QUERY TYPE IDENTIFICATION USING ISOLATED QUESTION WORD RECOGNITION

Raji Sukumar .A [1], Firoz Shah.A[2], Babu Anto.P [3]

School of Information Science and Technology, Kannur University

[1]*rajivinod.a@gmail.com*, [2]*firozathoppil@gmail.com*, [3]*bantop@gmail.com*

ABSTRACT

Speech is the predominant mode of human communication for every day interaction. Speech will also be the preferred mode for human-machine interaction. The study of human-machine communication comes in the area of intelligent query processing. This paper discusses a novel technique for recognition of the isolated question words from speech queries, which leads to the first step to identify the question type. Here we have created and analyzed a database of 250 question words from Malayalam (one of the south Indian language) speech queries. We have used Daubechies wavelets for the speech feature extraction purpose. We have used the Multi Layer Perceptron (MLP) Architecture and Piecewise Network Classifier (PLN) is used for classification and recognition. A recognition accuracy of 72% could be achieved from this experiment for Multi Layer Perceptron classifier and 60% for the Piecewise Network Classifier.

KEYWORDS

Discrete Wavelet Transform, Artificial Neural Network, Multi Layer Perceptron, Piecewise Network Classifier.

1. INTRODUCTION

The most natural way a human can communicate with a machine is dialogue-based and lead to intelligent query processing. Speech recognition systems falls into two classes isolated word recognition and continuous speech recognition. Isolated question word recognition from speech queries will be helpful to identify the type of the question. While answering a question, the system carries out a detailed query analysis that produces the identification of question type, i.e. time, location, person place, size number etc. The categorization involves morpho-syntactic analysis of the questions to identify the question type which defines what constitutes relevant data, which helps other modules to correctly locate the answer. Even though no universally admitted classification of possible question type exists. One of the various different taxonomies have been proposed by researchers are Factoid questions, whose answer is a name or a figure. To identify the question type, first step is recognizing the question word. The wide spread access to Internet makes the users assume that there is nothing easier than searching in the world wide web. Nevertheless a very large part of the world population does not have access to computers or internet. Hence speech query processing system has much importance in the area of Information Extraction. This system involves query in a natural language based on a grammar, which explain how words are combined in a meaning full way based on certain known. Spoken language understanding involves two primary component technologies, Speech Recognition (SR), and Natural Language (NL) understanding. The integration of speech and natural language has grater advantages and leads to a speech understanding system.

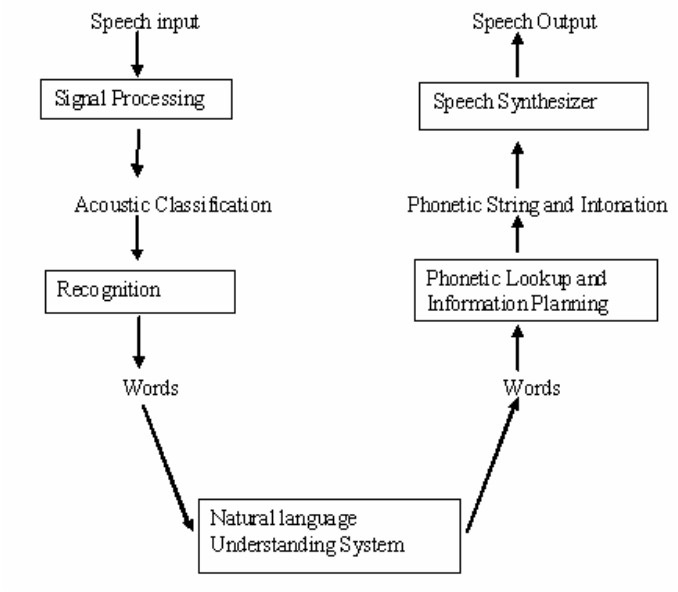

Figure 1:The architecture of speech understanding system

The basic architecture of spoken language understanding system is shown in the figures1 shows the entire natural language understanding. The sounds produced by the speaker are converted into digital form by an analog-to-digital converter. This signal is then processed to extract various features using the feature extraction method like Discrete Wavelet Transform. These features serve as the input to the speech recognition system which consider speech features as patters. Here we are using two classifiers, Multi Layer Perceptron and Piecewise Network Classifier to identify the most likely sequence of words to serve as input to the natural language understanding system.

