ISSN (Print) : 0974-6846
Indian Journal of Science and Technology, *Vol 9(33), DOI: 10.17485/ijst/2016/v9i33/95628, September 2016* ISSN (Online) : 0974-5645

Environmental Noise Classification using LDA, QDA and ANN Methods

Safdar Tanweer*, Abdul Mobin and Afshar Alam

Jamia Hamdard (Hamdard University, Deemed to be University), Hamdard Nagar, New Delhi - 110062, Delhi, India; safdartanweer@yahoo.com, amobin_ansari@yahoo.com, aalam@jamiahamdard.ac.in

Abstract

Objective: The impact of feature based environmental noise classification and their efficiencies are discussed in this paper. The aim is to recognize, compare, classifyandidentify with the help of computation and applied techniques. **Methods/Statistical Analysis:** To perform the task of classification of noise sources, LPC and MFCC were used as input to the classifiers in experimental work. LDA, QDA and ANN are tested for the classification purpose. **Findings:** Once the source is identified we can address these untoward noisiness class and to minimize their impact to the human perceptions by some means or in speech recognition task to enhance the system recognition efficiency. The source can be identified by analysis of various categories of classifier in association with specific feature of noise source.LDA used with LPC gives an overall efficiency of classification is about 65.1% and with MFCC it is about 77.9%. QDA used with LPC gives an overall efficiency of 72.7% and with MFCC it is about 86.3%. ANN with LPC gives an overall efficiency of classification 83.2% and with MFCC it is about 90%. The MSE's (mean squared error) of ANN with MFCC are found to be 4.94838-e^{-2} (training), 5.33561-e^{-2} (validation) and 6.95805-e^{-2} (testing) and the %error for the same are 9.40265 (training), 10.02949 (validation) and 14.45427 (testing). **Application/improvement:** The performance of LDA, QDA and ANN with LPC and MFCC is analyzed. It is evident that ANN in combination with MFCC gives the best result and showing efficiency about 90%.

Keywords: ANN, Classifier, Environmental Noise, LDA, LPC, MFCC, QDA

1. Introduction

The noisiness degrades the performance of speechrecognition task, therefore by designing noise classifier that can classify the class of noise to address the effect[1]. Keeping in mind we have analyzed these problems that motivate us to develop an optimal solution to combat the over said problem. Most environmental noise investigations begin with measuring instruments using sound level meter, noise level analysis for measurement and analysis of environmental noise. These instrument are capable of measuring equally continue Level (L), Sound Exposure Level (SEL), Noise Potential Level (NPL), Traffic Noise Index (TNI), Day Night Average Sound Level (DNL) and Noise Equivalent Level (NEL)[2].

In the recent past, Noise Monitoring System (NMS) is pillared on electronics and digital processing technology. Although the computational efficiency is enhancing day by day, therefore NMS needs to be treated more sophisticated for noise data that they record to process. Research into the field of Automatic Noise Recognition (ANR) opens new window to develop new feature measurement for these noise data set to elaborate and identify noise source that are present in acoustic environment[3]. In our proposed model the problem of noise sources are categorized (viz. industry, bus, market). Features are extracted from the recorded noise samples that are then used by classifier to make a decision on the type of source of noise samples. There are various techniques so far available for classification. We have used LDA, QDA and ANN for the classification.

The major source of noise is car, bus, train, aircraft, market, office, industry etc. Several means and ways are used to investigate those noises by their proper classification. Once they are classified, static and dynamic volume controller can be employed that might suits particular category of noise to recognize[1]. The spectrum of some class of noise always remains constant with time in case of stationary, however they vary suddenly with rapid pace in case of non-stationary objects. Besides different

** Author for correspondence*

parameterization methodology in speech recognition we have selected two main categories LPC and MFCC for computation, although LPC possess many merits and few demerits in contrast to MFCC[4]. Three categories of noise samples are taken for the parameterization. Because of the presence of environmental noise at the front end input stage that severely degrade the performance of speech signal during the process of recognition. In this sense, implementation of recognition system to address type of noise source present and their contribution to the overall noise level measured. The utility of those systems might be of focal point of interest for long term measurement[5]. Research in general environmental noise classification has received some interest in the last two decades[6] but still considered in growing stage in comparison to speech, music, musical instrument recognition system.

