Online): 2278 – 8875

International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Website: www.ijareeie.com

Vol. 8, Issue 3, March 2019

into a crowd. This truck attack at Christmas market left 12 were killed and 48 people injured. Besides, on 2017 October 31st, a deadly vehicle attack near the World Trade Cente in Lower Manhattan was occurred [8]. That terror truck attack killed eight people and injured eleven in Manhattan, New York City. In relation to those phenomenons, a new system for monitoring, managing and controlling traffic in order to increase road safety and urban security is needed. Automated vehicle detection and recognition is one example, and a system that assists the task of vehicle detection and recognition is being considered here. These factors are also motivated to implement the audio classification on moving vehicles not only for protection of lives but also as a tool to improve infrastructure security. This paper describes the two major portions: feature extraction and classification. Among many feature extraction methods for classification, MFCCs are the most popular acoustic features used in audio identification. The MFCCs assume that the audio signal is stationary within a given time frame and may therefore lack the ability to analyze the localized events accurately. The Multi-Layer Feedforward Neural Network (MLFFNN) is used for classification and recognition of the type of vehicles.

II. SYSTEM METHODOLOGY

For audio classification, it was first extracted the features from the input signal. The main aim of the feature extraction step was to encapsulate the most relevant and discriminate attributes of the signal to acknowledge these features. In this system, MFCCs and its Delta energies are extracted as features of the passing vehicle classification.

***Mel-Frequency Cepstral Coefficients (MFCCs)*:** The stages of the MFCC, its Deltas and Delta-Deltas in the proposed system are demonstrated by Fig. 1 to perform the step by step execution. The description of MFCC computation with the relevant figures is also briefly explained below.

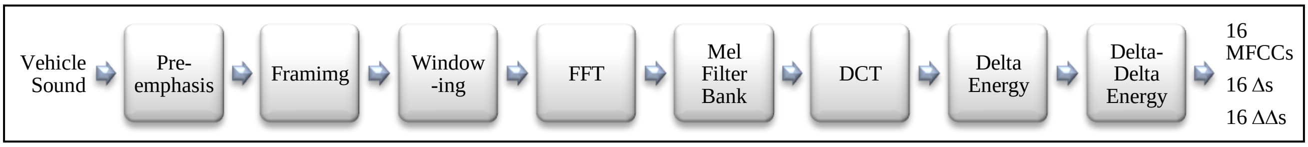

Fig. 1 Phases of MFCC

- **Pre-emphasis:** The first step of MFCC is to enhance the quantity of energy in the high frequencies. In this stage, the input vehicle signal is passed through a filter which produces higher frequencies. The pre-emphasis step of the bus vehicle is illustrated in Fig. 2.

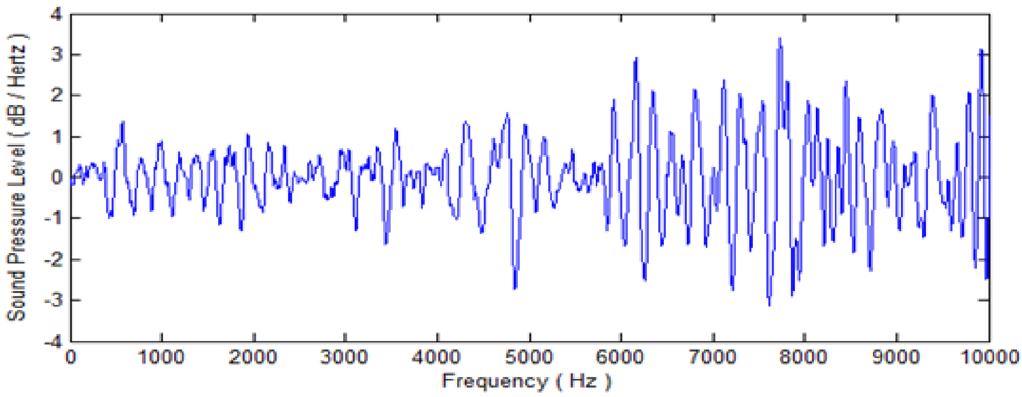

Fig. 2 Pre-emphasis step of MFCC

- **Framing:** The input signal is segmented into small duration blocks of 20-30 ms known as frames. Audio signal is divided into N samples and adjacent frames are being separated by M (M<N). Typical values for M=100 and

ISSN (Print) : 2320 – 3765
ISSN (Online): 2278 – 8875

International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Website: www.ijareeie.com

Vol. 8, Issue 3, March 2019

N=256. Framing is required as speech is a time varying signal but when it is examined over a sufficiently short period of time, its properties are fairly stationary. Therefore, short time spectral analysis is done [9].

