



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

(An ISO 3297: 2007 Certified Organization)

Vol. 2, Issue 11, November 2013

AN APPROACH TO FEATURE SELECTION ALGORITHM BASED ON ANT COLONY OPTIMIZATION FOR AUTOMATIC SPEECH RECOGNITION

C.Poonkuzhali¹, R.Karthiprakash¹, Dr.S.Valarmathy², M.Kalamani³

PG Student [AE], Dept.of ECE, Bannari Amman Institute of Technology, Sathyamangalam Tamilnadu, India¹

Prof. and Head, Dept.of ECE, Bannari Amman Institute of Technology, Sathyamangalam, Tamilnadu, India²

Asst. Prof. (Sr.G), Dept.of ECE, Bannari Amman Institute of Technology, Sathyamangalam, Tamilnadu, India³

ABSTRACT: Speech is one of the most promising models by which people can express their emotions like anger, sadness, and happiness. These states can be determined using various techniques apart from facial expressions. Acoustic parameters of a speech signal like energy, pitch, Mel Frequency Cepstral Coefficient (MFCC) are important in finding out the state of a person. In this project, the speech signal is taken as the input and by means of MFCC feature extraction method, 39 coefficients are extracted by using MFCC. The large amount of extracted features may contain noise and other unwanted features. Hence, an evolutionary algorithm called as Ant Colony Optimization (ACO) is used as an efficient feature selection method. By using Ant Colony Optimization technique the unwanted features are removed and only best feature subset is obtained. It is found that the total number of features extracted get reduced considerably. The software used is MATLAB 13a.

Keywords: Ant Colony optimization, MFCC, feature selection, speech recognition

I. INTRODUCTION

Research in speech processing and communication for the most part, was motivated by people those desire to build mechanical models to emulate human verbal communication capabilities. Speech is the most natural form of human communication and speech processing has been one of the most exciting areas of the signal processing. Speech recognition technology has made it possible for computer to follow human voice commands and understand human languages. The main goal of speech recognition area is to develop techniques and systems for speech input to machine. Most of today's Automatic Speech Recognition (ASR) systems are based on some type of Mel-Frequency Cepstral Coefficients (MFCCs), which have proven to be effective and robust under various conditions. To enhance the accuracy and efficiency of the extraction processes, speech signals are normally pre-processed before features are extracted. Speech signal pre-processing covers digital filtering and speech signal detection.

The objective of this paper is to optimize the features extracted from the Mel Frequency Cepstral Coefficients (MFCC) using Ant Colony Optimization (ACO) algorithm. This can improve the performance of the Automatic Speech Recognition (ASR). Automatic speech recognition has made enormous strides with the improvement of digital signal processing hardware and software. Although significant advances have been made in speech recognition technology, it is still a difficult problem to design a speech recognition system for speaker independent, continuous speech. One of the fundamental questions is whether all of the information necessary to distinguish words is preserved during the feature extraction stage. If vital information is lost during this stage, the performance of the following classification stage in the ASR is inherently crippled and can never measure up to human capability. Thus, efficient techniques for feature extraction and feature selection have to be used in order to increase the speed of recognition. As a result, the performance of the Automatic Speech Recognition system can be improved. It is shown that as the number of iterations increased, the number of features get reduced. Section II explains about the overview of Automatic Speech Recognition

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

(An ISO 3297: 2007 Certified Organization)

Vol. 2, Issue 11, November 2013

(ASR) is presented. In section III, extraction of features using MFCC is presented. The feature selection algorithm called Ant Colony Optimization (ACO) is described in section IV. The results are discussed in section V. Conclusion and future work is presented in section VI.

II. OVERVIEW OF ASR

Speech Recognition (also known as Automatic Speech Recognition (ASR) or computer speech recognition) is the process of converting a speech signal to a sequence of words which is shown in figure 1 and it is implemented as algorithm in computer.

