IJESRT

INTERNATIONAL JOURNAL OF ENGINEERING SCIENCES & RESEARCH TECHNOLOGY

An Implementation of Text Dependent Speaker Independent Isolated Word Speech Recognition Using HMM

Ms. Rupali S Chavan[*1], Dr. Ganesh S. Sable[2]

[*1, 2] Department of electronics engineering, Savitribai Phule Women's Engineering College, Aurangabad - 431005, Maharashtra, India

Chavanrupali452@gmail.com

Abstract

Speech is the powerful tool of information exchange. There are different aspects related to speech like speech recognition, speech verification, speech synthesis, speaker recognition, speaker identification etc. Speech Recognition is the process of determining which speech is present based on individual's utterance. This paper gives an implementation& performance evaluation of Text Dependent Speaker Independent Isolated Word Speech Recognition System is using HMM. Here MFCC is used for speech parameterization. These feature parameters are used for HMM training. In HMM modeling forward backward algorithm with EM principle is used for parameter estimation & optimization. Initially K means algorithm is used for dividing feature dataset into smaller parts and then means are calculated.GMM is used to model the distribution of speech features for each state of HMM. Finally the computed HMM parameters for words are stored in respective HMM models as a reference database. To recognize the spoken word the likelihood of generation of test speech observations and the most likely path sequence through each stored HMM model models is calculated. Finally the one with maximum likelihood path is selected as recognized word. The system is implemented in MATLAB 7.9.The system is trained with own created database consisting 60 speech samples of selected words & is tested for speaker independent mode. For selected three words recognition accuracy of 92%, 92% & 88% respectively is obtained in noisy environment.

Keywords: Automatic Speech Recognition, Gaussian Mixture Model, Hidden Markov Model, Mel Frequency Cepstral coefficients, Speech recognition.

Introduction

The process of automatically recognizing spoken words of speaker based on information in speech signal is called Speech Recognition. In automatic speech recognition computer captures the words spoken by a human with a help of microphone. These words are then recognized by automatic speech recognizer, and in the end, system displays the recognized words on the screen.

The real time automatic speech recognition system faces big challenge of increasing accuracy and recognition speed. The performance of speech recognition system degrades due to noise. It also gets affected by varying speech data due to dependence on gender of speaker, environmental conditions and speaking styles. The recognition accuracy depends on the method of feature extraction and training so recognition accuracy is one of the core issues of speech recognition research. The objective of this system is to develop speech recognition system with less recognition time and high recognition accuracy. Also to improve system performance in presence of noise with the use of noise robust feature extraction and training technique.

This system can be used to provide security to different systems. One can access the system if the spoken word is recognized. Also this project can be easily extended for voice dialing applications, voice command recognition systems, voice interactive systems and appliance control system through voice. Speech recognition technology is one from the fast growing engineering technologies. It has a number of applications in different areas and provides potential benefits. Nearly 20% people of the world are suffering from various disabilities; many of them are blind or unable to use their hands effectively. The speech recognition systems in those particular cases provide a significant help to them, so that they can share information with people by operating devices through speech input.

