ISSN (online): 2349-784X

# Speech Based Access for Agricultural Commodity Prices in Tamil

**S. Yogapriya**
*PG Student*
*Department of Electronics & Communication Engineering*
*Saranathan College of Engineering, Tiruchirappalli*
*Tamil Nadu, India*

**Dr. P. Shanmugapriya**
*Associate Professor*
*Department of Electronics & Communication Engineering*
*Saranathan College of Engineering, Tiruchirappalli*
*Tamil Nadu, India*

## Abstract

Machine-Human interaction plays an important role for many users in order to access the information. The interactions could be in the combination of text, speech, facial expressions etc., or separately. The objective of the proposed work is to develop a speech based agricultural commodity prices accessing system in which recognition of user queries is implemented using i-vector approach. In the proposed work, i-vector based approach is adopted for speech recognition. The recognition tasks can be conducted for both speaker dependent and speaker independent case. The experimental results show that the Equal Error rate (EER) obtained for the proposed system is improved and the performance of the system is improved when increasing the number of Gaussian mixture components and also number of i-vector components.

**Keywords: GMM-UBM, i-vector, MFCC, Speech Recognition**

## I. INTRODUCTION

Speech recognition is the process in which the system can understand and respond to the sound produced by the human speech. Studies of speech recognition have been performed for decades. Speech is the low frequency, aperiodic one dimensional signal which composed of sequence of sounds which have the transformation that reveal the information. It consists of 3 important parts: Pre-processing, Feature extraction, and pattern recognition. Pre-processing is the process of conversion of speech input into the form in which the speech recognizer can process. Feature extraction is the process to extract the features from the input speech signal that help the system in identifying speaker. Speech recognition is the process in which identification of what has been said at the input side. Various methods have been proposed for efficient extraction of speech parameter for recognition of speech. Among them, MFCC (Mel-Frequency Cepstral Coefficients) [1] [2] plays an important role. Along with the MFCC, the recognition such as HMM is mostly used. In the proposed work GMM-UBM [3] is used as a model for the speech data and the performance of the speech recognition system is analyzed. Similar to speech recognition, speaker identification, speaker verification also plays an important role. In the field of biometric identification, various studies have been performed. Mostly speaker identification systems use the model called Gaussian Mixture Model (GMM) [5]. Many approaches have been proposed in order to improve the discriminative qualities of GMM. Among them, Gaussian mixture model-universal background model (GMM-UBM) is a popular method. In UBM, the speaker models are adapted by using Bayesian adaptation [6].This model is the global model. A UBM is formed from the speech data set containing all the words. During training, by using Maximum a posterior adaptation (MAP), GMM model is adapted from the UBM [6]. The state-of-the-art method for speaker recognition systems uses i-vector along with GMM-UBM model [7]. i-vector is the low dimensional fixed length vector which can be used as the compact representations of speaker utterances which is the modified version of Joint factor analysis (JFA) [8].

## II. RELATED WORKS

This section provides various work done in the field of speech recognition and speaker recognition. Speech recognition is the process in which the speech sound produced by the human can be identified and responds by the system. Speaker recognition involves two tasks namely speaker verification and speaker identification [9]. Identification of sound produced by the human from the known speakers is called speaker identification. Speaker verification is the process of reject or accepts the identity claim of the speaker who uttered the sound. For speech recognition, along with the MFCC, HMM is predominantly used. In the proposed work, GMM-UBM is used as a model of speech and the performance of the system is analyzed. The research is going on in speech recognition for many years in order to accomplish the agricultural task domain. In [10], Speech Based Interaction system (SBC) system was proposed in Telugu language by IIT Hyderabad which is implemented using sphinx recognition system. They have used the method called context dependent tri-phone HMM with the eight Gaussian mixtures per state to model a speech data. 96 speakers are used for recording the data where each speaker uttered 500 words (lists of commodity, markets and district). Total number of words recorded is 9600. The accuracy as 77.7% is obtained. In [11], the authors used the models called CDHMM and SGMM with the sixteen Gaussian mixtures per state in order to investigate the problem of ASR acoustic modelling for agricultural tasks domain. They have developed ASR in four Indian languages namely Assamese, Bengali, Marathi and Hindi and compares

the performance of the ASR modeled using SGMM and CDHMM. They concluded that for the languages assamese and Bengali, the performance of ASR system using both CDHMM based acoustic modeling and SGMM methods is comparable due to small vocabulary size. Similarly for Hindi and Marathi due to large vocabulary size, SGMM provides the better performance. They have implemented using HTK toolkit.