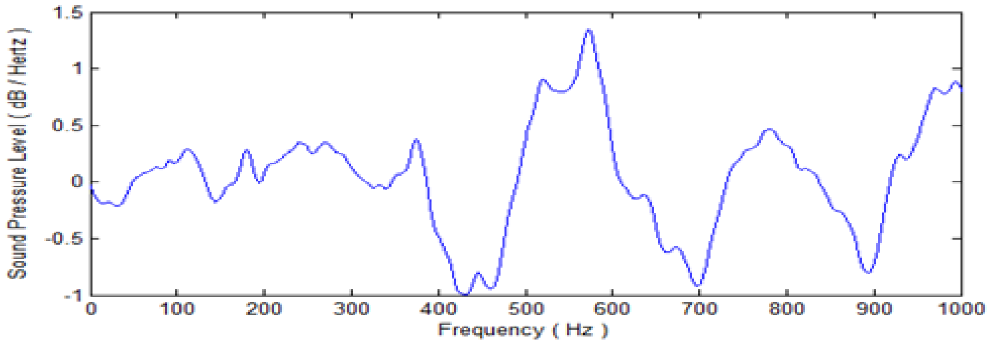

Fig. 3 Framing of the input signal

- **Hamming Windowing:** Each of the above frames is multiplied with a hamming window in order to keep continuity of the signal. So to reduce this discontinuity, window function is applied. Basically the spectral distortion is minimized by using window to taper the voice sample to zero at both beginning and end of each frame.

$$Y(n) = X(n) * W(n) \tag{1}$$

In Equation (1), where W(n) is the window function [9]. Indeed the simplest window is a rectangular window, but it can cause the problems because it abruptly cuts of the signal at its boundaries. These discontinuities create problems when Fourier analysis is done. For this reason, a more command window used in MFCC extraction is the Hamming window, which shrinks the values of the signal toward zero at the window boundaries, avoiding discontinuities.

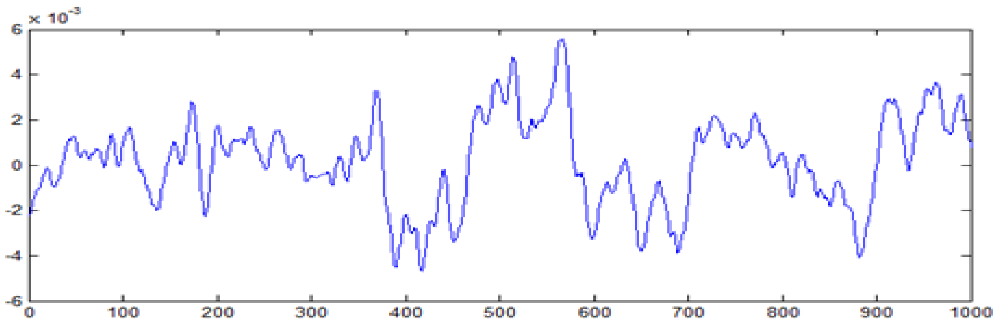

Fig. 4 Hamming Windowing of the signal

- **Fast Fourier Transform (FFT):** The next step is to extract spectral information for the hamming windowed signal; that is required to distinguish how much energy the signal contains at different frequency bands. The tool for extracting spectral information for discrete frequency bands for a discrete-time (sampled) signal is the Discrete Fourier Transform of DFT. A commonly used algorithm for computing the DFT is the Fast Fourier Transform (FFT). FFT is a process of transforming time domain into frequency domain to produce the magnitude frequency response of each frame. The output of processing FFT is a magnitude spectrum or periodogram of the signal (in fig. 5).