2. FEATURE EXTRACTION METHOD

The most studied attributes for speech based studies are MFCCs, LPCs and prosodic parameters. Literature says that in case of the above said parameters the feature vector dimensions and computational complexity are higher to a greater extent. In this work we introduced Discrete Wavelet Transforms (DWT) for speech based query processing. Using this technique we can successfully reduce the computational complexity, because the size of the feature vector is very less. DWT is the transformation of a signal to the high frequency and low frequency components by using digital filtering techniques. DWT is the most promising mathematical transformation which provides both the time and frequency information of the input signals. DWT is performed by using digital filter banks to obtain the low frequency components of the input signals. Wavelet transform is a technique to transform an array of N numbers from their actual numerical values to an array of N wavelet coefficients. Each wavelet coefficient represents the closeness of the fit (or correlation) between the wavelet function at a particular size and a particular location within the data array. DWT is any wavelet transform for which the wavelets are discretely sampled. It captures both frequency and location information. DWT is performed by using digital filter banks to decompose the low frequency and high frequency components of the input signals. The DWT of a signal is calculated by passing it through digital filter banks. DWT uses filter banks to construct a multi resolution time-frequency plane. The number of filter banks increases the lower frequency component resolution. In DWT a discrete signal x[k] is filtered by using a quadrate mirror filter, which will separate the signals to detail coefficients and approximation coefficients. The filter outputs are then sub sampled by 2.

The definition of the scaling function $\phi_{j,k}(t)$ and wavelet function $\Psi_{j,k}(t)$ is given by

$$\phi_{j,k}(t) = 2^{j/2}\phi(2^j t - k) \qquad j,k \in Z \qquad (1)$$

$$\Psi_{j,k}(t) = 2^{j/2}\psi_{j,k}(2^j t - k) \quad j,k \in Z \qquad (2)$$

Where Ψ (t) is the basic analyzing function called the mother wavelet. As the number of filter banks increases the frequency resolution becomes arbitrarily good at low frequencies. The filter bank operation on the signal can be depicted by the following equations:

$$Y_{high}[k] = \sum nx[n]g[2k-n] \qquad (3)$$

$$Y_{low}[k] = \sum nx[n]h[2k-n] \qquad (4)$$

Where Y_{high} (detail coefficients) and Y_{low} (approximation coefficients) are the outputs of the high pass and low pass filters.

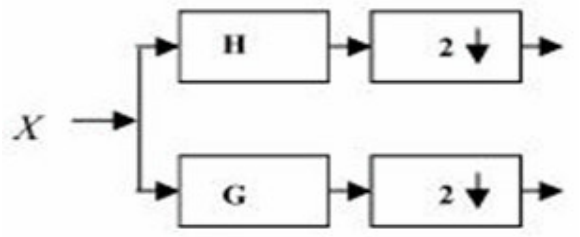

Figure 2 : Wavelet decomposition

Where Y high and Y low are the outputs of the high pass and low pass filters obtained by sub sampling by 2 ,as shown by the figure 2.

3. QUESTION WORD CORPUS FOR MALAYALAM

For this experiment we have created and analyzed a speaker dependent database for our experiment. The queries used for the database creation include the factoid question words which will helps for the classification of questions a database consisting of 250 single session speaker dependent speech query utterance by using Malayalam. The question words used to create the database and their corresponding IPA formats are given in Table1. [1].

Table 1 : Question word Speech database and IPA format

Words in English	Words in Malayalam	IPA format
Aaranu	ആരാണ്	//ɑ□/ ɔr/ɑ□ /n /ʊ//
enthanu	എന്താണ്	//□ /n /t /h/ɑ□/n /ʊ//
engineaanu	എങ്ങിനെയാണ്	//□ /n/ ŋ/ɣ /ɪ/ n /æ/ ɔɪ/ ɑ□/n /ʊ//
ethrayanu	എത്രയാണ്	//□/t/h/r/□ □/ɔ□/ ɑ□/n /ʊ//
eppozhanu	എപ്പോഴാണ്	//□ /p /p/ □/ z /h/ ɑ□/ /n /ʊ//