In [12], six consortium members are involved in developing a speech based automated commodity price helpline in six Indian languages namely Assamese, Bengali, Hindi, Marathi, Telugu and Tamil. They have used CDHMM for acoustic modelling configuration for 10 commodity and 10 district names in each languages. Still the research is going on. In [13], the author developed the telephony based automatic speech recognition system for Bodo language implemented using asterisk server and sphinx recognition toolkit. Total number of words collected are 31(4 digits (0, 1, 2, 3)), 25 commodity names and yes/no word from 100 different speakers (70 male speakers and 30 female speakers). ASR can be modelled using tied state tri-phone model with sixteen Gaussian mixtures per state. They have obtained the accuracy of 77.24% in training phase and 72.12% in testing phase. Similarly various research works are being done in speaker recognition.

In [14], speaker verification task is implemented using i-vector based approach and GMM based approach in order to reduce the confusion errors. For this, the speech data collected from 50 speakers in a laboratory environment. The performance can be obtained in the form of Equal Error Rate (EER). EER can be decreased by 4 and 4.5%. In [14], the authors, proposes a speaker identification system which provides 2 important parts namely feature extraction and sparse representation classifier (SRC). For classification purpose sparse representation classifier is used. The databases like NIST2005, NIST2006, and NIST2008 are used. Total number of speakers is 136 and the Gaussian mixture component of 512. The length of the i-vector is 400. Here the author tries to enhance the performance of SRC which compensates variations of session and channel in order to make the dictionary discriminative. In [15], the author proposed i-vector based method for signature verification using the database of SigWiComp2013. i-vector is extracted and used for template making. In order to reduce the within class variance WCCN is used and cosine similarity used for scoring performance. The author analyzes the performance by varying the number of mixture component and dimension. As a result the better performance obtained for larger mixture component with i-vector of larger dimension.

In the proposed work, by considering small account of work in agricultural domain, the speech recognition is performed using i-vector based approach and the performance of the system is analyzed. EER obtained for proposed method is compared with state of art method GMM-UBM for various i-vector dimensions.

## III. METHODOLOGY

### *A. Speech Recognition through i-vector Approach*

In the proposed work, i-vector approach can be adopted for speech recognition. The proposed recognition system consists of three parts namely, feature extraction, modelling using GMM-UBM and i-vector extraction. Fig.1 shows the block diagram of proposed speech recognition system.

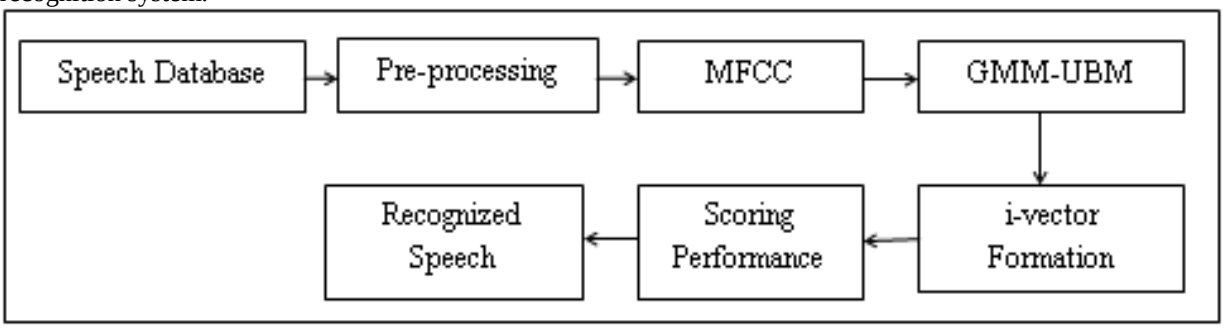


Fig. 1: Block Diagram of Proposed Recognition System

The description of the block of i-vector based approach is given as follows:

The isolated Tamil words which are recorded through normal microphone and saved as a wave (.wav) file using audio editing software/tool named Audacity. These recorded wave files are stored in a particular directory which is collectively called speech database. The speech signal from the database is pre-processed and the MFCC features from the speech signal are extracted. After that, i-vector is extracted for extracting the features which compensates the variations due to channel, speaker or source. After that, score of i-vector based models are obtained log-likelihood ratio between corresponding model and test data. Performance of the recognition system is analyzed in the form of detection error trade off (DET) curve. DET curve which shows the relation between false positive rate and false negative rate.

