ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 8, August 2015

Speech Processing For Secluded Marathi Words Recognition Using MFCC Features

Bharatratna P. Gaikwad, Pawan S. Kamble

Assistant Professor, Department of CS and IT, Dr. B. A.M. University, Aurangabad (MS), India

ABSTRACT: This paper proposes a speaker recognition system for secluded word recognition system based on a dynamic features of the Marathi speech signals by exploring Mel Frequency Cepstral Coefficient (MFCC) methods for feature extraction techniques. In this paper numerous technique are extracting feature from marathi speech signal using MFCC. The feature parameters are extracted from Marathi speech Signals depend on speaker.Speaker Recognition is a method of automatically recognizing who is speech utterance basis of the separate information encompassed in speech signal. Speaker Recognition is one of the most beneficial biometric recognition techniques in this world where insecurity is a major threat. The implementation of speech recognition as medium security admittance control to constrained services such as phone banking system, voicemail or access to database services. The main focus of this research is speaker identification, which compared speech signal from unknown speaker to a database of known speaker using text-dependent utterances. From the experimental results, this method has showed that it was able to recognize the correct voice perfectly. The system is tested with the aid of Mel Frequency Cepstral coefficient with 12 and classification is done with Euclidian Distance. The performance of the system is tested on the basis of different three groups of speaker as 05Male speaker, 05 Female speaker, randomly selected speaker. The performance of system is calculated using FAR and FRR. The accuracy of the Male subject group is less than the accuracy of female subject group. The overall performance of the selected subject group is 81.33%.This is the overall accuracy rate of our Marathi Speaker Recognition System for Isolated Words (MSRSFIW) system.

KEYWORDS: Speech Recognition, Speaker Recognition, Types of Speech Recognition System, Mel-Frequency Cepstral Co-efficient.

I. INTRODUCTION

Speech is the most commonly and extensively used form of communication between humans. There are various spoken languages which are used throughout the world. The communication among the human being is mostly done by vocally, therefore it is natural for people to expect speech interfaces with computer [1].

Recognizing a person by her/his voice is known as speaker recognition. Development of Speech recognition systems has attained new heights but robustness and noise tolerant recognition systems are few of the problems which make speech recognition systems inconvenient to use [2]. Many Research projects have been completed and currently in progress around the world for the development of robust speech recognition systems. There are various languages in the world that are spoken by human beings for communication [3]. The computers system which can understand the spoken language can be very useful in various areas like agriculture, health care and government sectors etc. Speech recognition refers to the ability of listening spoken words and identifies various sounds present in it, and recognizes them as words of some known language [4].

Speech signals are quasi-stationary signals. When speech signals are examined over a short period of time (5-100 msec), its characteristics are stationary; but, for a longer period of time the signal characteristics changes; it reflects to the different speech sounds being spoken. Features are extracted from the speech signals on the basis of short term amplitude spectrum (phonemes). Feature extraction is the most important phase in speech recognition system. There are some problems which are faced during the feature extraction process because of the variability of the speakers [5].The Based on major advances in statistical modelling of speech, automatic speech recognition systems currently find widespread application in tasks that require human machine interface, such as automatic call processing in telephone networks, and query based information systems that provide updated travel information, stock price

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 8, August 2015

quotations, weather reports, Data entry, voice dictation, access to information: travel, banking, Commands, Avoinics, Automobile portal, speech transcription, Blind people supermarket, railway reservations etc. Speech recognition technology was progressively used within telephone networks to automate as well as to improve the operator services.

Literature Survey

In previous works several features are extracted for classifying speech affect such as energy, pitch, formants frequencies, etc. all these are prosodic features. In general prosodic features are primary indicator of speaker's emotional state. In feature extraction techniques are used for features extraction as Mel Frequency Cepstral Coefficient (MFCC) and Mel Energy spectrum Dynamic coefficients (MEDC), LPCC, PCA, and LDA [4].all various types of feature extraction methods for speech analysis for showing in following table 1.

Table 1. Feature Extraction Methods

Sr. No.	Techniques	Property	Methods for execution
1	Principal Component analysis (PCA)	Non-linear feature extraction method, Linear map, fast, eigenvector-based	Traditional, eigenvector base method, also known as karhuneu-Loeve expansion; good for Gaussian data
2	Linear Discriminate Analysis(LDA)	Non-linear feature extraction method, Supervised linear map; fast, eigenvector-based	Better than PCA for classification
3	Linear Predictive coding	Static feature extraction method,10 to 16 lower order coefficient,	It is used for feature Extraction at lower order
4	Mel-frequency scale analysis	Static feature extraction method, Spectral analysis	Spectral analysis is done with a fixed resolution along a Subjective frequency Scale i.e. Mel-frequency Scale.
5	Mel-frequency cepstrum (MFFCs)	Power spectrum is computed by performing Fourier Analysis	This method is used for find our features
6	Dynamic feature extractions i)LPC ii)MFCCs	Acceleration and delta coefficients i.e. II and III order derivatives of normal LPC and MFCCs coefficients	It is used by dynamic or runtime Feature

Types of Speech Recognition:

A. ***Isolated Words***: Isolated word recognizers regularly need each utterance to have quiet (absence of an audio signal) on both sides of the sample window. It accepts single words or single utterance at a time. These systems have "Listen/Not-Listen" states, where they require the speaker to wait between utterances. Isolated Utterance might be a better name for this class.