ISSN (Print) : 2320 – 3765
ISSN (Online): 2278 – 8875

International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Website: www.ijareeie.com

Vol. 8, Issue 3, March 2019

- **Mel Filter Bank:** The filter bank has a triangular bandpass frequency response, and the spacing as well as the bandwidth is determined by a constant mel-frequency interval. The frequencies range in FFT is very wide and voice signal does not follow the linear scale. To compute the Mel scale for the given frequency (f) in Hz:

$$m = 2595 * \log_{10}(1 + f / 700). \quad (2)$$

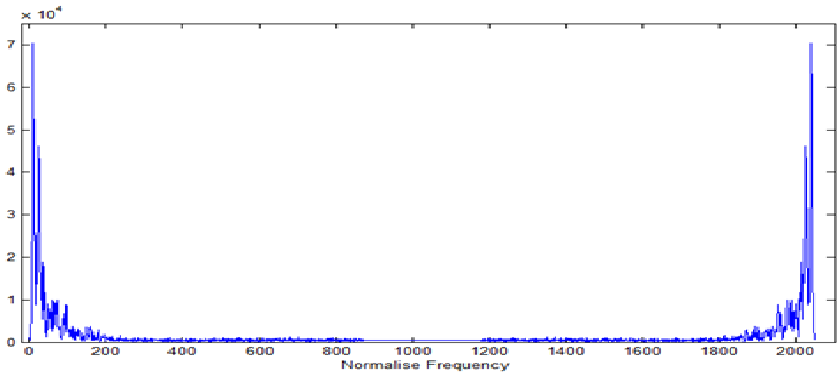

Fig. 5 Magnitude Spectrum of the signal

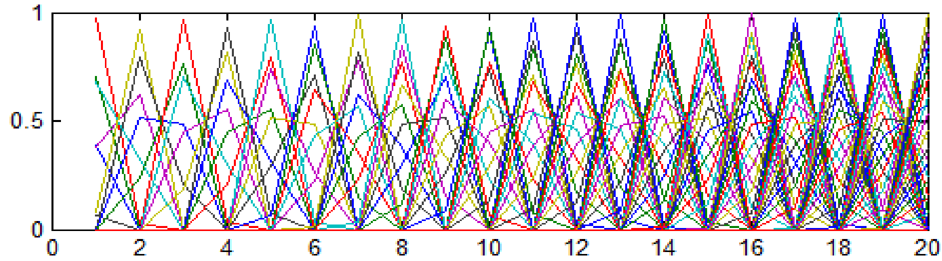

Fig. 6 Triangular filter bank with 20 band pass filters

- **Discrete Cosine Transform (DCT):** This is the process to convert the log Mel spectrum into time domain (time-like domain called quefrency domain) using DCT. In this final step, the log mel spectrum is converted back to time domain. The result is called the Mel Frequency Cepstrum Coefficients (MFCCs).

$$C(n) = \Sigma^{k}_{k=1}(\log S_k) \cos\{n(k - ½)* \pi/k\} \quad (3)$$

Equation (3) was that n = 1, 2,…, k, whereas S_k , k =1, 2, …, k are the outputs of last step.

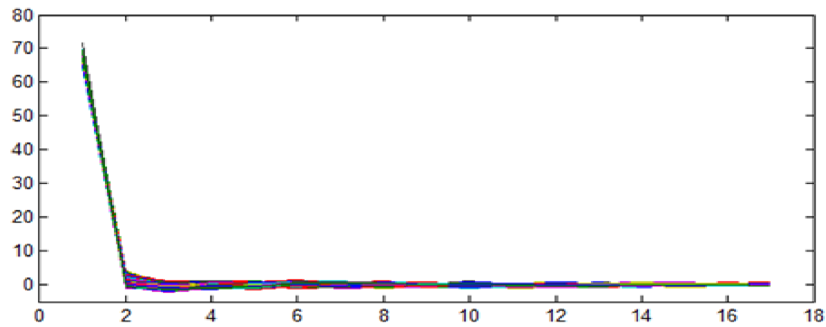

Fig. 7 Log filter bank energies to cepstral coefficients through DCT

- **Delta and Delta-Delta Energy:** Delta and Delta-Delta coefficients are also called differential and acceleration coefficients. In this proposed system, the first and second derivatives of the 16 feature vectors of MFCCs are