### *B. Feature Extraction Process*

Feature extraction is the important part in speech recognition system. It is the representation of original input speech signal by calculating the series of feature vectors. It helps to identify the corresponding word uttered by extracting the features from the input speech signal. There are various feature extraction technique are available [16], namely Linear Predictive Cepstral Coefficients

(LPCC), Perceptual Linear Prediction Coefficients (PLPC) and Mel-Frequency Cepstral Coefficients (MFCC). Among these features, MFCC is used in the proposed work.

### C. *Mel-Frequency Cepstral Coefficients*

From the input speech signal, MFCC (Mel-frequency cepstral coefficients) technique is used to extract the useful information. The MFCC consists of important feature called Mel filter bank. The important characteristics of Mel filter bank is, at low frequency it can be spaced linearly and at high frequency it can be spaced logarithmically in order to capture the useful characteristics in the input speech signal. Fig.2 shows the process of MFCC Extraction.

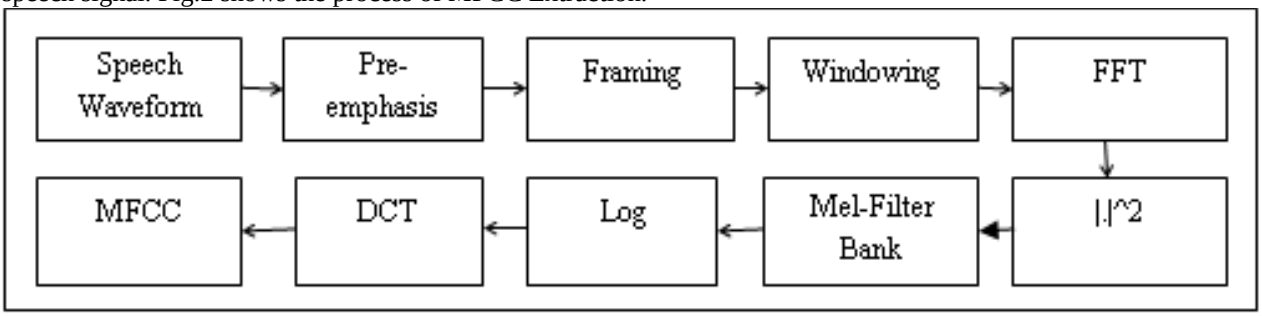


Fig. 2: MFCC Extraction Process

The first step in feature extraction is pre-processing. The steps involved in pre-processing are: pre-emphasis, Framing and windowing. Frequency normalization is performed through pre-emphasis. Then frame blocking is performed on pre-emphasized signal. The frames are windowed in order to maintain the signal continuity. Usually hamming window is used for this purpose. In order to obtain the power spectrum of the signal, Fast Fourier Transform (FFT) is applied to the each frame. After that the spectrum of the signal is passed to the Mel-filter bank. The Mel frequency can be calculated using the following formula (1),

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

Similar to the frequency scale of human ear Mel filter is an auditory scale. After that, logarithmic operation is performed. Finally, discrete cosine transform is used to calculate the cepstrum coefficients. Table 1 shows the parameters used for feature extraction.

Table – 1
Parameters of MFCC Extraction

| *Parameter* | *Value* |
|---|---|
| *Sampling Rate* | *8000 Hz* |
| *Frame Rate* | *100 f/s* |
| *Window Size* | *256* |
| *Dimensions of the Features* | *12* |
| *Total Number of Filters in the Mel Filter Bank* | *40* |