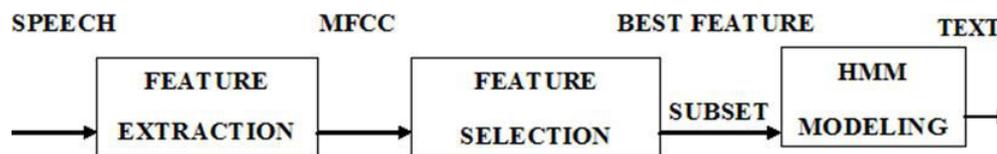


Fig.1 Block diagram of Basic components of ASR system

In the first step, the Feature Extraction, the sampled speech signal is parameterized. The goal is to extract a number of parameters ('features') from the signal that has a maximum of information relevant for the following classification. That means features are extracted that are robust to acoustic variation but sensitive to linguistic content. Put in other words, features that are discriminate and allow distinguishing between different linguistic units (e.g., phones) are required. On the other hand the features should also be robust against noise and factors that are irrelevant for the recognition process (e.g., the fundamental frequency of the speech signal).

In the modeling phase the feature vectors are matched with reference patterns, which are called acoustic models. The reference patterns are usually Hidden Markov Models (HMMs) trained for whole words or, more often, for phones as linguistic units. HMMs cope with temporal variation, which is important since the duration of individual phones may differ between the reference speech signal and the speech signal to be recognized. A linear normalization of the time axis is not sufficient here, since not all phones are expanded or compressed over time in the same way. In between the feature extraction and modeling phase, features selection algorithm is used. Algorithms like Evolutionary algorithms, Genetic algorithm and Neural Network based algorithms can be used for selecting best subset among the whole feature set.

III. FEATURE EXTRACTION BY MFCC

Feature extraction can be understood as a step to reduce the dimensionality of the input data, a reduction which inevitably leads to some information loss. Typically, in speech recognition, speech signals are divided into frames and extract features from each frame. During feature extraction, speech signals are changed into a sequence of feature vectors. Then these vectors are transferred to the classification stage.

MFCC is mostly used for Automatic Speech Recognition because of its efficient computation and robustness. Filtering includes pre-emphasis filter and filtering out any surrounding noise using several algorithms of digital filtering. Finally 36 coefficients are extracted from the Mel Frequency Cepstral Coefficient Method. The block diagram representing MFCC is shown in figure 2. MFCC consists of seven computational steps. Each step has its function and mathematical approaches as discussed briefly in the following:

A. Pre-emphasis

This step processes the passing of signal through a filter which emphasizes higher frequencies. This process will increase the energy of signal at higher frequency.

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

(An ISO 3297: 2007 Certified Organization)

Vol. 2, Issue 11, November 2013

$$Y[n] = X[n] - 0.95 X[n - 1] \quad (1)$$

Assume a = 0.95, which make 95% of any one sample is presumed to originate from previous sample.

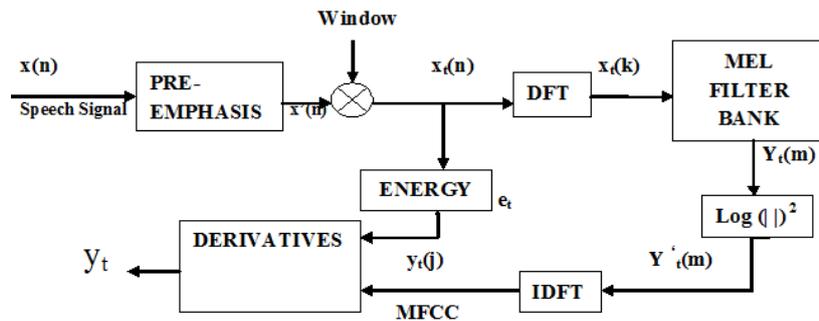


Fig.2 Block diagram representing Mel Frequency Cepstral Coefficient analysis

B. Framing

The process of segmenting the speech samples obtained from analog to digital conversion (ADC) into a small frame with the length within the range of 20 to 40 msec. The voice signal is divided into frames of N samples. Adjacent frames are being separated by M (M<N). Typical values used are M = 100 and N= 256.