ISSN (Print) : 2320 – 3765
ISSN (Online): 2278 – 8875

International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Website: www.ijareeie.com

Vol. 8, Issue 3, March 2019

produced as the 16 Delta energies and 16 Delta-Delta energies. The following formula is applied to work out the delta coefficients:

$$d_t = \Sigma^N_{n=1} n(c_{t+n} - c_{t-n}) / 2\Sigma^N_{n=1} n^2 \quad (4)$$

where d_t is a delta coefficient, from frame t computed in terms of the static MFCCs c_{t+n} to c_{t-n}. A typical value for N is 2. Delta-Delta (Acceleration) coefficients are calculated in the same way, but these are calculated from the deltas, not the static MFCC coefficients.

***Classification*:** Working with the classification algorithms, the primary datasets are divided into training datasets and experimental (testing) datasets. The model is made by using training datasets and the experimental dataset is used for evaluation and computation of the model accuracy [10]. Some common classification methods are Support Vector Machine (SVM), k-Nearest Neighbor (k-NN), Genetic Algorithm (GA), Neural Network and Bayesian classification. The system is to exploit one of the machine learning algorithms: Multi-Layer Feedforward Neural Network architecture implemented to classifying the type of vehicle based on acoustic sounds. And, this is used in this system to recognize and identify the main sounds associated with the type of vehicles, is discussed in the following section.

***Feedforward Neural Network (FFNN)*:** Along with the machine learning area, a Feedforward Neural Network (FFNN) is a specific type of Artificial Neural Network (ANN) that is composed of a series of layers. These layers can be constructed by the first input layer, multiple hidden layers and an output layer. The information of neurons is always moved from the input layer to the output layer in one direction and never goes backward. Feedforward only refers to the topology of the network, the prorogation and learning algorithm used can vary. In this paper, one of the Artificial Neural Network architectures called the Multi-Layer Feedforward Neural Network (MLFFNN) is exploited as shown in Fig. 8.

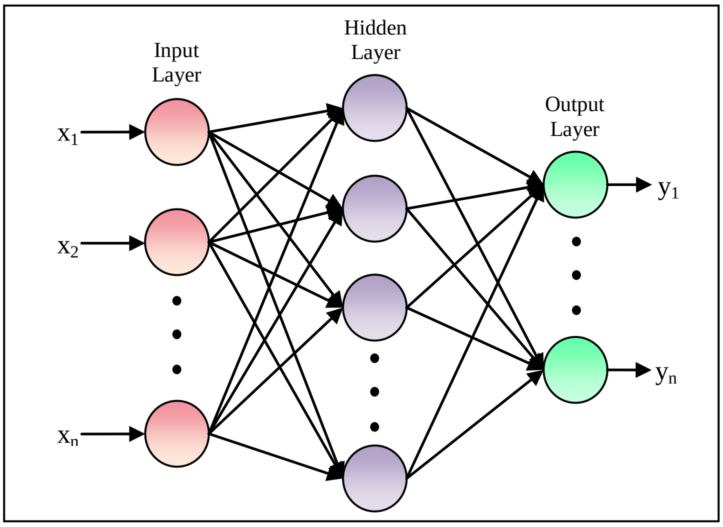

Fig. 8 The structure of a simple Multi-Layer Feedfoward Neural Network

III. PROCESS OF SYSTEM MODEL

The process model of the proposed system is partitioned into two main categories. Firstly, the input vehicle audio files are collected and then pre-processed to get a certain wave file (.wav) format. Secondly, significant features are extracted for both training and testing phase. In the training phase, the extracted features are stored in the datasets according to the types of vehicle. In the testing stage, the extracted features are classified by using the feature datasets. MFCCs and its Deltas, Delta-Deltas feature extraction method is used to extract the features for both training and testing phases. For the classification of vehicle sounds, Multi-Layer Feedforward Neural Network (MLFFNN) is used. The system is expected to classify the main sounds associated with vehicles such as bus, car, motorcycle and truck.