### D. *i-vector Formation*

i-vector was proposed initially For automatic speaker recognition [18], and after that it was adopted for other applications such as emotion recognition [21], language identification [18][19] etc. In order to overcome the session and channel variabilities in speaker recognition joint factor analysis (JFA) [8] is used. When compared to JFA, i-vector performs better [6][14][18]. i-vector makes the speech signal as fixed length. In speaker recognition, i-vector is extracted from the input speech signal. By using this extracted vector, scores can be computed and then recognized. Same approach is applied to the proposed recognition of speech. There are two reasons for adopting this i-vector for speech recognition. The first reason is each input word uttered has variable lengths. Hence by using i-vector, variable length of the input signal can be converted into fixed length features. The second reason is the speech signal of a person is slightly different in each time. These variations lead to increase in false rejection rate (FRR). These variations can be reduced by LDA. Fig. 3 shows the block diagram of steps involved in the recognition of i-vector approach. After extracting feature from the speech signal using MFCC, a model called universal background model (UBM) [22][23][24]is developed. There are different models have been used as UBM. Normally for text independent speaker verification task GMM is used [17] [21] and for text dependent speaker verification task HMM is used [22] [23]. In this proposed speech recognition system, in order to analyze the performance of the recognition system, GMM is used.

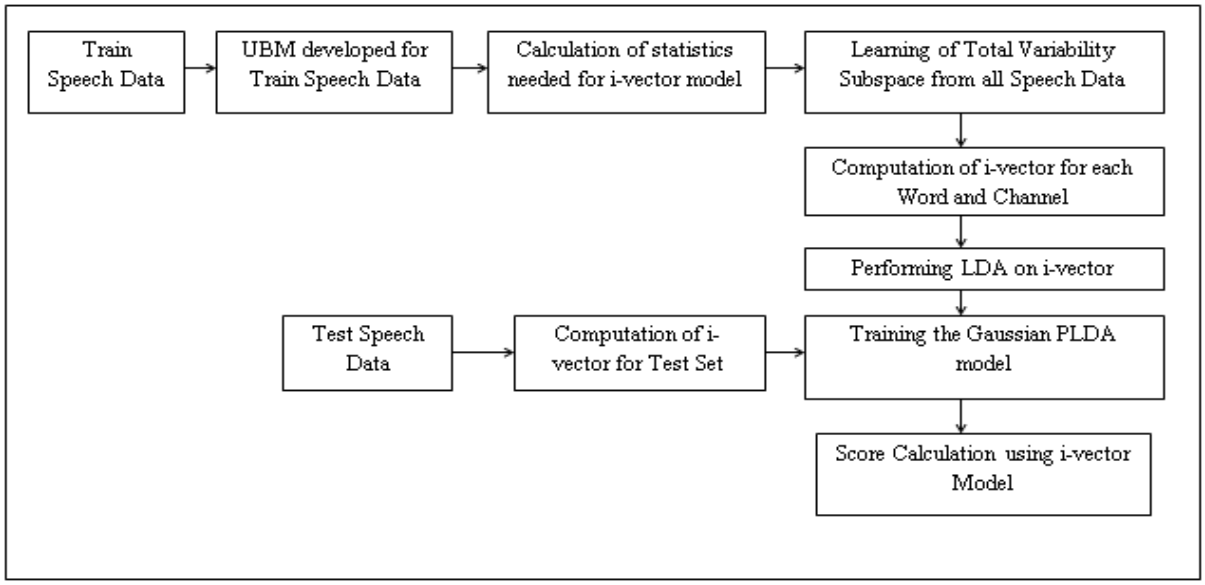


Fig. 3: i-vector Based Recognition System

The process of extraction of i-vector for speech recognition is given as follows:

- Initially GMM is created for the extracted speech vectors by using Expected Maximization algorithm.
- By using maximum a posteriori adaptation (MAP), adaptation of specific speech GMM model from the UBM model. [Here UBM is nothing but mean, covariance matrix and weights of the entire feature vector].
- Then, by using Baum Welch Method statistics can be created for entire feature vectors.
- Total variability subspace is obtained from the extracted vectors through EM algorithm.
- i-vector are extracted from the extracted feature vectors of training set.
- In order to minimize the within class variances and maximizes the between class variances, linear transformation i.e. , linear discriminant analysis (LDA) is performed.
- Again, using EM algorithm Gaussian PLDA model is developed from the extracted i-vector set (i.e., trained i-vector).
- Similarly for testing set, MFCC features are extracted and this extracted feature set are used for the computation of statistics.
- By using the statistics, i-vector for the test speech set is extracted.
- Scoring Performance is obtained by taking the log-likelihood between the model and test data which is given in the form of detection error trade off (DET) curve.