In early research Mahdi Shaneh and Azizollah Taheri suggested the "Voice command recognition system based on MFCC and VQ algorithms". They designed a system to recognition voice commands. They used MFCC algorithm for feature extraction and VQ (vector quantization) method for reduction of amount of data to decrease computation time. In the feature matching stage Euclidean distance is applied as similarity criterion. Because of high accuracy of used algorithms, they got the high accuracy voice command system. They trained initially with one repetition for each command and once in each in testing sessions and got 15% error rate. Secondly they increased the training samples then got zero error rate [2].In their research H.P. Combrinck and E.C. Botha reported "On The Mel-scaled Cepstrum''. They reported on superior performance of MFCC especially under adverse conditions. Also concluded that it represents a good trade-off between computational efficiency and perceptual considerations [3]. In their research "Voice recognition algorithms using MFCC and DTW" Lindasalwa Muda, Mumtaj Begam and I. Elamvazuthi concluded that with these techniques the particular speaker was correctly recognized. They found that DTW is best for non linear time (speech) sequence alignment [4]. Another research was done by Ahmad A. M. Abushariah, Teddy S. Gunawan, and Othman O. Khalifa in their paper **"**English Digits Speech Recognition System Based on Hidden Markov Models". Two modules were developed, namely the isolated words speech recognition and the continuous speech recognition. Both modules were tested in both clean and noisy environments and showed a successful recognition rates. These recognition rates are relatively successful if compared to similar systems. The recognition rates of multi-speaker mode performed better than the speaker-independent mode in both environments. For multi speaker mode they got 99.5% whereas for speaker independent 79.5% accuracy [5]. Then Ibrahim Patel and Dr Y. Shrinivasa Rao in their research paper "Speech recognition using Hidden Markov Model with MFCC Subband technique" concluded that with these methods quality metrics of speech recognition with respect to computational time, learning accuracy get improved [6].In another research M. A. M. Abu Sarah, R. N. Anion, R. Zainuddin, and O. O. Khalifa in their research paper "Human computer interaction using isolated-words speech recognition technology" developed an isolated-word automatic speech recognition (IWASR) system based on vector quantization (VQ). They showed that system receives, analyzes searches and matches an input speech signal with the trained set of speech signals which are stored in the database/codebook, and returns matching results to users. Their experimental results showed that the recognition rate has been improved with the increase of codebook size and showed that the codebook size of 81 feature vectors had a recognition rate exceeded 85% [7].In another research of "Voice Recognition using HMM with MFCC for secure ATM" by Shumaila Iqbal, Tahira Mahboob and Malik Sikandar recognition accuracy was found to be 86.67% [8]. In another research by Ms. Vrinda, Mr. Chander Shekhar in "Speech Recognition system for English language'' concluded that the accuracy of the system depends upon the training time. Time is directly proportional to accuracy .If training time increases then accuracy will increase automatically [9]. Sharada C. Sajjan, Vijaya. C gave "Comparison of DTW and HMM for isolated word recognition" with LPC and MFCC as feature extraction. They concluded that MFCC analysis provides better recognition rate than LPC as it operates on a logarithmic scale which resembles human auditory system whereas LPC has uniform resolution over the frequency plane. Experimentally they observed that recognition accuracy is better for HMM compared with DTW. They used the database TI-46 isolated word corpus zero-nine from Linguist Data Consortium. They showed that recognition accuracy is 69% for LPC & DTW, 86% For LPC & HMM, 77% for MFCC & DTW and 90% for MFCC & HMM [10].

In this paper a Text Dependent Speaker Independent, Isolated English Word Recognition Speech Recognition System is implemented and its performance is evaluated using own created database based on correct acceptance and correct rejection rate. This system gives speech interface which helps us to tackle various issues like firstly problems of Physically challenged people to handle/use various kinds of systems(computer, household appliances etc),secondly to prevent unauthorized access to human's computerized and electronic belongings by his/her speech providing security, Thirdly for some applications like voice dialing, voice responsive systems etc.

The system uses MFCC for feature extraction, K means algorithm for finding means, GMM for acoustic modelling, in which each speech state is distributed by a Gaussian Mixture Model (GMM), HMM for training/learning and Viterbi algorithm for decoding. The use of MFCC for extracting speech features and HMM/GMM for recognition provides an improved accuracy in real time scenario. The system is implemented using MATLAB 7.9. As system is text dependent three words as three names Kavita, Sonali & Yogesh are selected. For each word 20 samples are collected from different speakers. Thus database consists of total 60 samples. These samples are collected through pre processing step which involves recording & sampling. After preprocessing features are extracted through MFCC & these features are trained using HMM. Using viterbi decoding test speech is matched with reference HMM models by finding most probable state sequence & recognition decision is made. The testing is carried out using 5 speakers whose samples were not used in database (speaker independent).Then wordwise correct acceptance and correct rejection rate is found.

Section II demonstrates proposed HMM/GMM Based System & Methodology, Section III gives Experimental Results & Section IV gives conclusion.