ISSN (Print) : 2320 – 3765
ISSN (Online): 2278 – 8875

International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Website: www.ijareeie.com

Vol. 8, Issue 3, March 2019

Data Collecting and Pre-processing: The acoustic signatures for vehicle type classification may be considered as a kind of pattern recognition process. It consists of data acquisition, feature extraction, and classification. Initially, the sound of passing vehicles are recorded with the assistance of Sony Cyber-Shot (16.2 mega pixels) camera which is placed on the Mandalay-Lashio-Muse asphalt highway (Asian Highway No.14 or AH14) roadside (as shown in Fig. 9). The acoustic signatures are also varying largely depending on speed, acceleration, gear position, and even the aspect angle of the sensor. However, vehicle acoustic signals are collected from different situations, in case, windy, rainy and also minimum background noise by the side of the road. The recorded video files are intended to facilitate the manual labeling of the four types of vehicles: bus, car, motorcycle and truck. In addition, it is also required to train classifier for labeling and analyzing the vehicle acoustics. Next, Format Factory application is used to convert the (.wav) file format from the recorded video file. The audio files are converted in separate WAV file (16 bits, mono) at a sampling rate of full 48 kHz without downsampling. For the reason that it appeared to have significant spectral characteristics observed in a high frequency range upon visual inspection on the spectrogram. Subsequently, all wave files are cut in the range about 5 to10 seconds (High Quality).

Fig.9 Data acquisition at Mandalay-Lashio-Muse asphalt highway roadside

IV. RESULT AND DISCUSSION

The whole system process has been implemented in the Matlab Platform. In support of vehicle type classification, twenty percent (100 out of 500) of the primary dataset is taken into account as the testing sample and eighty percent (400 out of 500) as the training sample according to these vehicle dataset. The numbers of vehicles observed and used in this analysis are depicted in Table 1. The system training method consists of three phases. By using MFCC, its Delta and Delta-Delta energies, the first feature dataset called, MFCC is used the analysis hamming window duration of 25 ms with 50% overlapping. The first 16 coefficients were kept, excluding the 0^{th} order coefficient. The second and third are MFCCD and MFCCDD correspondingly, these features are calculated using a window length of 9 frames. Thus, the result of feature vector is 32 and 48 dimensions, so it could be only differed from the number of coefficients compared with MFCC and other parameters are the identical. The developed feature dataset is then used to model a multi-layer feedforward neural network system trained by Levenberg-Marquardt algorithm since it acquires a smaller amount of training time. Thus, this algorithm also appears to be the fastest method for training moderate-sized feedforward neural networks (up to several hundred weights). Mean-Squared Error (MSE) function is utilized to measure the performance of network and error histrogram of the network training is also illustrated in Fig. 10 (a) and (b), respectively.