C. Hamming windowing

Hamming window is used as window shape by considering the next block in feature extraction processing chain and integrates all the closest frequency lines.

y(n) = Output signal

x (n) = input signal

w(n) = Hamming window, then the result of windowing signal is shown below:

$$y(n)=x(n)* w(n) \quad (2)$$

D. Fast Fourier Transform

To convert each frame of N samples from time domain into frequency domain. The Fourier Transform is to convert the convolution of the glottal pulse u[n] and the vocal tract impulse response h[n] in the time domain. This statement supports the equation below:

$$Y(w)= FFT [h(n) *u(n)]=H(w)*U(w) \quad (3)$$

E. Mel-Scaled Filter Bank

The filter bank analysis consists of a set of band pass filter whose bandwidths and spacing's are roughly equal to those of critical bands and whose range of the centre frequencies covers the most important frequencies for speech perception. The filter bank is a set of overlapping triangular band pass filter, that according to mel-frequency scale, the centre frequencies of these filters are linear equally-spaced below 1 kHz and logarithmic equally-spaced above. The speech signal consists of tones with different frequencies. For each tone with an actual Frequency, f, measured in Hz, a subjective pitch is measured on the 'Mel' scale. We can use the following formula to compute the mels for a given frequency f in Hz:

$$mel(f) = 2595 * \log_{10} \left(1 + \frac{f}{700} \right) \quad (4)$$

F. Discrete Cosine Transform

This is the process to convert the log Mel spectrum into time domain using Discrete Cosine Transform (DCT). The result of the conversion is called Mel Frequency Cepstrum Coefficient. The set of coefficient is called acoustic vectors. Therefore, each input utterance is transformed into a sequence of acoustic vector. The IDFT needs complex arithmetic,

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

(An ISO 3297: 2007 Certified Organization)

Vol. 2, Issue 11, November 2013

the DCT does not. The DCT implements the same function as the FFT more efficiently by taking advantage of the redundancy in a real signal. The DCT is more efficient computationally.

IV. FEATURE SELECTION BY ACO

The main focus of this algorithm is to generate subsets of salient features of reduced size. ACO Feature Selection utilizes a hybrid search technique that combines the wrapper and filter approaches. In this regard, ACO Feature Selection modifies the standard pheromone update and heuristic information measurement rules based on the above two approaches. The reason for the novelty and distinctness of ACO Feature Selection algorithm versus previous algorithms like PSO, GA, lies in the following two aspects.

First, ACO Feature Selection emphasizes not only the selection of a number of salient features, but also the attainment of a reduced number of them. ACO Feature Selection selects salient features of a reduced number using a subset size determination scheme. Such a scheme works upon a bounded region and provides sizes of constructed subsets that are smaller in number. Thus, following this scheme, an ant attempts to traverse the node (or, feature) space to construct a path (or, subset). However, a problem is that, feature selection requires an appropriate stopping criterion to stop the subset construction. Otherwise, a number of irrelevant features may be included in the constructed subsets, and the solutions may not be effective. To solve this problem, some algorithms, define the size of a constructed subset by a fixed number of iteration for all ants, which is incremented at a fixed rate for following iterations. This technique could be inefficient if the fixed number becomes too large or too small. Therefore, deciding the subset size within a reduced area may be a good step for constructing the subset while the ants traverse through the feature space.

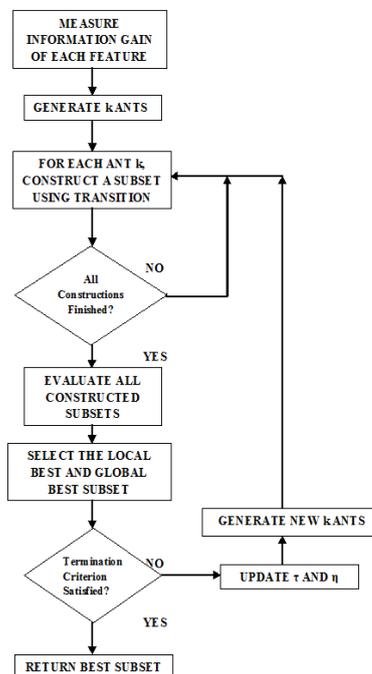


Fig.3 Flow Chart for ACO feature selection

The main structure of ACOFS is shown in figure 3. However, at the first stage, while each of the k ants attempt to construct subset, it decides the subset size r first according to the subset size determination scheme. This scheme guides the ants to construct subsets in a reduced form. Then, it follows the conventional probabilistic transition rule for selecting features as follows,

$$P_i^k(t) = \begin{cases} \frac{[\tau_i(t)]^\alpha \cdot [\eta_i]^\beta}{\sum_{u \in J^k} [\tau_u(t)]^\alpha \cdot [\eta_u]^\beta} & \text{(if } i \text{ belongs to } J^k \text{)} \\ 0 & \text{or else} \end{cases} \quad (5)$$

where,

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

(An ISO 3297: 2007 Certified Organization)

Vol. 2, Issue 11, November 2013

J^k = set of feasible features

η_i = pheromone value

τ_i = heuristic desirability associated with feature i

α and β = two parameters that determine the relative importance of the pheromone value and heuristic information.