4. PATTERN RECOGNITION BY USING ARTIFICIAL NEURAL NETWORK

Artificial Neural Network is a powerful pattern recognition technique, which closely relates to the human neural method of understanding. A neural net has a parallel-distributed architecture with larger number of neurons. Collective behaviour of neurons is the property behind the intelligence of the neural networks. Since the performance of each neuron is very limited, the parallel processing of the neurons helps them to overcome the individual slow performance of each of them. Artificial neural network is a computational model that tries to simulate the structure or functional aspects of biological neural networks. Just like human neuron, ANN understands the things. It consists of an interconnected group of artificial neurons and processes information using a connectionist approach to computation. In most cases an ANN is an adaptive system that changes its structure based on external or internal information that flows through the network during the learning phase. Neural networks are non-linear statistical data modelling tools. They can be used to model complex relationships between inputs and outputs or to find patterns in data. The architectural structure of neuron is a network model of parallel distributed architecture. A neural network is intelligent because the collective behaviour of neurons and which allows to overcome the slow performance of individual neurons.

4.1 Pattern recognition by using Multi Layer Perceptron

In the first experiment used artificial neural network for pattern classification and recognition. We have used Multi Layer Perceptron (MLP) architecture to train and test the network. A multilayer perceptron is a feed forward neural network with at least one hidden layer. The MLP consisting of nodes arranged in layers with only forward connections to units in subsequent layers. Each signal travelling along a link is multiplied by its weight. It can deal with nonlinear classification problems because it can form more complex decision regions. The architecture of the MLP is as shown in Figure 3.

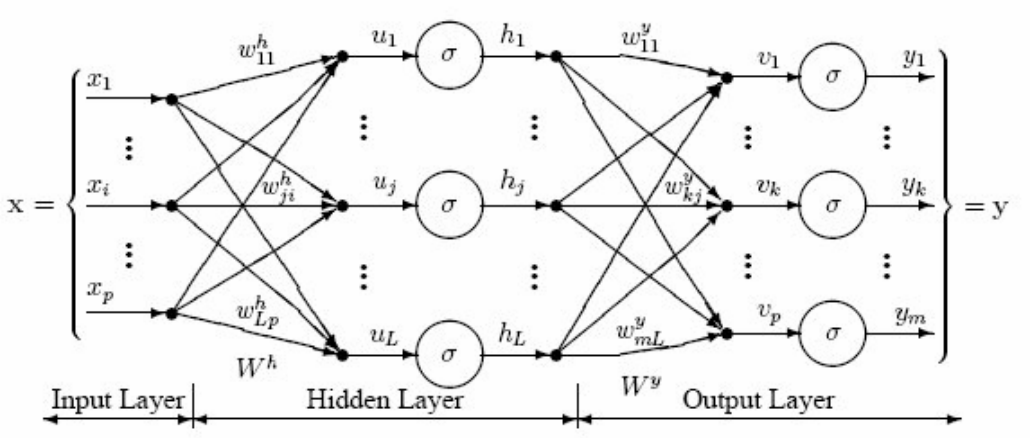

Figure 3 : Multi Layer Perceptron network

The diagram illustrates a perceptron network with three layers, network has an input layer (on the left) with three neurons, one hidden layer (in the middle) with three neurons and an output layer (on the right) with three neurons. There is one neuron in the input layer for each predictor variable. In the case of categorical variables, N-1 neurons are used to represent the N categories of the variable.

Input Layer: A vector of predictor variable values (x1...xp) is presented to the input layer. The input layer (or processing before the input layer) standardizes these values so that the range of each variable is -1 to 1. The input layer distributes the values to each of the neurons in the hidden layer. In addition to the predictor variables, there is a constant input of 1.0, called the

bias that is fed to each of the hidden layers; the bias is multiplied by a weight and added to the sum going into the neuron.

Hidden Layer: — Arriving at a neuron in the hidden layer, the value from each input neuron is multiplied by a weight (w_{ij}), and the resulting weighted values are added together producing a combined value (u_j). The weighted sum (u_j) is fed into a transfer function, σ, which outputs a value (h_j). The outputs from the hidden layer are distributed to the output layer.

Output Layer: Arriving at a neuron in the output layer, the value from each hidden layer neuron is multiplied by a weight (w_{kj}), and the resulting weighted values are added together producing a combined value (v_j). The weighted sum (v_j) is fed into a transfer function, σ, which outputs a value (y_k). The y values are the outputs of the network.