Proposed HMM/GMM Based System

The system implemented here consists of four main phases Pre processing, Feature Extraction, Pattern Training and Decision Logic. For the proposed system 3 words are taken. The system uses MFCC technique to describe the acoustic features of spoken words. Then GMM with k means clustering is used to make observation likelihood vector for mixtures of Gaussians of HMM. K means algorithm is used for Clustering, by which large sets of features (observations) is grouped into k clusters of smaller sets of similar observations and it finds means required to find observation likelihood vectors by GMM.GMM is used to model the distribution of speech features to each state of HMM. HMM uses EM algorithm with forward backward estimation technique to train these observations to compute the maximum likelihood HMM parameters and to find the log likelihood for each word .In recognition the likelihood value of test observation with each HMM model is computed using Gaussian probability density function. Then viterbi algorithm is used to find find the most-probable (Viterbi) path through each HMM state trellis. This procedure is applied for all HMM models and finally the one with maximum probable path is selected and the word corresponding to that particular HMM is the recognized spoken word. This section explains the proposed system block diagram and methodology step by step.

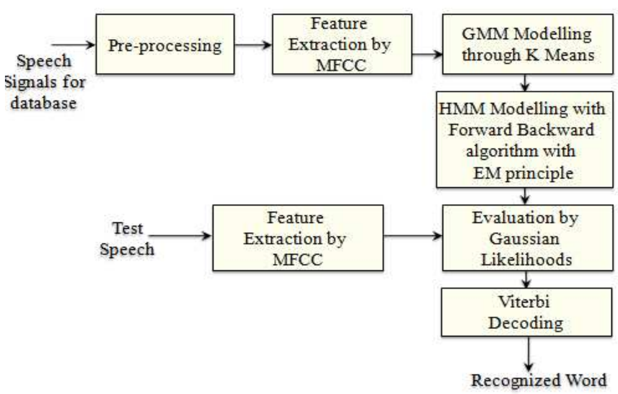


Figure 1: System Architecture

i) Pre-processing

In pre processing the 16 bits of speech wave is recorded for 2 sec & sampled at 2xFs.Here Fs=11025, so speech is sampled at 22050 Hz.

ii) Feature Extraction by MFCC

The purpose of this module is to convert the speech waveform to some type of parametric representation. MFCC is used to extract the unique features of speech samples. It represents the short term power spectrum of human speech.

The sampled signal is given to the feature extractor(MFCC).In the very first step the sampled signal is passed through the one coefficient pre emphasis filter where a is selected to be 0.95 .In this step the energy of signal is increased at higher frequency. After this the pre emphasized signal is segmented into 171 frames each of N=256 samples. The frame size for 11025 Hz signal is 23.21ms.Adjacent frames are separated by M=128 samples or by 11.6 ms (M<N).Then each individual frame is windowed so as to minimize the signal discontinuities at the beginning and end of each frame. Hamming window is used as window and it integrates all the closest frequency lines. The length of window is N=256.The Fast Fourier Transform of windowed signal is taken to convert each frame of 256 samples from time domain into frequency domain. It requires number of samples N to be power of 2 (e.g. N=256=2^8). Then the resultant signal is processed with 40 Mel Filter Bank. The frequency range in FFT spectrum is very wide and voice signal does not follow the linear scale. It gives a weighted sum of filter spectral components so that the output of process approximates to a Mel scale. The output is Mel spectrum consists of output powers of these filters. Then its logarithm is taken and output is log mel spectrum. Taking the log of each of the 40 energies leaves us with 40 log filter bank energies. Then to obtain MFCC coefficients Discrete Cosine Transform is taken. This is the process to convert the log Mel spectrum into time domain. The Discrete Cosine Transform (DCT) of the 40 log filter bank energies gives 40 cepstral coefficients. For ASR, only the lower 14 of the 40 coefficients are kept. The resulting features (14 numbers for each frame) are called Mel Frequency Cepstral Coefficients.

iii) HMM/GMM Training

The features extracted by MFCC are used for pattern training by HMM/GMM. The feature data sets is divided into 3 smaller sets and for each set 14 means are found by k means clustering. Gaussian mixture model is created with specified architecture with dimension 14, n-centers 1, centers (means) 14, covariance diagonal.GMM evaluates the Observation likelihood and the matrices initial state probability matrix {π}, state transition matrix {A} and observation probability matrix {B} required for HMM are initialized. Forward backward algorithm with EM estimates & optimizes the HMM parameters.

iv) Recognition/Decoding

The viterbi algorithm is used to find the most likely state sequence when HMM models for all words and unknown observations (test speech features) are given. Then from these three most likely sequences the one with maximum likelihood is selected. Then the word corresponding to the HMM with which most probable state sequence was observed, is nothing but recognized word.