The approach used by the ants in constructing individual subsets during Subset Construction (SC) can be seen in figure 4.

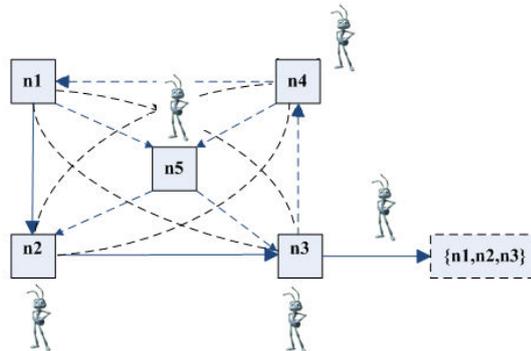


Fig.4 Representation of Subset Constructions

A quantity of pheromone, on each node is given as:

$$\Delta \tau_i^k(t) = \begin{cases} \phi \cdot \gamma(S^k(t)) + \frac{\phi \cdot (n - |S^k(t)|)}{n} & \text{if } i \text{ belongs to } S^k(t) \text{ or else } = 0 \end{cases} \quad (6)$$

where,

$S^k(t)$ = feature subset found by ant k at iteration t

$|S^k(t)|$ = feature Subset length.

The addition of new pheromone by ants and pheromone evaporation are implemented by the following rule applied to all the nodes:

$$\tau_i(t+1) = (1 - \rho) \tau_i(t) + \sum_{k=1}^m \Delta \tau_i^k(t) + \Delta \tau_i^g(t) \quad (7)$$

where,

m = number of ants at each iteration

$\rho(0,1)$ = pheromone trail decay coefficient.

V. RESULTS AND DISCUSSIONS

A. Implementation of Feature Extraction Algorithm

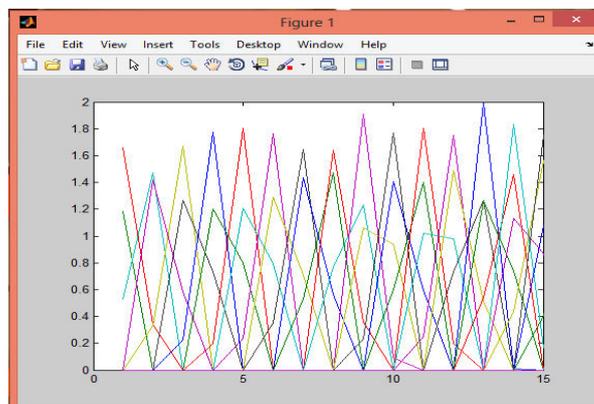


Fig. 5 Mel Filter Bank

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

(An ISO 3297: 2007 Certified Organization)