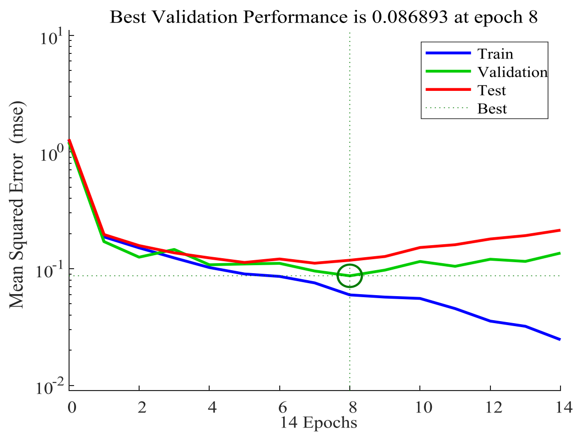

Fig. 10 (a) Network training performance validation

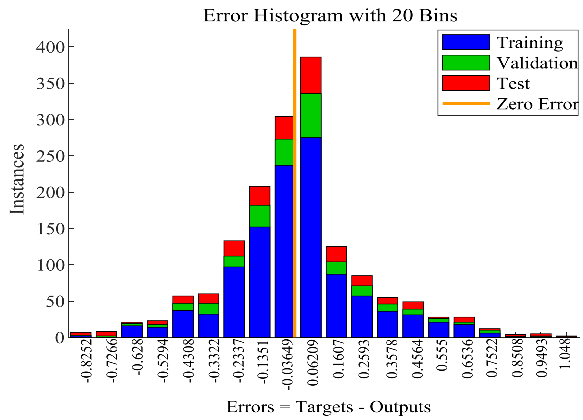

Fig. 10 (b) Network training error histrogram

Table 1. The number of vehicles observed in the proposed system

Type of Vehicle	No. of Training Samples	No. of Testing Samples
Bus	100	25
Car	100	25
Motorcycle	100	25
Truck	100	25

Table 2. The structure of MLFFNN Model in training phase

	MFCC	MFCC+Δ	MFCC+Δ+Δ
No. of Input Neurons	16	32	48
No. of Hidden Neuron (1st Layer)	25	25	25
No. of Hidden Neuron (2nd Layer)	15	15	15
No. of Output Neurons	4	4	4
Network Training Function	Levenberg-Marquardt Algorithm		
Network Performance Function	Mean-Squared Error (MSE)		
Ratio of Vectors (training, validation, testing)	0.7, 0.15, 0.15		
Maximum no. of Epochs	1000		

Table 2 describes the structure of neural network model that is developed in this system. The classification results for the developed MLFFNN network models are shown in Tables 3 corresponding to MFCC and its Delta, Delta-Delta coefficients. From Table 3, it can be observed that the neural network model tested using 20% data samples of MFCCs and its Deltas, Delta-Deltas (MFCC+Δ+ΔΔ) has better classification accuracy when compared to the MFCC coefficients only and MFCC and its Deltas, respectively. Therefore, the system chosen for the appropriate and better features of vehicle classification is the MFCC plus its Delta plus its Delta-Delta coefficients. Classification performance is measured using overall accuracy: the number of correctly classified segments among the total number of test segments, and then the evaluation of system performance (described in Table 4) is also estimated with the

ISSN (Print) : 2320 – 3765
ISSN (Online): 2278 – 8875

International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Website: www.ijareeie.com

Vol. 8, Issue 3, March 2019

precision, recall, specificity and F-score. In relation to the classification results on testing samples, the overall performance of the vehicle classification is got 82.0% with the representation of confusion matrix as shown in Fig. 11.

Table 3. The classification accuracy for each vehicle using MLFFNN

	Bus	Car	Motorcycle	Truck	Overall Accuracy
MFCC	68%	56%	88%	28%	**60%**
MFCC+Δ	80%	80%	88%	48%	**74%**
MFCC+Δ+Δ	80%	88%	92%	68%	**82%**

Table 4. The performance measure of the MLFFNN with MFCCDD testing features

	Precision	Recall	Specificity	F-Score
Bus	0.67	0.80	0.87	0.73
Car	0.92	0.88	0.97	0.89
Motorcycle	0.92	0.92	0.97	0.92
Truck	0.81	0.68	0.95	0.74