If a regression analysis is being performed with a continuous target variable, then there is a single neuron in the output layer, and it generates a single y value. For classification problems with categorical target variables, there are N neurons in the output layer producing N values, one for each of the N categories of the target variable.

4.2 Multilayer Perceptron Architecture

The network diagram shown above is a full-connected, three layer, feed-forward, perceptron neural network. "Fully connected" (means that the output from each input and hidden neuron is distributed to all of the neurons in the following layer). "Feed forward" (means that the values only move from input to hidden to output layers; no values are fed back to earlier layers).

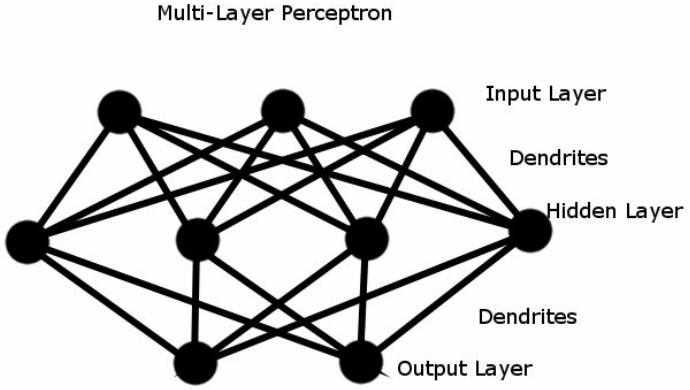

Figure 4: The Architecture of Multi Layer Perceptron

Output vectors and the intermediate layers or hidden layers A Neural Network (NN) is usually structured into an input layer of neurons, one or more hidden layers and one output layer. Neurons belonging to adjacent layers are usually fully connected and the various types and architectures are identified both by the different topologies adopted for the connections and by the choice of the activation function. Such networks are generally called Multi Layer Perceptron (MLP). The MLP network, one of most popular NN, is suited to a wide range of application, such as pattern recognition, prediction, process modelling, etc. An MLP network comprises a

number of identical units organized in layers, with those on one layer connected to those on the next layer.

4.3 Pattern classification using Piecewise Linear Network

A Piecewise Linear Network (PLN) is discussed which classifies N-dimensional input vectors. The network uses a distance measure to assign incoming input vectors to an appropriate cluster. Each cluster has a linear classifier for generating class discriminants. A training algorithm is described for generating the clusters and discriminants. Classification deals with associating a class label to a point, according to the class to which the point is believed to belong. Classification is typically based on a distance measure between the point under consideration and other points belonging to known classes in point space

It is not hard to think of configurations of points for which simple linear classification will perform poorly, however. There may exist no hyper plane capable of uniquely separating point classes, because of geometrical reasons or simply because there are more than 2 classes. Figure 5 shows an example of this problem. The PLNC has input elements in the first layer, the hidden units in the second and the outputs in the third layer.