Vol. 2, Issue 11, November 2013

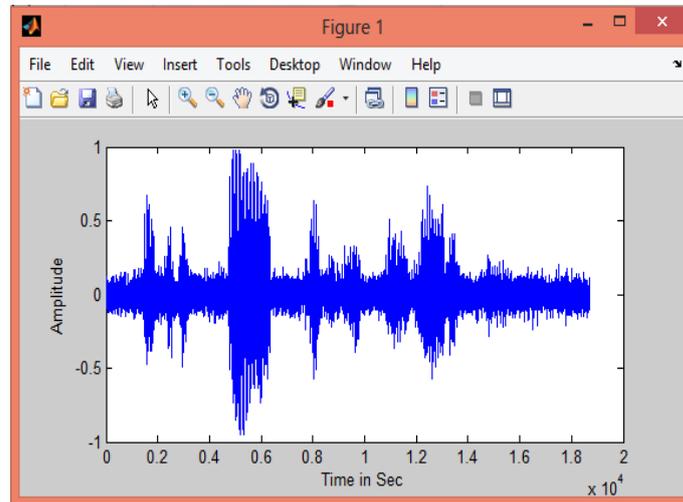


Fig. 6 Input Signal

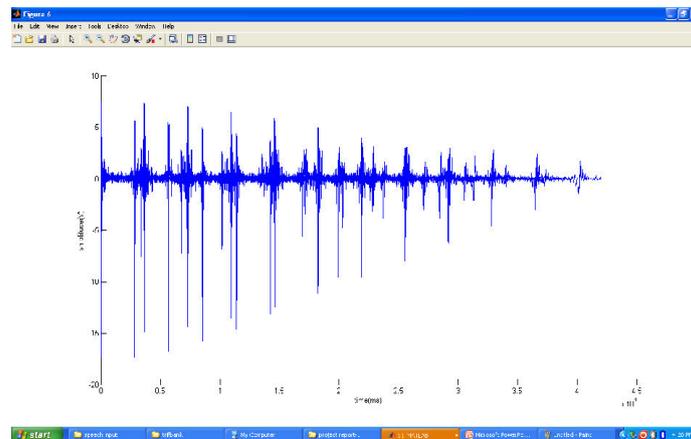


Fig. 7 MFCC Output

Figure 5 represents the group of filters used in the proposed work. Totally 24 filters are designed in which filters with cut off frequency upto 1 KHz are linear and above 1 KHz are logarithmic. Figure 6, shows the input speech signal for the feature extraction stage. Figure 7, shows the Mel Frequency Cepstral Coefficient (MFCC) output for the applied input speech signal. Initially Mel filter bank is implemented and then the MFCC output is obtained.

B. Implementation of Feature Selection Algorithm

In the implementation of ACO- Feature Selection algorithm, initially for 100 numbers of maximum iterations and for 6, 12, 13, 26 and 39 coefficients the best feature subset is obtained. Then the length of the best feature subset is calculated. The above procedure is performed for 200 and 300 iterations also. The length of the feature subset is also calculated for those MFCC coefficients separately for all 300 iterations. The total features taken are about 312.

The resulted values are tabulated and the ratio of length of feature subset obtained in 200 iterations and 300 iterations for 39 MFCC coefficients is calculated. Table 1 shows the length of the best feature subset for maximum number of iterations 100, 200 and 300 for the corresponding number of Mel Frequency Cepstral Coefficients.



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

(An ISO 3297: 2007 Certified Organization)

Vol. 2, Issue 11, November 2013

TABLE I
LENGTH OF THE BEST FEATURE SUBSET

No. of Iterations	No. of MFCC Coefficients	Length of the Best Feature Subset
100	6	72
	12	68
	13	58
	26	55
	39	48
200	6	68
	12	64
	13	60
	26	50
	39	45
300	6	64
	12	62
	13	55
	26	53
	39	40

From the table, we have observed that the number of features get reduced to about 16.6% in 300 iterations compared to 100 iterations. Compared to other optimization algorithms the ACO performs well.

VI. CONCLUSION & FUTURE WORK

In this project, the problem of optimizing the acoustic feature set by Ant Colony Optimization (ACO) technique for Automatic Speech Recognition (ASR) system is addressed. Some modifications of the algorithm are done and apply it to larger feature vectors containing Mel Frequency Cepstral Coefficients (MFCC) and their delta coefficient, and two energies. Ant Colony Optimization algorithm selects the most relevant features among all features in order to increase the performance of Automatic Speech Recognition system. From the tabulated results it is observed that the number of features get reduced when number of iterations increased and also number of MFCC coefficients increased. Compared to number of features obtained in 100 iterations, the features get reduced to 16.6% in 300 iterations. Ant Colony Optimization is able to select the more informative features without losing the performance.

Future work is to apply the best feature subset obtained from the proposed Ant Colony Optimization (ACO) algorithm to the modeling phase.

ACKNOWLEDGEMENT

Authors would like to thank Dr. S.Valarmathy and Ms. Kalamani for their support in implementation of this project.