B. ***Connected Words***: Connected word systems are related to isolated words, but allows separate utterances to be run combine with a minimal pause between them.

C. ***Continuous Speech***: Continuous speech recognizers allow users to speak almost naturally, while the computer determines the content. Recognizers with continuous speech proficiencies are some of the most difficult to create because they utilize special methods to define utterance boundaries.

II. RELATED WORK

Anatomical structure of the vocal tract is unique for every person and hence the voice information available in the speech signal can be used to identify the speaker. Recognizing a person by her/his voice is known as speaker recognition. Since differences in the anatomical structure are an intrinsic property of the speaker, voice comes under the category of biometric identity. Using voice for identity has several advantages. One of the major advantages is remote person authentication. The feature extraction plays an important role in speaker recognition. The figure 1 describes the typical speaker recognition system.

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 8, August 2015

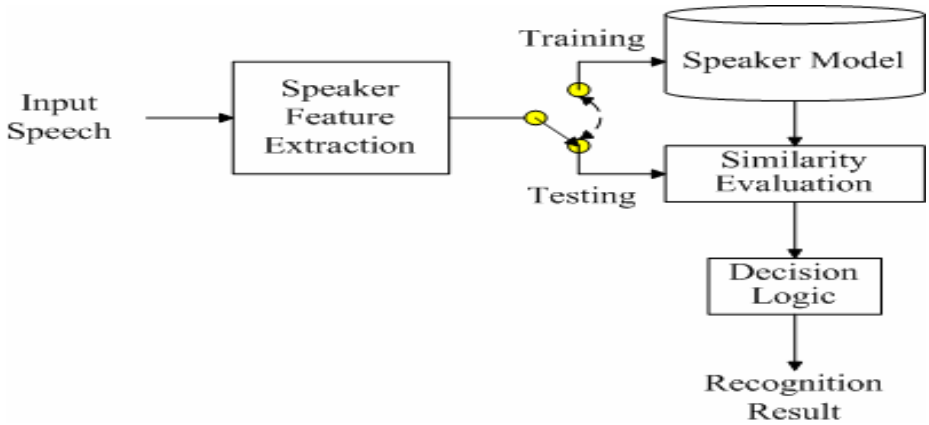

Fig.1.Illustrative speaker verification system

Training is the process of familiarizing the system with the voice characteristics of the speakers registering. Testing is the actual recognition task. Feature vectors representing the voice characteristics of the speaker are extracted from the training utterances and are used for building the reference models. During testing to check similarity feature of previous training sample for evaluation as shown in figure 1, similar feature vectors are extracted from the test utterance, and the degree of their match with the reference is obtained using some matching technique.

III. PROPOSED SYSTEM

A. Database Development :

The database for research is recorded by 50 UG and PG students of High-Tech College of computer science and Department of computer science and IT, Dr. Babasaheb Ambedkar Marathwada University. All databases are recorded by 05 Male and 05 Female speakers. The 05 word selected which are started from Vowel such as Aadarniy, Amrut etc. The database is recorded in Day time at classroom environment. The database is recorded using Praat software and Iball 2.0 Microphone. The detail technical specification of the database is given below table 2.

Table 2. Detail technical specification of the database

Sr.No.	Parameter	Values
1	Speaker type	Students (UG/PG)
2	Gender	05 Male, 05 Female
3	Basic language	Marathi
4	Accent	Marathi
5	Isolated word	05
6	Uttrances	05
7	Region	Marathwada
8	Environment	Classroom Environment

The detail vocabulary word of database development is given below. The word is selected as vowel based word. The detail pronunciation and word are described in table 3.

Table 3: The detail vocabulary word of database creation

Sr.No.	**Word**	**Marthi Pronunciation**
1	Aadarniy	आदरणीय
2	Aairani	ऐरणी
3	Amrut	अमृत
4	Antara	अंतरा
5	Aushadh	औषध

The detail digitization of recorded vocabulary database is given below. The graphical representation of Aadarniy, Aairani, and Eshanya are described in figure2 to figure 4 respectively.