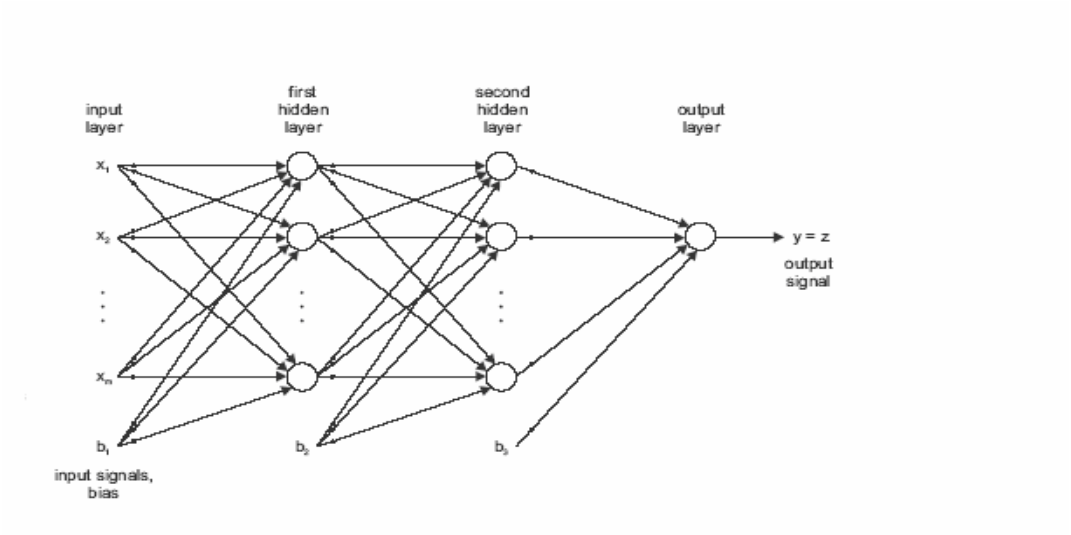

Figure 5: Piecewise Linear Network

As shown in figure 5, the PLNC has input elements in the first layer, the hidden units in the second and outputs in the third. The N-dimensional input vector X_f has elements $X_f(n)$ with means and standard deviations $\mu(n)$ and $\sigma(n)$ respectively where $1 \leq n \leq N$. First, x_f is normalized as

$$X_f(n) \leftarrow \left[X_f(n) - \mu(n)\right] / \sigma(n) \qquad (5)$$

The normalized $(N+1)$-dimensional vector x is formed by augmenting x_f as

$$X = \left(X_f^T : 1\right)^T \qquad (6)$$

The hidden layer consists of K clusters, each having an N- dimensional mean vector $\mathbf{m}_c$, where $1 \leq c \leq K$. Each cluster also has a weight matrix $\mathbf{A}_c$ of dimension N_c by $(N+1)$, where N_c is the number of classes in the classification problem.

Given an input vector $\mathbf{x}$, we find cluster index c such that $d(\mathbf{x},\mathbf{m}_c)$ is minimized. Then we form the output vector $\mathbf{y}$ as

$$\mathbf{y} = A_c \,.\, \mathbf{x} \qquad (7)$$

The estimate of the correct class i_c is given by

$$i_c^i = \arg\max[y_i] \qquad (8)$$

where y_i is the i^{th} element of output vector **y** and $1 \le i \le Nc$.

5. RESULT AND DISCUSSIONS

In this experiment, isolated question word recognition from speech queries is done by using a hybrid approach of discrete wavelet transform technique and artificial neural networks. A database is created by using five isolated interrogative words in Malayalam language. We have created a speaker dependent speech corpus for the database. The words recorded by using a high quality microphone at a sampling rate of 8 KHz band limited. The sentences containing the questions words are recorded first and then the question words are isolated and pre-processed to use in the database. The database consists of a total number of 250 question word speech samples. Discrete Wavelet Transformation (DWT) was performed on the samples in the database by using the Daubechies wavelets. We have performed successive decomposition on the signals by using Db4 wavelets. In all the cases we have taken the approximation coefficients from the thirteenth level of decomposition. According to the order of wavelets used the feature extraction; the obtained feature vectors also have different sizes.

In the first experiment we have used Artificial Neural Network for machine learning. We used the Multi Layer Perceptron architecture for training and testing of the neural network and a database proportion of 80:20 for training and testing of the neural network respectively. The same proportion is used for training and testing of the entire feature vector developed by different order Daubechies wavelets. After successful training of the neural network the classifier can recognize the four different question words. When the testing of the classifier was performed an overall recognition accuracy of 80% was achieved. A confusion matrix is also added to indicate the performance of the recognition accuracies.

Table 2 : Confusion matrix obtained for MLP classifier

Question Class	Q1	Q2	Q3	Q4	Q5
Q1	90%	0%	0%	20%	0%
Q2	0%	100%	0%	0%	0%
Q3	0%	0%	100%	0%	0%
Q4	0%	0%	100%	0%	0%
Q5	20%	0%	0%	60%	40%

Confusion matrix obtained for Db8 wavelet

In the second experiment Artificial Neural Network for machine learning is done by using Piecewise Linear Classifier. We use the same for training and testing of the neural network. Here we used and a database proportion of 80:20. for training and testing of the neural network respectively. The same proportion is used for training and testing of the entire feature vector developed by different order Daubechies wavelets. After the training the neural network classifier can recognize the five different question words. When the testing of the classifier was performed an overall recognition accuracy of 65% was achieved. A confusion matrix is also added to indicate the performance of the recognition accuracies.