REFERENCES

- [1]. A. Biem and S. Katagiri, "Cepstrum-based filter-bank design using discriminative feature extraction training at various levels," in Proc. IEEE Int. Conf. Acoustics, Speech, and Signal Processing, 1997, pp. 1503–1506.
- [2]. B. Milner and X. Shao, "Prediction of fundamental frequency and voicing from mel frequency cepstral coefficients for unconstrained speech reconstruction," in proc. of international conference IEEE Trans. Audio, Speech, Lang. Process., vol. 15, no. 1, pp. 24–33, Jan. 2007.
- [3]. Christian Blum, "Ant colony optimization: Introduction and recent trends", in Elsevier journal, Physics of Life Reviews 2 (2005) 353–373.
- [4]. Chulhee Lee, Donghoon Hyun, Euisun Choi, Jinwook Go, and Chungyong Lee, "Optimizing Feature Extraction for Speech Recognition", IEEE Trans. Audio, Speech, Lang. Process., Vol. 11, No. 1, pp.80, January 2003.
- [5]. D. R. Sanand and S. Umesh, "VTLN Using Analytically Determined Linear Transformation on Conventional MFCC", IEEE Transactions on Speech and Audio Processing, VOL. 20, NO. 5, pp.1573, JULY 2012.
- [6]. Daniele Giacobello, Mads Græsbøll Christensen, Manohar N. Murthi, Søren Holdt Jensen and Marc Moonen, "Sparse Linear Prediction and Its Applications to Speech Processing", IEEE Transactions on Speech and Audio Processing, Vol. 20, No. 5, pp.1644, July 2012.
- [7]. Dimitrios Dimitriadis, Petros Maragos and Alexandros Potamianos, "On the Effects of Filter bank Design and Energy Computation on Robust Speech Recognition", IEEE Transactions on Audio, Speech and Language Processing, Vol. 19, No. 6, August 2011.
- [8]. Dipmoy Gupta, Radha Mounima C. Navya Manjunath, Manoj PB, "Isolated Word Speech Recognition Using Vector Quantization (VQ)",



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

(An ISO 3297: 2007 Certified Organization)

Vol. 2, Issue 11, November 2013

in International Journal of Advanced Research in Computer Science and Software Engineering, Volume 2, Issue 5, May 2012 ISSN: 2277 128X, pp. 164-168.

- [9]. D. Giacobello, M. G. Christensen, M. N. Murthi, S. H. Jensen, and M. Moonen, “Enhancing sparsity in linear prediction of speech by iteratively reweighted 1-norm minimization,” in Proc. IEEE Int. Conf. Acoust., Speech, Signal Process., 2010, pp. 4650–4653.
- [10]. D. Chazan, R. Hoory, G. Cohen, and M. Zibulski, “Speech reconstruction from mel frequency cepstral coefficients and pitch frequency,” in Proc. ICASSP, 2000, vol. 3, pp. 1299–1302.

BIOGRAPHY



C.Poonkuzhali received her BE Degree in Electronics and Communications from Anna University, Chennai in May 2012. Currently she is pursuing her ME Degree in Applied Electronics from Anna University, Chennai. Her research interest includes Signal Processing, Wireless Sensor Networks and Evolutionary Computing. She has published 2 papers in International conferences and 1 paper in National Conference.



R.Karthiprakash received his BE Degree in Electronics and Communications from Anna University, Chennai in May 2012. Currently he is pursuing his ME Degree in Applied Electronics from Anna University, Chennai. His research interest includes Signal Processing, image processing and biometrics. She has published 1 paper in International conference and 1 paper in National Conference.



Dr.S.Valarmathy received her Ph.D. degree at Anna University, Chennai in the area of Biometrics in 2009. She is presently working as Professor & Head in the department of Electronics and Communication Engineering, Bannari Amman Institute of Technology, Sathyamangalam. She is having a total of 20 years of teaching experience in various engineering colleges. Her research interest includes Biometrics, Image Processing, Soft Computing, Pattern Recognition and Neural Networks. She is the life member in Indian Society for Technical Education and Member in Institution of Engineers. She has published 14 papers in International and National Journals, 48 papers in International conferences and National Conferences.



M.Kalamani received her M.E degree at Anna University, Chennai in Applied Electronics in 2009. She is currently pursuing her Ph.D. degree at Anna University, Coimbatore in ICE. She is presently working as Assistant Professor (Sr.G) in the department of Electronics and Communication Engineering, Bannari Amman Institute of Technology, Sathyamangalam. She is having a total of 7.5 years of teaching experience in Bannari Amman Institute of Technology. Her research interest includes Signal Processing, Speech Signal Processing and Neural Networks. She has published 10 papers in International and National Conferences.