Experimental Results

i) Speech Database

Speech recognition is carried out using our own created speech database of three word category. This system had four different speakers whose speech samples were collected for training. The speakers include two female & two male speakers belonging to different ages & genders.

Speakers Profile for creation of database is given in table 1

Table 1: Speakers Profile

Age	Number of speakers	Gender
Adults ranging from 24 to 35 years	2	Male
	2	Female

- The selected words for the system are Kavita, Sonali, and Yogesh.
- Samples for each word: 20
- Total Samples in database: 3*20= 60.
- Word wise Speaker Profile for database is as given bellow in Table

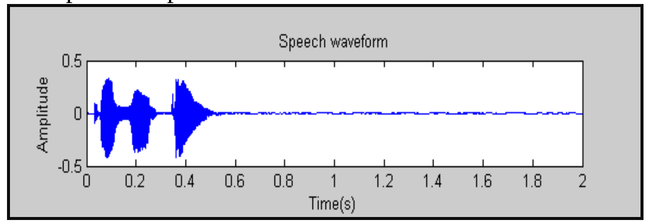
Table 2: Wordwise Speaker Profile

	Kavita	Sonali	Yogesh
Samples	1Male-10 samples	1 Male-5 samples	1Male-10 samples
	1 Female-10 samples	2 Female-15 samples	1 Female-10 samples
Total	20	20	20

Here the results are shown step by step for one sample of word Kavita, kv1.wav (Female speech sample) which was stored in database.

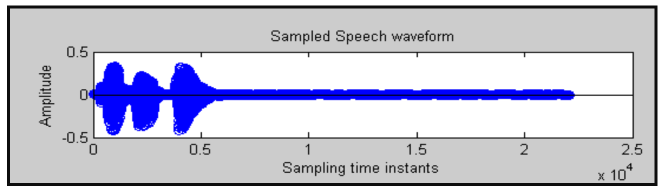
ii) Pre-Processing

The speech samples are taken using microphone & sampled with rate 22050Hz.The speech signal is recorded for 2 sec. For first sample of word kavita recorded file is saved as kv1.wav. The speech waveform of recorded kv1.wav of sample kavita is as given below. The speech amplitude is shown for 2 sec of recorded file.



Graph 1: Speech Waveform

The speech signal is sampled at the rate of 2xFS. Thus for each speech wav file we get sampled file of size<22050 x 1>. The graph given below shows sampled amplitude of speech waveform for 22050 time instants for wave file kv1.wav.

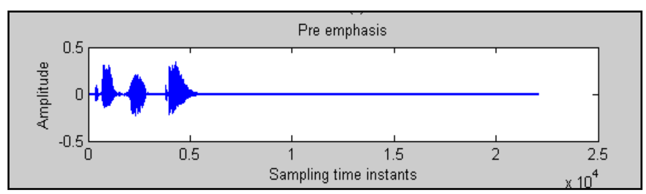


Graph 2: Sampled Speech Waveform

iii) Performance Analysis by MFCC

Step 1: Pre-emphasis

The sampled signal is passed through the one coefficient pre emphasis filter where 'a' is selected to be 0.95. In this step the energy of signal is increased at higher frequency.

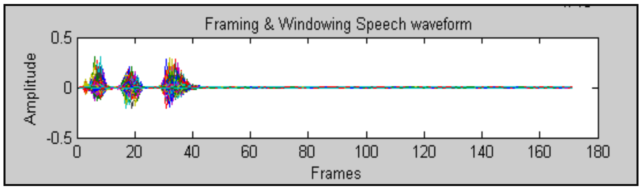


Graph 3: Pre-emphasis

Step 2: Framing & Windowing

The pre emphasized signal is segmented into 171 frames each of N=256 samples. The frame size for 11025 Hz signal is 23.21ms.Adjacent frames are separated by M=128, 11.6ms (M<N).Each individual frame is windowed by multiplying with Hamming Window. The length of window is N=256. The graph given below shows the 171 windowed frames for wave file kv1.wav.