Speech disfluency and stuttering assessment through classification were studied. The comparison of three feature vector MFCC, LPC and PLP were tested for classification of two types of speech disfluency, repetition and prolongation. k-NN, LDA and SVM were used to evaluate the performance and SVM with MFCC outperformed k-NN and LDA[7]. The Performance analysis of speech recognition were evaluated recognition of English word corresponding to digit 0-9 spoken by two speaker are evaluated MFCC and NN performance are used for recognition purpose[8]. Fore arm EMG (electromyogram) signal for wrist motion direction were collected using two channel EMG systems. The DAMV were used to construct feature map. LDA, QDA and k-NN algorithm verified for classification[9]. MFCC do not completely reflect the time varying feature of non-stationary non speech signal proposed 2D feature set based on pitch range PR for non-speech sound and autocorrelation function. The results are compared with MFCC using SVM and radial basis function NN[10]. Several multi class classifier based on GLDA (Generalized LDA) algorithm were proposed marginal LDA, Bayesion LDA and one dimensional LDA are introduced for classification matrix is directly utilized for multi class clasasification[11]. The performance of an empirical feature analysis for audio environment characterization and propose to use the Matching Pursuit (MP) Algorithm to obtain effective time frequency features. The MP-based method utilizes a dictionary of atoms for feature selection, resulting in a flexible, intuitive and physically interpretable set of features. The MP-based feature is adopted to supplement the MFCC features to yield higher recognition accuracy for environmental sounds[12].

2. Methods

2.1 Feature Extraction Methods

The purpose of extraction of important feature from the samples is to identify the original noise/speech samples with those features. By the introduction of such technique, complexity of processing is drastically reduced. There are various feature extraction methods available such as LPC, MFCC, PLP, LPCC etc., among them LPC and MFCC are tested in our experiment.

2.1.1 LPC (Linear Predictive Coding)

One of the most powerful speechanalysis methods is LPC (Linear Predictive Coding)[9]. In LPC, short term correlation between speech samples are modeled and removed by short order filters[10]. LP model uses past samples to predict the present state with the adjustment of weight[13]. It is a kind of model for processing of speech signal coding. Such samples are estimated as linear combination of past 'p' samples, where 'p' represents the LPC order.

LPC model[12] are sensitive to noise and similar kind of noise like variability. The most common form of spectral analysis model for speech frames and can be represented as

$$H(z) = 1/(a_1 z^{-1} + a_2 z^{-2}+ \qquad (1)$$

H(z) = pth order polynomial with z-transform coefficients a_1, a_2-----------a_p, are constant through out the speech frame.

Using excitation G u(n), we get

$$s(n) = \Sigma a_i s(n-i) + G\, u(n) \text{ where } i= 1\ldots\ldots\ldots p \qquad (2)$$

Where u (n) is normalized excitation, G is the gain of excitation and s(n) is approximated as linear combination of 'p' past samples.

Expressing Equation (2) in z domain

$$S(z) = \Sigma a_i z^{-i} S(z) + G \qquad (3)$$

Where i = 1…………p

Yields to transfer function

$$H(z) = G/(1- \Sigma a_i z^{-i}) \quad (4)$$

Speech signals are highly non-stationary in nature the parameters of speech signals vary with time, however individual frames are stationary. We have selected 20 ms frames for the parametrisation.

2.1.2 MFCC (Mel Frequency Cepstral Coefficient)

Mel Frequency Cepstral Coefficient (MFCC) was designed to adapt human perception. During computation of speech signal pass through linear filters which are passed linearly over Mel scale[14]. MFCC also have been process to be one as the successful feature extraction method in speech disfluencies[7]. A detail description of this process with block diagram and algorithm can be found in many papers[4,8]. Feature extraction is categorized into Frequency based feature extraction and Time frequency based feature extraction. They fall in the category of stationary feature extraction. However time frequency feature extraction is non-stationary[11]. MFCC is characterized in frequency domain since most environmental noise is non-stationary in nature. In this paper, experimentally analyzed and parameterized coefficients of three categories of noise viz. industry noise, bus noise and market noise. The computation algorithm of proposed MFCC consist of five steps[8,10] summarized to formulate the problem addressed:

- Frame blocking: since speech signal is a continuous signal therefore by proper analyzing the behavior of speech signal, their parameter are divided into frames.
- Windowing: Speech signals are highly non stationary in nature and the previous study shows that their parameter changes normally at every 20 ms approx so we are windowing the frames at energy 20 ms. However, in case of MFCC, these are logarithmic in nature.
- FFT: It is very efficient tool in signal processing for transforming time domain signal into frequency domain. It is computationally very efficient to calculate DFT of any signal. It greatly saves processing time and power.
- Mel filter bank transformation: Mel filter Bank is a logarithmic scale which resembles the Human auditory system that is also logarithmic in nature and is very robust in speech detection.
- Mel(f) = 2595log (1+f/1000), Where f = frequency of noise signal.
- DCT: DCT applied for Mel filter Bank to obtain MFC. It minimizes the distortion in frequency domain.