Based on Baum Welch method [21], statistics can be created using UBM for each features of each signal.

Consider $X_i$ is the entire extracted feature vectors for the training input signal $i$, the $0^{th}$ (N) and $1^{th}$ (F) order statistics for $c^{th}$ component of Universal Background Model [16] which is calculated as in equations (2) and (3),

$$N_c(X_i) = \sum_t \gamma_{i,t}^c \tag{2}$$

$$F_c(X_i) = \sum_t \gamma_{i,t}^c (X_{i,t} - m_c) \tag{3}$$

Where, $X_{i,t}$ - $t^{th}$ Vector of entire feature of speech signal,

$m_c$ -mean of $c^{th}$ component,

$\gamma_{i,t}^c$ -posterior probability of UBM mixture component given extracted feature.

$$\gamma_{i,t}^c = p\left\langle c \middle| X_{i,t} \right\rangle = \frac{w_c \mathrm{N}\left\langle X_{i,t} \middle| m_c, \sum_c \right\rangle}{\sum_{j=1}^{C} w_j \mathrm{N}\left\langle X_{i,t} \middle| m_j, \sum_j \right\rangle} \tag{4}$$

Where, N denotes the normal distribution.

Total variability subspace can be created using EM algorithm [24]. For this, assume factor analysis model of the form,

$$M = m + T.X \tag{5}$$

Where, $M$ -adapted mean super vector,

$m$ -UBM mean super vector,

$T$ -total variability matrix with low rank,

$X$ -i-vector.

i-vector compensate the variations due to session and speaker , by creating the total variability subspace 'T'. To reduce the session variability which is unwanted in i-vector space, inter class compensation methods like LDA followed by PLDA is used. From [26] the separation of speaker class defined by the linear discriminant analysis in the direction of B which is given by the following equation (6),

$$\lambda = \frac{B^T S_b B}{B^T S_w B} \tag{6}$$

Where, B is the projection matrix which contains $k$ Eigon vectors.

$S_b$ and $S_w$ are the within and between class covariance matrix.

## IV. EXPERIMENTAL SETUP AND PERFORMANCE ANALYSIS

In the proposed work, experiment is conducted for both speaker dependent and speaker independent case. For speaker dependent case, the words collected from a single speaker is used. Total number of utterance collected from the single speaker is 4725. Among them 4200 utterances can be used for training and 525 utterance used for testing. For speaker independent case, the speech is collected from 3 different speakers. Total number of utterance collected from each speaker is 675. Among them, 450 utterances are used for training and 225 utterances are used for testing. For a single speaker, the number of words used for recording purpose is 35 which include agricultural commodity names. For speaker independent, number of uttered words are 17. The sampling rate of each speech signal is 8 kHz which can be digitized at 16 bits per sample. Each utterances remains the duration of 1 second. Each input speech signal is converted into the feature vectors called Mel-frequency cepstral coefficients which have the dimensions of 12 with the frames of 96. After that, universal background model can be created by using speech data which produce the weight, mean and covariance of the training speech data can be obtained. Then first and the second order statistics can be obtained. The length or i-vector dimension must be varied between 50 and 200 and the mixture component must be varies between 32 and 1024, the performance can be analyzed. The LDA and PLDA dimension must be 16 and 34 for speaker independent and speaker dependent case. After that features extracted from the speech data for testing, these features can be compared to the PLDA Gaussian model which is developed for train speech data. Finally scoring verifications can be obtained in the form of equal error rate. The performance of the i-vector based approach can be analyzed by varying the number of Gaussian mixture components and dimension of the i-vector. For each mixture component various equal error rate (EER) can be obtained. The following table 2 represents the various equal error rate (EER) by varying different number of mixture components and number of dimensions for a different speakers (speaker independent case) who uttered the commodity names.