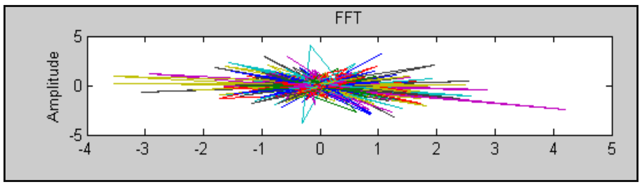
ISSN: 2277-9655
Impact Factor: 1.852



Graph: 4 Framing & Windowing Speech Waveform

Step 3: Fast Fourier Transform

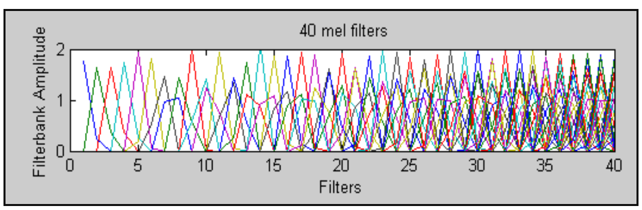
The Fast Fourier Transform of windowed signal is taken to convert each frame of 256 samples from time domain into frequency domain. It requires number of samples N to be power of 2 (e.g. N=256=2^8). The graph given below shows the FFT for wave file kv1.wav.



Graph 5: Fast Fourier Transform

Step 4: Mel Filter Bank Processing

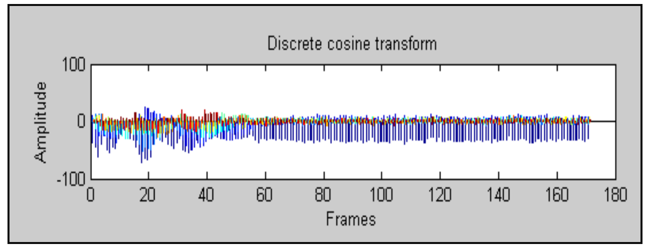
The FFT spectrum is processed with 40 Mel Filter Bank. The output is mel spectrum consists of output powers of these filters. Then its logarithm is taken and output is log mel spectrum. Taking the log of each of the 40 energies leaves us with 40 log filter bank energies. The graph given below shows the 40 Mel filter amplitude for wave file kv1.wav.



Graph 6: 40 Mel Filters

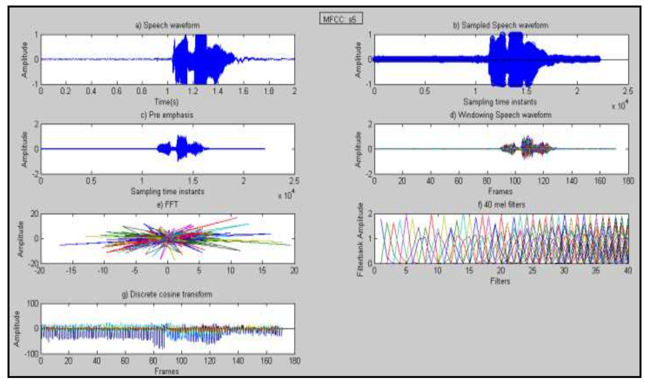
Step 5: Discrete Cosine Transform

The Discrete Cosine Transform (DCT) of the 40 log filter bank energies gives 40 cepstral coefficients. For ASR, only the lower 14 of the 40 coefficients are kept. The graph given below shows the discrete cosine transform (DCT) i.e 14 MFCC coefficients for 171 frames for wave file kv1.wav



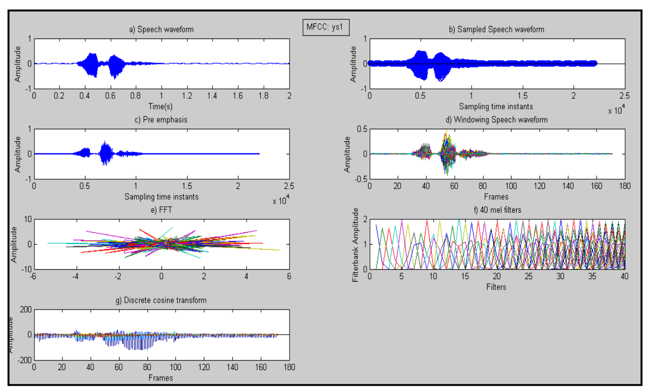
Graph 7: Discrete Cosine Transform

Similarly the MFCC result for one sample of word Sonali (s5.wav) is given below in one MATLAB figure window.