2.2 Classifiers

The classifier classifies the input data according to some rule. The success of classifier largely depends upon the distribution, density, size and type etc. of input data set that is to be classified. The data can be linearly separable or not. According to the nature of data set the rule of classification would be decided. Various types of classifiers are available such as GMM, HMM, ANN, LDA, QDA etc., are proved to be good classifier. We have studied, analyzed and compared the performance of LDA, QDA and ANN in our experiment.

2.2.1 ANN (Artificial Neural Network)

ANN is inspired from human neural network also known as biological neural network. In their structure, billions of neurons are connected with trillions of synapsis to address complex computation task. The ANN based pattern classifier works well to the input data presented[6]. ANN is made up of man-made neural network that is the replica of very small section of the biological NN. The structure is comprises of three layers[9]: Input layer, hidden layer and output layer. A simple architecture of an ANN is as shown in Figure 1.

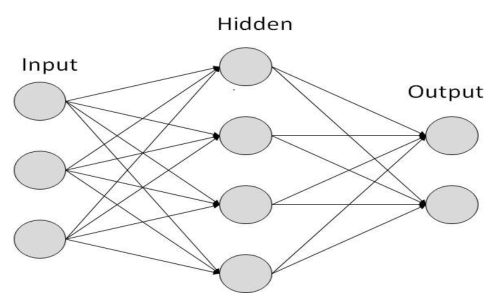


Figure 1. A simple architecture of an ANN.

Input layer is excited by the input data which is then transferred to hidden layer. It processes the information and feds it to the output layer. ANN can be explained by considering the following sequence of parameters:

- Network Architecture.
- Weight Adjustment.
- Activation Function.
- Network Architecture: The size of architecture

depends upon the amount and type of data. The network is expected to process. Accordingly neurons are adjusted doing so pattern/mode of connection within and in between various layers can be marked as architecture of neural network. There are various modes of representation such as feed forward network, feedback network, fully interconnected network, competitive network.

- Weight Adjustment: During the course of learning and training the synaptic weights are varied between neurons to get expected output. This training algorithm is termed as learning.
- Activation function: for getting optimum response of network the collective sum of weighted input signal is presented for activation function on the basis of that they calculate output response. They are classified into linear and non-linear activation function. The linear activation function is simple and generally used for single layer network however non-linear activation function is widely used for multilayer complex network.

Now days, they are widely used in speech processing, pattern recognition because they have adaptive learning ability based on input data provided for training. They organized them for parallel computational work as the biological neural networkdoes for real time applications.

2.2.2 LDA (Linear Discriminative Analysis) and QDA (Quadratic Discriminative Analysis)

The objective of LDA is to perform dimensionality reduction while preserving as much class discriminatory information as possible[7]. It is a statistical technique to classify objects into mutually exclusive and exhaustive group based on a set of measurable objects feature. The solution proposed by Fisher is to maximize a function that represents the difference between the means normalized by a measure of within class scatter. Linear Discriminant Analysis (LDA) is a generalization of Fisher's linear discriminant, a method used in statistics, pattern recognition and machine learning to find a linear combination of features that characterizes or separates two or more classes of objects or events. A classification rule they are represented by max gate function g(X)[9]. The πi is the prior probability of class i with conditional density of x in class i is f(x), wherex is multi-variable. μ_i is uniformly distributed group matrix Σ for LDA and Σi is for QDA denotes group specific with assumption gi(X)>gj(X), for j≠ i. Bay's rule explored to get linear and quadratic discriminant function[7,9,12].

$$f_i(X)=\frac{1}{(2\pi)^{p/2}\,|\Sigma_i|^{1/2}}\exp\left[-\frac{1}{2}(X-\mu_i)^T\sum_i^{-1}(X-\mu_i)\right]$$

-Linear discriminant function

$$g_i(X)=X^T\sum^{-1}\mu_i-\frac{1}{2}\mu_i^T\sum^{-1}\mu_i+\log(\pi_i)$$

-Quadratic discriminant function

$$g_i(X)=-\frac{1}{2}(X-\mu_i)^T\sum_i^{-1}(X-\mu_i)-\frac{1}{2}\log\left(|\sum_i|\right)+\log(\pi_i)$$

Where P is a dimension factor (1 for LDA and 2 for QDA), and T is a transpose operator.