Table – 2
EER for Speaker Independent Case (Different speakers) for Different Mixture components and Dimensions

| *Dimension* | *Various Mixture Component* | | | | | |
|---|---|---|---|---|---|---|
| | *32* | *64* | *128* | *256* | *512* | *1024* |
| *40* | *37.50* | *35.43* | *31.61* | *30.14* | *37.50* | *30.88* |
| *50* | *34.55* | *34.55* | *33.04* | *32.12* | *34.55* | *33.82* |
| *60* | *41.91* | *38.23* | *34.46* | *31.61* | *32.26* | *30.56* |
| *70* | *38.97* | *39.70* | *36.53* | *34.55* | *33.08* | *31.43* |
| *80* | *35.47* | *37.77* | *37.50* | *35.38* | *29.36* | *34.55* |
| *90* | *40.34* | *36.02* | *35.43* | *35.43* | *33.08* | *34.55* |
| *100* | *41.63* | *43.38* | *38.97* | *37.40* | *35.29* | *34.88* |

From the table 2 it is observed that the performance of difference speaker (speaker independent case) must be achieved better when increasing number of mixture components .Here larger mixture component provides the less EER. For different speakers (speaker independent case), i-vector dimension of 80 with the mixture component of 512 provides the less equal error rate of 29.36. The implementation of the i-vector based approach can be done in the Matlab R2013a by using the MSR identity toolbox. Similarly the following table 3 represents the various equal error rate (EER) by varying different number of mixture components and number of dimensions for a single speaker who uttered the commodity names.

Table – 3
EER for Speaker Dependent Case (Single Speaker) for Different Mixture Components and Dimensions

| *Dimension* | *Various Mixture Component* | | | | |
|---|---|---|---|---|---|
| | *32* | *64* | *128* | *256* | *512* |
| *50* | *35.35* | *37.85* | *34.33* | *35.00* | *35.7* |
| *60* | *32.50* | *34.46* | *34.28* | *34.45* | *33.62* |
| *70* | *33.17* | *33.92* | *32.84* | *32.85* | *35.35* |
| *80* | *36.23* | *36.09* | *33.57* | *31.92* | *35.12* |
| *90* | *33.50* | *33.35* | *33.29* | *34.39* | *31.20* |
| *100* | *38.21* | *34.64* | *33.21* | *32.50* | *30.86* |
| *110* | *37.72* | *36.05* | *36.42* | *33.47* | *33.43* |
| *120* | *37.85* | *34.64* | *35.99* | *33.21* | *31.42* |
| *130* | *40.35* | *36.42* | *34.28* | *35.35* | *32.89* |

| | | | | | |
|---|---|---|---|---|---|
| *150* | *42.50* | *40.94* | *37.14* | *34.45* | *32.55* |
| *200* | *41.15* | *40.71* | *41.07* | *37.46* | *34.03* |

From the table 3 it is observed that for the i-vector dimension of 100 with the mixture component of 512 achieves the better performance. Therefore performance must be achieved better when increasing number of mixture components. When increasing the number of mixture components EER can be reduced. In the proposed work, for speaker dependent case, obtained lowest EER is 30.86 with the dimension of 100.

Fig 4 shows the detection error trade off curve for both single speaker (speaker dependent) and different speakers (speaker independent) and it is observed that increasing the mixture component reduces equal error rate. Fig 4 and 5 shows the detection error trade off curve for both single speaker (speaker dependent) and different speaker (speaker independent) with the i-vector dimension of 100 and 80 and it is observed that increasing the mixture component reduces equal error rate. In this curve 'X' axis represents the false positive rate (FPR) and the 'Y' axis represents the false negative rate (FNR). The linear line which cuts the curve in the graph which gives the rate called equal error rate (EER).