Graph 8: MFCC for s5 (Male Speech Sample: Sonali)

The MFCC result for one sample of word Yogesh (ys11.wav) is given below in one MALAB figure window.



Graph 9: MFCC for ys11 (Male Speech Sample: Yogesh)

Thus for each of 60 speech samples we get feature vector of size<14x171>.The calculated MFCC features for word Kavita are stored in data_1,for word Sonali in data_2,for word Yogesh in data_3 as given in Table 4.7.

Table 3: MFCC Result Storage

Sr. No	Word	MFCC features <1 x 20 cell> array
1	Kavita	data_1
2	Sonali	data_2
3	Yogesh	data_3

The data_1 is array of size <1x20 cell>, consists of feature vectors for all 20 samples of word Kavita. So there are 20 cells in a array and each cell is of size<14 x171>.Similarly for data_1 and data_3 consists of 20 cells for 20 samples, each cell is of size <14x171>.

iv) HMM

The features extracted by MFCC are used for pattern training by HMM. The HMM model is represented by λ = {π, A, B}.Here 3 state left to right phoneme model is created. So for our first word Kavita trained data will be stored in HMM_1 similarly for second word Sonali HMM_2 and for third word Yogesh HMM_3.

Table 4: HMM Result Storage

Sr. No	Word	Training data
1	Kavita	HMM_1
2	Sonali	HMM_2
3	Yogesh	HMM_3

Each HMM models consists of means, variances, initial state matrix {π}, state transition matrix {A}, Observation probability matrix {B}.

For word Kavita trained HMM_1 model consisting of A, B, π & means, variances is as given below.

HMM_1 <1x1 struct>

Field	Value	Min	Max
pi	[1.0000;5.5524e-14;0]	0	1.0000
trans	[0.9825,0.0175,0;0,0.9...	0	1
means	<14x3 double>	-33.8093	7.4200
vars	<14x14x3 double>	0	516.51...
mm	[1;1;1]	1	1

Figure 2: HMM_1 for Word Kavita

For word Sonali trained HMM_2 model consisting of A, B, π & means, variances is as given below.

HMM_2 <1x1 struct>

Field	Value	Min	Max
pi	[1;8.3458e-129;0]	0	1
trans	[0.9761,0.0239,7.2917...	0	1
means	<14x3 double>	-42.1542	7.0864
vars	<14x14x3 double>	0	644.40...
mm	[1;1;1]	1	1

Figure 3: HMM_2 for Word Sonali

For word Yogesh trained HMM_3 model consisting of A, B, π & means, variances is as given below.

HMM_3 <1x1 struct>

Field	Value	Min	Max
pi	[1;8.0457e-117;0]	0	1
trans	[0.9808,0.0192,4.6542...	0	1
means	<14x3 double>	-55.1927	5.6300
vars	<14x14x3 double>	0	1.6378...
mm	[1;1;1]	1	1

Figure 4: HMM_3 for Word Yogesh

v) Evaluation and Success Criterion

At first the recognition accuracy of our system will be evaluated with three performance metrics (recognition rate, false recognition rate, recognition speed). The results are tested against the specified objectives of proposed system. For testing, test samples from 3 male & 2 female were taken. These are the speakers whose samples were not used in database (speaker independent).The developed system is tested by taking five speech samples continuously without repetition from each speaker for each word. The testing is done in noisy environment. Testing was carried in noisy rooms where TV & fan was ON; also for some speakers testing was done in noisy laboratories. Accuracy rate shows the percentage of correctly identified test samples by the system per word.

a) Correct Recognition Rate/Recognition accuracy

It is defined as follows:

$$\text{Recognition accuracy} = \frac{\text{No. of correctly recognized speech samples}}{\text{Total no. of testing samples}} \times 100$$

The results of testing are as given below. The correctly recognized samples are labeled as T (True) and the samples which are not recognized correctly are labeled as F (False).