3. Result and Discussion

In our experiment noise source is classified by using LDA, QDA and ANN. Three types of environmental noise are considered: industry, bus, and market. The noise features are used for training validation and testing purposes. Features (LPC, MFCC) are extracted using PRAAT software, and MATLAB is used for classification. 12 LPC and 13 MFCC coefficients are evaluated from the chosen noise samples (10 of each class) of equal size of 600 ms. Two different features (LPC and MFCC) are computed for the identification of noise source, considered the frame size is of 20 ms for parameterization.

Experimental procedure carried out as follows:

- Noise samples are recorded at three different places viz. industry, bus, market.
- Input samples of 3 categories of noiseshave 10 samples, each sample is of 600 ms duration (for the ease of experiment).
- Window size is kept of 20ms duration.
- Feature vectors:
  - LPC feature extraction: Each sample has 12 feature vectors.
  - MFCC feature extraction: Each sample has 13 feature vectors.
- Classification methods adopted
  - LDA
  - QDA
  - ANN

The experimental result shows the overall efficiency of the LDA classifier is 65.1%, when 12 LPC coefficients are used for the training, validation and testing data set for the classifier. Table 1 shows the confusion matrix for the three classes of noises which are in order as ind, bus, mkt. A total of 3390 elements are used for the training, testing and validation of the classifier in which 1130 elements belongs to the each class of noise sample (during the experimental work).

Table 1. Confusion matrix of LDA using LPC

Target / Output	Ind.	Bus	Mkt.	Overall
Ind.	766	218	146	67.8%
	22.6%	6.4%	4.3%	32.2%
Bus	76	844	210	74.7%
	2.2%	24.9%	6.2%	25.3%
Mkt.	200	334	596	52.7%
	5.9%	9.9%	17.6%	47.3%
Overall	73.5%	60.5%	62.6%	65.1%
	26.5%	39.5%	37.4%	24.9%

LDA is used with MFCC gives an improvement in the efficiency as compared to LDA with LPC. In this case overall efficiency is found to be 77.9%. The result of the classifier is represented in the form of confusion matrix as shown in Table 2.

Table 2. Confusion matrix of LDA using MFCC

Target / Output	Ind.	Bus	Mkt.	Overall
Ind.	908	83	139	80.4%
	26.8%	2.4%	4.1%	19.6%
Bus	43	920	167	81.4%
	1.3%	27.1%	4.9%	18.6%
Mkt.	149	168	813	71.9%
	4.4%	5%	24%	28.1%
Overall	82.5%	78.6%	72.7%	77.9%
	17.5%	21.4%	27.3%	22.1%

QDA is used with LPC (of the same database as for LDA) to give an overall efficiency 72.7% as clear from the confusion matrix shown in Table 3. The result shows it is better classifier in contrast to LDA with LPC.

Table 3. Confusion matrix of QDA using with LPC

Target / Output	Ind.	Bus	Mkt.	Overall
Ind.	815	168	147	72.1%
	24.0%	5.0%	4.3%	27.9%
Bus	41	1001	88	88.6%
	1.2%	29.5%	2.6%	11.4%
Mkt.	178	302	650	57.5%
	5.3%	8.9%	19.2%	42.5%
Overall	78.8%	68.0 %	73.4%	72.7%
	21.2%	32.0%	26.6%	27.3%

The confusion matrix for QDA is shown in Table 4, when it uses MFCC for the training validation and testing dataset for the classifier. It shows an overall improved efficiency of 86.3%. It gives better result in our experiment as compared with LDA.