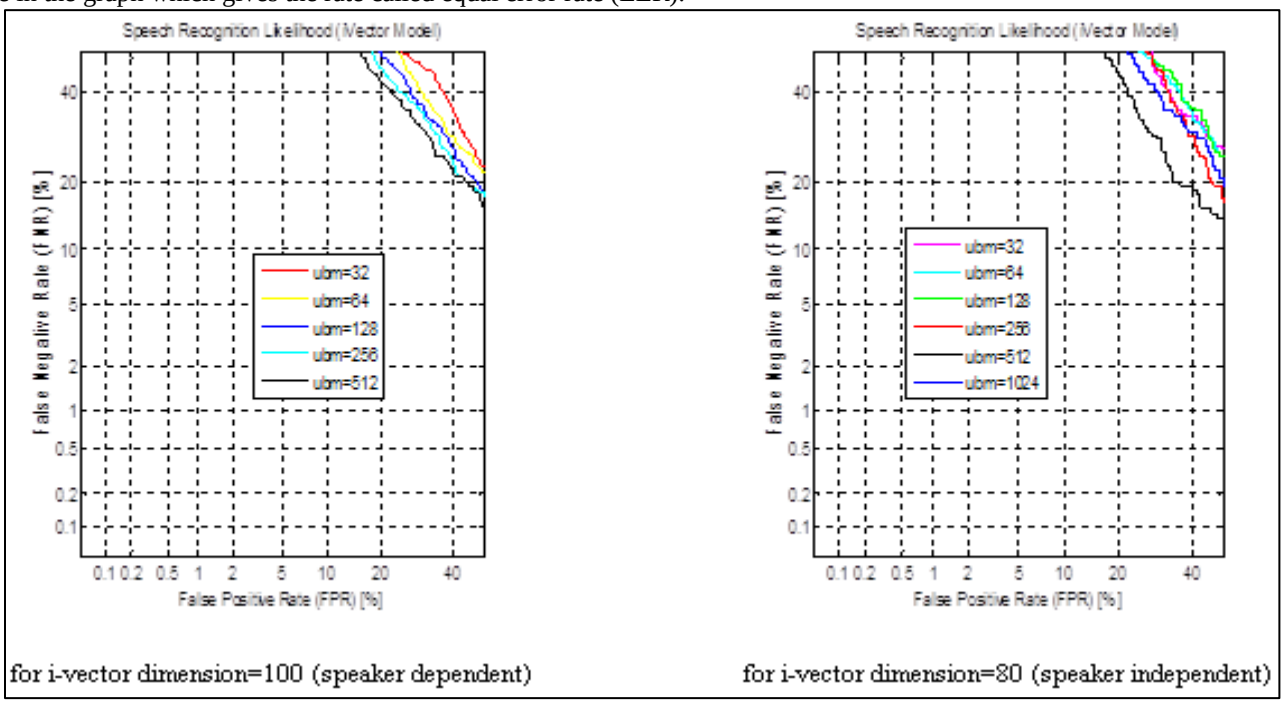


Fig. 4: DET Curve for both Speaker Dependent Case (Single Speaker) and Speaker Independent Case (Different Speakers)

## V. CONCLUSION

In this proposed method, for the applications of speech recognition the author adopting this i-vector method. As a result, we have observed that increasing number of training data will increase the performance of the system. And also by increasing the number of Gaussian mixture component performance must be improved. One of the disadvantages of proposed method is it consumes time for larger mixture component and larger dimensions. In the future work, instead of using LDA we have an idea to use within class covariance (WCCN) and cosine similarity methods in order to improve the performance.

## REFERENCES

[1] S. B. Dhonde, S. M. Jagade, "Mel-frequency cepstral coefficients for speaker recognition:A Review" International Journal of Advance Engineering and Research Development, vol. 2, May 5 2015.
[2] S. B. Dhonde, S. M. Jagade, "Feature Extraction Techniques in Speaker Recognition:A Review" International Journal onRecent Technologies in Mechanical and Electrical Engineering (IJRMEE) ISSN:2349-7947 Volume.2 Issue:5 pp.104-106, May 2015.
[3] Aggarwal R. K and M. Dave, "Using Gaussian mixtures for hindi speech recognition system" International Journal of signal processing, Image Processing and Pattern recognition vol. 4.4, pp. 157-170, 2011.
[4] Rabiner Lawerence R, "A tutorial on hidden Markov models and selected applications in speech recognitions" of the IEEE 77.2 pp. 257-286, 1989.
[5] D. A. Reynolds and D. C. Rose, "Robust text-independent speaker identification using Gaussian mixture speaker models," IEEE trans speech Audio process, vol. 3, no. 1, pp. 72-83, January 1995.
[6] B. Bharathi and T. Nagarajan, "GMM and i-vector based speaker verification using speaker-specific-text for short utterances", published in Tencon, October 2013.
[7] A. Kanagasundaram, R. Vogt, D. Dean, S. Sridharan, and M. Mason, "i-Vector Based Speaker Recognition on Short Utterances", In Proc.12th Annu. Conf. int. Speech Communication. Assoc., 2011, pp. 2341-2344.
[8] P. Kenny, "Joint Factor Analysis of Speaker and Session Variability: Theory And Algorithms," Tech. Rep. Crim-06/08-13, 2005. [Online]. Available: http://www.crim.ca/perso/patrick.kenny/
[9] J. H. L. Hansen and T.Hasan, "Speaker recognition by machines and humans: A tutorial review," IEEE signal process. Mag., vol. 32, no .6, pp. 74-99, nov.2015.