Table 5: Speech Recognition Test Results

Words	Sample No.	Speaker 1	Speaker 2	Speaker 3	Speaker 4	Speaker 5
Kavita	1	T	T	T	T	T
	2	T	T	T	T	F
	3	T	T	T	F	T
	4	T	T	T	T	T
	5	T	T	T	T	T
Sonali	1	T	T	T	T	T
	2	T	T	T	T	T
	3	T	T	T	T	T
	4	T	T	T	F	T
	5	T	T	T	F	T
Yogesh	1	T	T	T	T	T
	2	F	T	T	T	T
	3	T	F	T	T	T
	4	T	T	T	T	T
	5	F	T	T	T	T

Table 6: Correct Recognition accuracy

Words	Speaker 1	Speaker 2	Speaker 3	Speaker 4	Speaker 5	Avg accuracy
Kavita	100%	100%	100%	80%	80%	92%
Sonali	100%	100%	100	60	100%	92%
Yogesh	60%	80%	100%	100%	100%	88%

b) False Recognition Rate/Error rate/Correct Rejection

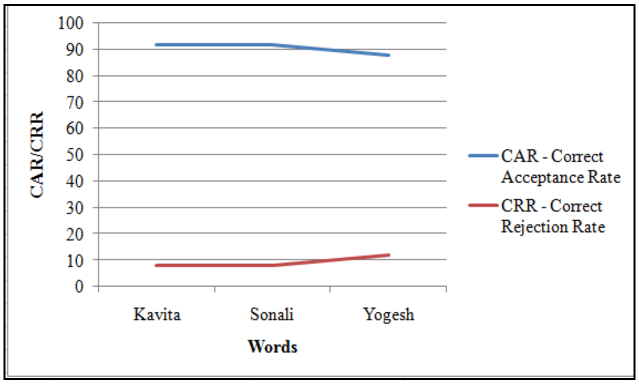
The error rate of total test samples that are being falsely recognized is calculated as follows:

$$\text{Error rate} = \frac{\text{Falsely recognized samples}}{\text{Total test samples}} \times 100$$

Table 7: False Recognition Rate

Words	Speaker 1	Speaker 2	Speaker 3	Speaker 4	Speaker 5	Avg Error rate
Kavita	0%	0%	0%	20%	20%	8%
Sonali	0%	0%	0%	40%	0%	8%
Yogesh	40%	20%	0%	0%	0%	12%

The graph of CAR Vs CRR in percentage for the system is given below. It shows that recognition rate is greater than error rate even when testing was done in noisy conditions.



Graph 10: CAR/CCR vs Words

vi) Recognition Time

The test speech is recorded for 2 sec. After that the system takes less than 2 sec to give recognized word. For word Kavita it is found approximately 1.92 sec, for word Sonali 1.84 sec and for Yogesh 1.45 sec. Thus the recognition time is less than 2 sec .This time is recorded for one test sample of Kavita, Sonali & Yogesh.

vii) Comparison with Existing Systems

The performance of our proposed system is compared with other existing system based on recognition accuracy. Following table gives comparison with existing systems.

Table 8: Comparison with Existing Systems

Ref. No.	Name of author	Technology Used	Remark
7	A. M. Abushariah, Teddy S. Gunawan, Othman O. Khalifa	MFCC & HMM	Recognition rate was 79.5% for clean speech & 67% for noisy speech speaker independent mode
9	Shumaila Iqbal, Tahira Mehboob, Malik	MFCC & HMM	Recognition rate was 86.67%
15	M. A. M. Abu Shariah, R. N. Ainon, R. Zainuddin, and O. O. Khalifa Ahmad	MFCC & VQ	Recognition rate was 88.88%
16	Sharada C.Sajjan, Vijaya C	LPC Vs MFCC DTW Vs HMM	Recognition accuracy 69% for LPC & DTW, 86% for LPC & HMM, 77% for MFCC & DTW, 90% for MFCC & HMM
Proposed System		MFCC & HMM/GMM	Recognition accuracy (88% to 92%), for first word 92%, for second word 92% & for third word 88%

For our system with MFCC & HMM the accuracy varies from speaker to speaker & word to word, here the accuracy obtained for first two words is 92% and for third word 88%.Here the accuracy is calculated by taking continuous 5 samples for a word for each speaker. Thus it shows that recognition accuracy is comparatively improved.