Table 3:Confution matrix obtained for different question words using PLN classifier

Question Class	Q1	Q2	Q3	Q4	Q5
Q1	100%	0%	0%	0%	0%
Q2	0%	90%	10%	0%	0%
Q3	0%	40%	60%	0%	0%
Q4	20%	0%	0%	40%	40%
Q5	40%	20%	0%	0%	40%

6. CONCLUSIONS

In this experiment isolated question word recognition from speech queries is done by using discrete wavelet transformation and artificial neutral networks. In experiment one we could achieve an overall recognition accuracy of 72% for five question words. And in experiment two we could achieve an accuracy of 62%. From both experiments it can be stated that Multi Layer Perceprtron can be selected which provide a better performance using discrete wavelet transform.

ACKNOWLEDGEMENTS

The authors would like to thank everyone, just everyone!

REFERENCES

[1]. Atsushi Fujii, Katunobu Itou, and Tetsuya Ishikawa, " A method for open-vocabulary speech driven text retrieval," in *proceedings of the 2002 conference on Emperical methods in Natural Language Processing*, pp. 188-195, 2002.

[2]. James Allen, "Natural language Understanding" second edition ISBN:0-8053-0334-0.

[3]. Satoshi Kaki, Eiichiro Sumitha, and Hitoshi Iida, "A Method for correcting speech recognition using the Statistical features Character Co-occurrence," *COLING-ACL '98*, pp. 653-657, 1998.

[4]. L. R. Rabiner and R. W. Schafer, "Digital Processing of speech signals," Prentice-Hall, 1978.

[5]. K.P. Soman K..I. Ramachandran "Insight into wavelets from theory to practice" second edition Prentice –hall of india ISBN 8L-203-2902-3

[6]. Gembac Lee, Jungyun Seo, Seungwoo Lee, Ilanmin Jung, Bong-IIyun Cho,Changki Lee, Byung-Kwan Kwak,Jeongwon Cha, Dongseok Kim, Joollui An, Ilarksoo Kim, and Kyungsun Kim, "SiteQ:Engineering Iligh performance QA System, using Lexico-Semantic Pattern Matching and Shallow NLP", *In proceedings of the 10th Text Retrival Conference(TREC-10),* Washington D. C.,2001.

[7]. "Neural Networks, Fuzzy Logic, and Genetic Algorithms, Synthesis and Applications", S. Rajasekeran, G. A. Vijayalakshmi Pai, Prentice-Hall of India, ISBN 8L-203-2L86-3, 2004.

[8]. "Neural Networks In Computer Intelligence", LiMin Fu, Tata McGraw-Hill publishing company limited ISBN 0-07-053282-6(2003)pp175-185.

[9]. M.S. Chen and M.T. Manry, "Conventional Modeling of the Multilayer Perceptron using Polynomial basis Functions", IEEE Trans. on Neural Networks, vol.4, No.1, Jan 1993, pp 164-166.

[10]. Fabio Crestani , Heather Du, Written versus spoken queries: A qualitative and quantitative comparative analysis, Journal of the American Society for Information Science and Technology, v.57 n.7, p.881-890, May 2006

[11]. S. Subbarayan, K. Kim, M. T. Manry, V. Devarajan and H. Chen,"Modular neural network architecture using piecewise linear mapping," 30th Asilomar Conference on Signals, Systems & Computers, vol. 2, pp. 1171-1175, Nov. 1996.

[12]. M.S. Chen and M.T. Manry, "Backpropagation representation theorem using powerseries", Proc. IJCNN, vol.I, 1990, pp 643-648.

Authors

Raji Sukumar. A, is a Research scholar working In the area of Natural Language Processing and Speech Processing at School Of Information Science And Technology Kannur University, Kannur District, Kerala State, India. He has received Msc. Degree in Computer Science from Kannur University, Kerala. His main research interest includes Query Processing, Artificial Intelligence and Signal Processing.

Firoz Shah A, is a Research scholar working In the area of emotional speech processing at School Of Information Science And Technology Kannur University, Kannur District, Kerala State, India. He has received Msc. Degree in Electronics from Mahatma Gandhi University, Kerala. His main research interest includes speech processing, Artificial Intelligence and Signal Processing.

Dr. Babu Anto P, is a Reader in information Science and Technology at the Department of Information Science And Technology, School Of Information Science And Technology, Kannur University, India. He holds his Msc. And Ph.D. Degrees in Electronics from CUSAT (Kochin University of Science And Technology), Kerala, India. His current research interests include Speech and Emotion recognition, Data Mining, Speaker Recognition, Visual Cryptography. He has published research papers in reputed national and international journals.