Table 4. Confusion matrix of QDA with MFCC

Target / Output	Ind.	Bus	Mkt.	Overall
Ind.	979	52	99	86.6%
	28.9%	1.5%	2.9%	13.4%
Bus	23	1033	74	91.4%
	0.7%	30.5 %	2.2%	8.6%
Mkt.	80	138	912	80.7%
	2.4%	4.1%	26.9%	19.3%
Overall	90.5%	84.5%	84.1%	86.3%
	9.5%	15.5%	15.9%	13.7%

ANN classifier used with the LPC (with the same data as for LDA and QDA) gives an overall efficiency of 83.2% as shown in Figure 2(a). The performance curve shows the best validation performance is at 157 epoch as clear from Figure 2(b). Mean Squared Error (MSE) is the squared average difference between outputs and targets. In Figure 2(c) training state of ANN is shown. Error histogram is shown in Figure 2(d). Error is the difference between outputs and targets.ROC (Receiver Operating Characteristics) of the ANN is shown in Figure 2(e).

ANN classifier used with the MFCC (with the same data as for LDA and QDA) gives an overall efficiency of 90% as shown in Figure 3(a). The performance curve shows the best validation performance is at 113 epoch as

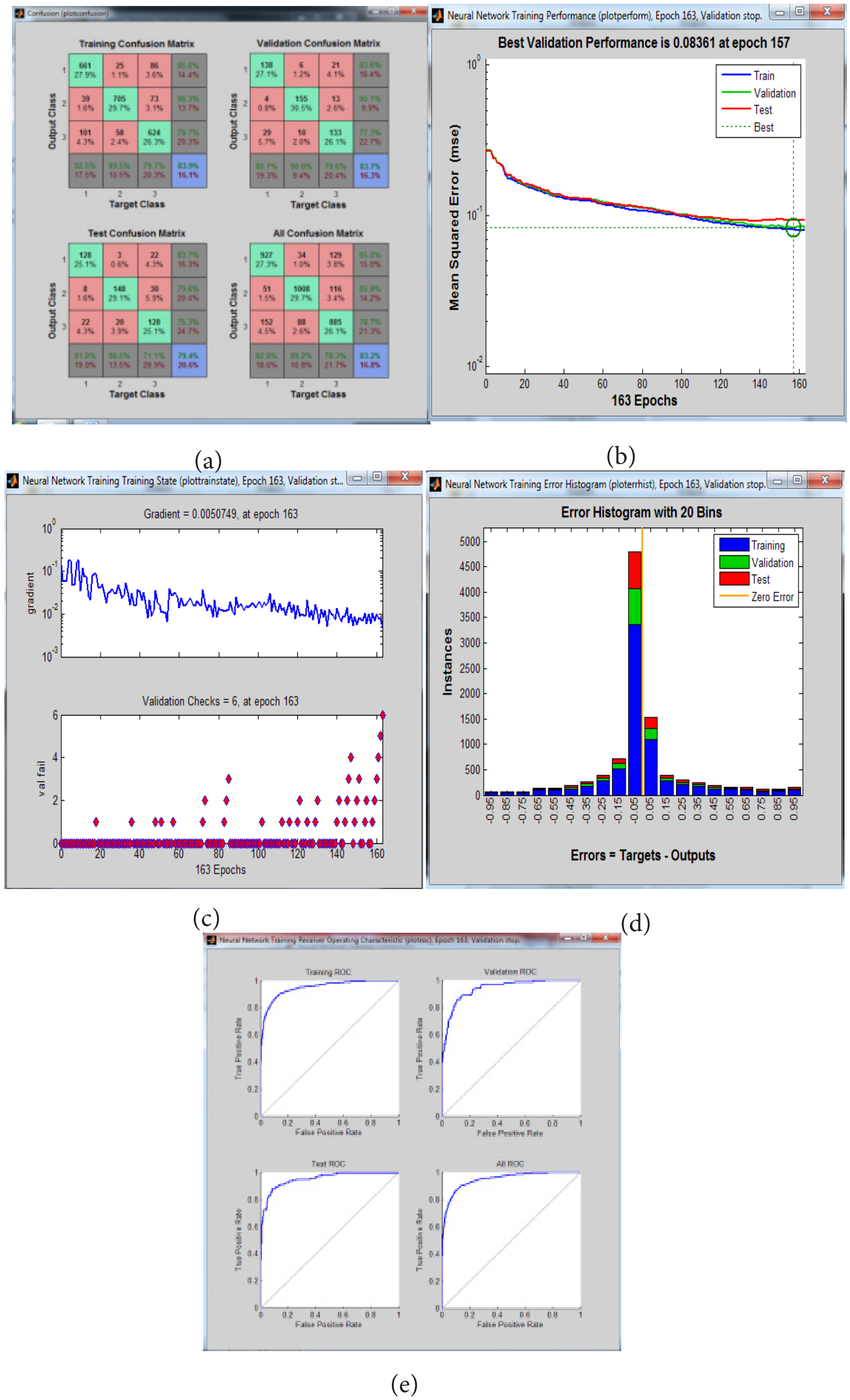
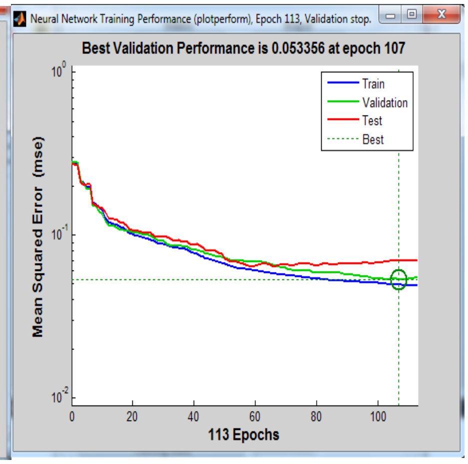
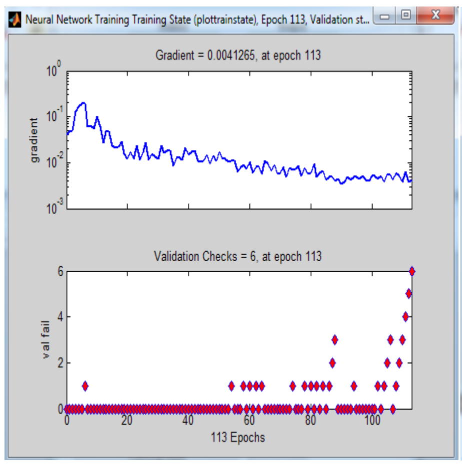
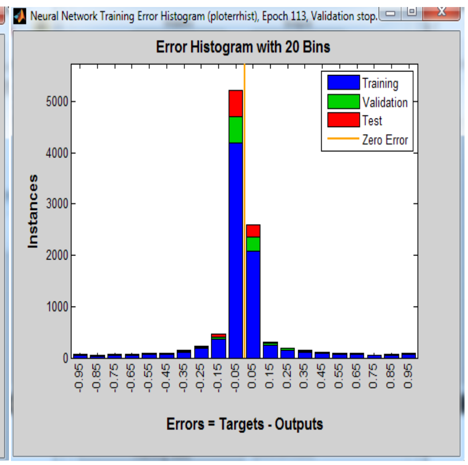