Conclusions

The method of MFCC for feature extraction and HMM for training is used in proposed system. It is tested upon own created database in different environments. The performance of the proposed method in terms of recognition accuracy is obtained. It shows that the proposed method has high recognition accuracy which illustrates the robustness of the proposed method against the speaking variations like speaking style, gender, speaking environment etc. Extensive training and testing experiments are carried out in order to demonstrate the effectiveness of the proposed method for speech recognition.

Speech recognition using HMM gives good result due to resemblance between architecture of HMM and varying speech data. The recognition accuracy increases due to combination of MFCC and HMM in noisy environments.

The performance is evaluated by finding word wise accuracy for five different speakers whose samples were not collected in data base .The accuracy varies from speaker to speaker & word to word, Here the accuracy obtained for first two words is 92% and for third word 88%.

References

[1] M. A. Anusuya, S. K. Katti “Speech Recognition by Machine: A Review” International journal of computer science and Information Security 2009.

[2] S D. B. Paul, Tutorial on,”Speech Recognition using Hidden Markov Model.”

[3] Mahdi Shaneh, and Azizollah Taheri,”Voice Command Recognition System Based on MFCC and VQ Algorithms”, World Academy of Science, Engineering and Technology 57 2009.

[4] L. R. Rabiner and B. H. jaung ,” Fundamentals of Speech Recognition Prentice-Hall, Englewood Cliff, New Jersy,1993.

[5] Lindasalwa Muda, Mumtaj Begam and I. Elamvazuthi, “Voice Recognition Algorithms using Mel Frequency Cepstral Coefficient (MFCC) and Dynamic Time Warping (DTW) Techniques” Journal of Computing, Volume 2, Issue 3, March 2010, ISSN 2151-9617.

[6] H. Combrinck and E. Botha, “On the mel-scaled cepstrum,” Department of Electrical and Electronic Engineering, University of Pretoria.,Journal of Computer Science 3 (8): 608-616, 2007 ISSN 1549-3636.

[7] Ahmad A. M. Abushariah,Teddy S. Gunawan, Othman O. Khalifa “English Digits Speech Recognition System Based on Hidden Markov Models”, International Islamic University Malaysia, International Conference on Computer and Communication Engineering (ICCCE 2010), 11-13 May 2010, Kuala Lumpur, Malaysia.

[8] Ibrahim Patel,Dr.Y.Srinivasa Rao, , “Speech recognition using Hidden Markov Model With MFCC-Subband Technique.” 2010 International Conference on Recent Trends in Information, Telecommunication and Computing.

[9] Shumaila Iqbal, Tahira Mehboob, Malik, ”Voice Recognition using HMM with MFCC for secure ATM”,IJCS Vol.8,Issue 6 Nov 2011.

[10] Ms. Vrinda ,Mr. Chander Shekhar,”Speech recognition for English language”, International Journal of Advanced Research in Computer and Communication Engineering Vol. 2, Issue 1, January 2013.

[11] Anjali Bala, ABHIJEET Kumar, Nidhika Birla, “Voice command recognition System Based on MFCC and DTW”, International Journal of Engineering Science and Technology,Vol.2(12),2010.

[12] J. W. Picone, "Signal modelling technique in speech recognition," Proc. Of the IEEE, vol. 81, no.9, pp. 1215-1247, Sep. 1993.

[13] Vimala C, Dr.V.Radha, “A Review on Speech Recognition Challenges and Approaches”, World of Computer Science and Information Technology Journal (WCSIT) ISSN: 2221-0741 Vol. 2, No. 1, 1-7, 2012.

[14] Lawrence R. Rabiner, Fellow, IEEE ‘A Tutorial On Hidden Markov Model And Selected Applications In Speech Recognition, Proceedings Of The IEEE, Vol. 77, No. 2, February 1989.

[15] M. A. M. Abu Sarah, R. N. Anion, R. Zainuddin, and O. O. Khalifa ,“ Human Computer Interaction Using Isolated-Words Speech Recognition Technology,” IEEE Proceedings of The International Conference on Intelligent and Advanced Systems (ICIAS’07), Kuala Lumpur, Malaysia, pp. 1173 – 1178, 2007.

[16] Sharada C. Sajjan, Vijaya C, Dept of E&TC ,SDM COE, Dharwad, Karnataka, India, ”Comparison of DTW and HMM for isolated word Recognition ”, Proceedings of International Conference on Pattern Recognition , March 2012.