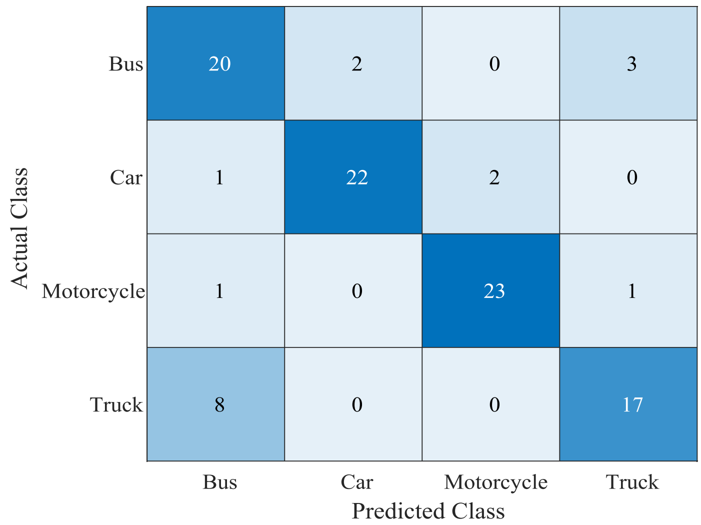

Fig. 11 Confusion matrix of the MLFFNN classifier evaluated on testing samples

V. CONCLUSION

The main aim of the system is to build a new vehicle sound dataset for acoustics-based classification. In this study, the input vehicle samples are employed while extracting the MFCC coefficients with the consideration of Delta energy function. As a result of adding the Delta and Delta-Delta coefficients, the MFCC feature extraction technique is more effective and robust. From the experimental results, it is inferred that the type and distance of moving vehicle can be identified using the Multi-Layer Feedforward Neural Network classifier. The performance of the system is obtained by 82.0% in the analysis of the experimentation. Although the truck vehicle has obtained a fewer accuracy result; the system can be able to adequately detect the other types of vehicle. Hence, it can be achieved the better accuracy upon

ISSN (Print) : 2320 – 3765
ISSN (Online): 2278 – 8875

International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Website: www.ijareeie.com

Vol. 8, Issue 3, March 2019

the all types of vehicle, the enhancement of input signals before extracting the features is required. As the future research activities, the proposed system will be able to extend the dataset to other sound events related with vehicles such as horn, siren of emergency vehicles, etc. Moreover, other deep structures of neural network model in deep learning algorithms will be implemented for the experimentation on the analysis of the proposed system. It is expected to become the better overall classification accuracy of the proposed system by constructing a huge dataset with more different vehicle features and different classifiers.

REFERENCES

[1] M. P. Paulraj, H. A. Abdul, S. Sathishkumar, B. A. R.Norasmadi, "Moving vehicle recognition and classification based on time domain approach," Procedia Engineering, vol. 53, pp. 405–410, 2013.

[2] N. Abdul Rahim, M. P. Paulraj, A. H. Adom, "Classification of moving vehicles using multi-classifier with time-domain approach," International Journal of Computer Applications (0975–8887), vol. 71, no. 1, pp. 12–17, Jun. 2013.

[3] Y. Nooralahiyan, H. R. Kirby, "Field trial of acoustic signature analysis for vehicle classification, " Transportation Research Part C: Emerging Technologies, vol. 5, Issues 3–4, pp. 165–177, Aug–Oct. 1997.

[4] Michael N. Johnstone, Andrew Woodward, "Automated detection of vehicles with Machine Learning ," Proceedings of the 11th Australian Information Security Management Conference, pp. 102–108, Dec. 2013.

[5] A. Averbuch, V. A. Zheludev, N. Rabin, and A. Schclar, "Wavelet-based acoustic detection of moving vehicles," Multidimensional Systems and Signal Processing, vol. 20, pp. 55–80, 2009.

[6] A. Averbuch, E. Hulata, V. Zheludev, and I. Kozlov, "A wavelet packet algorithm for classification and detection of moving vehicles," Multidimensional Systems and Signal Processing, vol. 12, pp. 9–31, 2001.

[7] "Berlin Christmas Market Attack Victims," https://edition.cnn.com/2016/12/23/europe/berlin-christmas-market-attack-victims/index.html, unpublished.

[8] "Terror Attack in Lower Manhattan ," https://www.nytimes.com/2017/10/31/nyregion/police-shooting-lower-manhattan.html, unpublished.

[9] P. S. Parwinder, R. Pushpa, "An approach to extract feature using MFCC," IOSR Journal of Engineering, vol.4, Issue 08, pp 21–25, Aug. 2014.

[10] A. H. Wahbeh¸ Q. A. Al-Radaideh¸ M. N. Al-Kabiand and E. M. Al-Shawakfa¸ "A comparition study between data mining tools over some classification methods," International Journal of Advanced Computer Science and Applications, vol. 35¸ pp 18–26, 2011.