Figure 2. Different curves and diagram of the ANN when LPC is used as training, validation and testing dataset. (a) Confusion matrix. (b) Performance curve. (c) ANN training state. (d) Error histogram. (e) ROC.

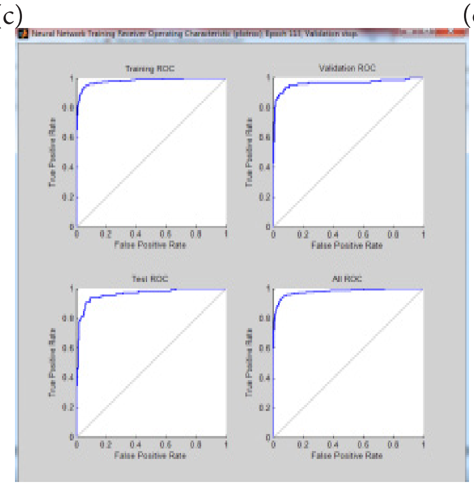


(a) (b)





(c) (d)



(e)

Figure 3. Different curves and diagram of the ANN when MFCC is used as training, validation and testing dataset. (a) Confusion matrix. (b) Performance curve. (c) ANN training state. (d) Error histogram. (e) ROC.

Table 5. Comparison of the performances in terms of efficiency for the different types of classifier with combination of feature vector for the three categories of noises

Classifier	LDA		QDA		ANN	
Noise Class	LPC	MFCC	LPC	MFCC	LPC	MFCC
Industrial	22.6%	26.78%	24.04%	28.87%	27.3%	30.3%
Bus	24.9%	27.13%	29.52%	30.47%	29.75	30.6%
Market	17.6%	23.98%	19.17%	26.90%	26.1%	29.2%
Overall efficiency	65.16%	77.89%	72.73%	86.24%	83.2%	90.0%

clear from Figure 3(b). Mean Squared Error (MSE) is the squared average difference between outputs and targets. In Figure 3(c) training state of ANN is shown. Error histogram is shown in Figure 3(d). Error is the difference between outputs and targets. ROC (Receiver Operating Characteristics) of the ANN is shown in Figure 3(e).

The entire experimental results are categorically placed into two parts. The first category is in the direction to find feature of the noise samples; among them two methods (LPC and MFCC) are analyzed. MFCC is found to be of better feature than LPC for noise classification as shown in Table 5. Second category of result depicts the classification of the noise source. Three methods (LDA, QDA and ANN) are analyzed on the same dataset. In the comparative analysis, it is evident that ANN dominates over the two classifiers as shown in Table 5.

4. Conclusion

The novelty of the paper is the usage of classifiers LDA, QDA and ANN for noise classification. The efficiencies of QDA and ANN are quite acceptable for the classification purpose. It is evident from the experimental results that the MFCC as feature of noise outperformed LPC for all the classes of noise as well as the methods adopted for the classifications and ANN as classifier for the classification of noise sources is the best among all the three classifiers used for the classificationas per our experiment. The maximum efficiency achieved is about 90% using ANN with MFCC, which can be further improved by adding a larger database of the similar kind of noise samples. We can add more class of noise sources for the classification as much as possible for the future prospects.

5. References

1. El-Maleh K, Samouelian A, Kabal P. Frame level noise classification in mobile environment. Proc IEEE Conf Acoustics, Speech, Signal Proc; phoenix AZ. 1999. p. 237–40.
2. Couvreur C, Fontain V, Gaunard P, Mubikangiey CG. Automatic classification of environmental noise event by Hidden Markov Model. Elsevier Applied acoustics. 1998; 54(3):187–206.
3. Zouhir Y, Ouni O. A bio inspired feature extraction for robust speech recognition. Springer Plus. 2014; 3:651.
4. Natrajan VA, Jothilaksmi S. Segmentation of continuous Tamil speech into syllable like units. Indian Journal of Science and Technology. 2015; 8(17):1–5.
5. ValeroX, Alias F. Hierarchical classification of environmental noise sources considering the acoustic signature of vehicle pass-bys. PAN archives of Acoustic. 2012; 37(4):423–34.
6. Gallant PJEL, Morin LE, Peopard P. Feature based classification of myoelectric signal using artificial neural network med. Boil Eng Comput. 1998; 36(4):485–9.
7. Fook CY, Muthusamy H, Chee LS, Yaacub SB, Adom AHB. Comparision of speech parameterization technique for the classification of speech dysfluencies. Turkish Journal of Electrical Engineering and Computer Science. 2013; 21:1983–94.
8. Tanweer S, Mubin A, Alam A. Analysis of combined use of NN and MFCC for specch recognition. International Science Index Computer and Information Engineering. 2014; 8(9):1–5.
9. Kim KS, Choi HH, Moon CS, Mun CW. Comparision of k nearest neighbor quadratic discriminant and linear discriminant analysis in classification of electromyogram signals based on wrist motion direction. Elsevier BV, Current Applied Physics. 2011; 11(3):740–5.
10. Chu S, Narayanan S, Kuo CJ. Environmental sound recognition with time frequency audio features. IEEE Transaction on Audio, Speech and Language Processing. 2009; 17(6):1142–58.
11. Uzkent B, Barkana BD, Cevikalp H. Non speech environmental sound classification using SVMS with a new set of features. International Journal of Innovative Computing, Innovation and Control. 2012; 8(5):841–8.
12. Kim H, Drake BL, Park H. Multiclass classifiers based on dimension reduction with generalized LDA. Elsevier Ltd Pattern Recognition. 2007; 40(11):2939–45.
13. Mohammed A, Masour A, Ghulam M, Mohammed Z, Mesellam TA, Malki KH, Mohamed F, Mohamed B. Automatic speech recognition of pathological voice. Indian Journal of Science and Technology. 2015; 8(32):1–6.
14. Sahidullah M, Saha G. Design analysis and experimental evaluation of block based transformation in MFCC computation for speaker recognition. Elsevier, Speech Communication. 2012; 54(4):543–65.