[10] Gautam Varma Mantena, S. Rajendran, B. Rambabu, Suryakanth V. Gangashetty, B. Yegnanarayana, Kishore Prahallad, "A Speech –Based Conversation System For Accessing Agriculture Commodity Prices In Indian Languages", IEEE Joint Workshop Hands-Free Speech Communication And Microphone Arrays (HSCMA), pp. 153-154 june 2011.
[11] Aanchan Mohan, Umesh S. Richard Rose, "Subspace Based Acoustic Modelling For Indian Languages", IEEE The 11th International Conference On Information Sciences, Signal Processing And Their Applications, 2012.
[12] "Speech- Based Automated Commodity Price Helpline in Six Indian Languages", http://asrmandi.wixsite.com/asrmandi
[13] Aniruddha Deka, Manoj Kumar Deka, "Speaker Independent Speech Based Telephony Service For Agro Service Using Asterisk And Sphinx 3", International Journal Of Computer Sciences And Engineering Open Access, vol 4. December 2016.
[14] Hao Chin, Jia-Ching Wang, Senior Member, IEEE, Chien-Lin Huang, Kuang-Yao Wang, And Chung-Hsien Wu, Senior Member, IEEE, "Speaker Identification Using Discriminative Features And Sparse Representation", published in IEEE at 2017.
[15] Hossein Zeinali, Bagher Babaali, "On The Usage Of i-Vector Representation For Online Handwritten Signature Verification" International Conference On Document Analysis And Recognition, At Kyoto, Japan.2017.
[16] Namrata Dave, "Feature Extraction Methods LPC, PLP and MFCC in Speech recognition", International Journal For Advance Research In Methods based on Perceptual, vol.1, July 2013.
[17] N.Dehak, P.Kenny, R.Dehak, P.Dumouchel, and P. Ouellet, "Front-end factor analysis for speaker verification," IEEE Transactions on Audio, Speech and Language Processing, vol.19, no.4 . pp.788-798, 2011.
[18] N. Dehak, P. A. Torres-Carrasquillo, D. A. Reynolds, and R. Dehak, "Language recognition via i-vectors and dimensionality reduction", in InterSpeech, 2011, pp. 857–860.
[19] D. Martınez, O. Plchot, L. Burget, O. Glembek, and P. Matejka, "Language recognition in i-vectors space", InterSpeech, pp. 861–864, 2011.
[20] R. Xia and Y. Liu, "Using i-vector space model for emotion recognition," in InterSpeech, 2012.
[21] H. Zeinali, A. Mirian, H. Sameti, and B. BabaAli, "Non-speaker information reduction from cosine similarity scoring in i-vector based speaker verification", Computers & Electrical Engineering, vol. 48, pp. 226–238, 2015.
[22] H. Zeinali, E. Kalantari, H. Sameti, and H. Hadian, "Telephony textprompted speaker verification using i-vector representation," , in Acoustics, Speech and Signal Processing (ICASSP), IEEE International Conference on, 2015, pp. 4839–4843.
[23] H. Zeinali, H. Sameti, L. Burget, J. Cernocky, N. Maghsoodi, and P. Matejka, "i vector/HMM based text-dependent speaker verification system for RedDots challenge", in InterSpeech, 2016, pp. 440–444.
[24] P. Kenny, G. Boulianne, and P. Dumouchel, "Eigenvoice modeling with sparse training data" ,Speech and Audio Processing, IEEE Transactions on, vol. 13, no. 3, pp. 345–354, 2005.
[25] Fahimeh Bahmaninezhad, John H.L. Hansen, "i-Vector/PLDA speaker recognition using support vectors with discriminant analysis" IEEE international conference in acoustics,speech and signal processing, 2017.