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 8, August 2015

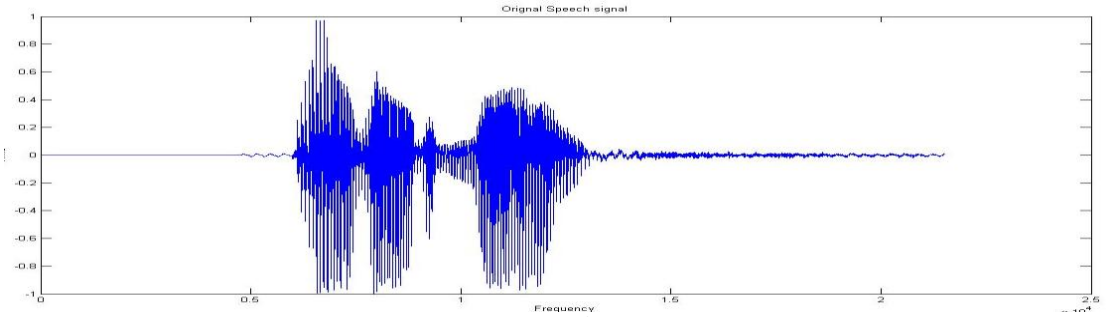

Fig.2.Graphical representation of vocabulary word 'Aadarniy'

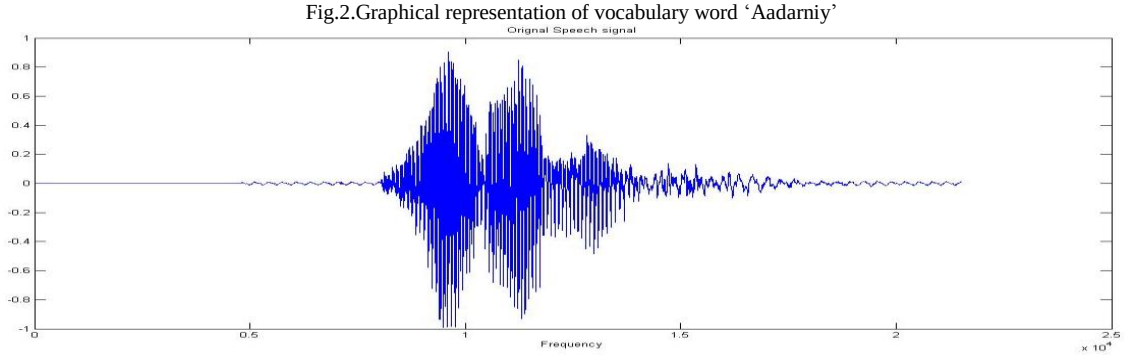

Fig.3.Graphical representation of vocabulary word 'Airani'

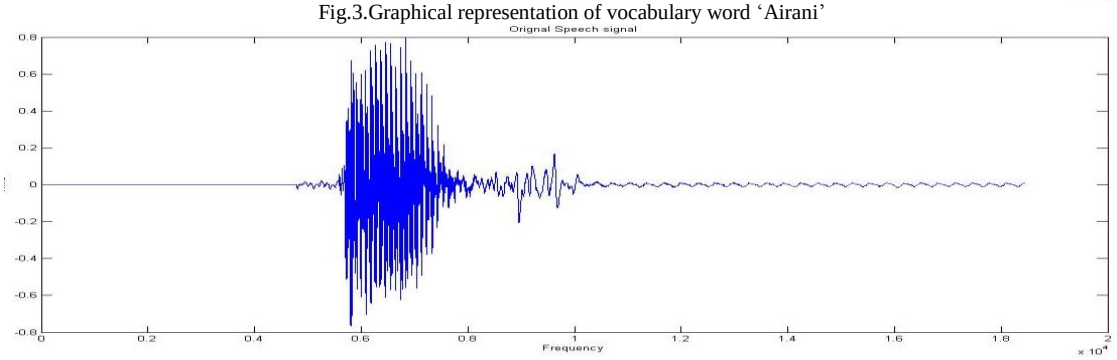

Fig. 4. Graphical representation of vocabulary word 'Eshanya'

The above three figurerepresented by the plot of frequency verses time of speech acquired signal by using spectrographic analysis.

B. Feature Extraction:

The focus of the proposed study is development of Standard speech database and using that developed database for development of Speaker Recognition System. For developing Speaker Recognition system we need to extract the feature from the acquired/recorded speech and then apply the recognition algorithm. However, initially we need to enhance the acquired/recorded speech signal sometime before we can extract the feature as it may contain noise [9]. There are different techniques that are used for the speech signal enhancement and speech feature extraction. Speech signal is analog. In the first place analog electrical signals are converted to digital signals. This is done in two steps, sampling and quantization [10]. So a typical representation of a speech signal is a stream of 8-bit numbers at the rate of 10,000 numbers per second. Once the signal conversion is complete, background noise is filtered to keep signal to noise ratio high. The signal is pre-emphasized and then speech parameters are extracted[11].

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 8, August 2015

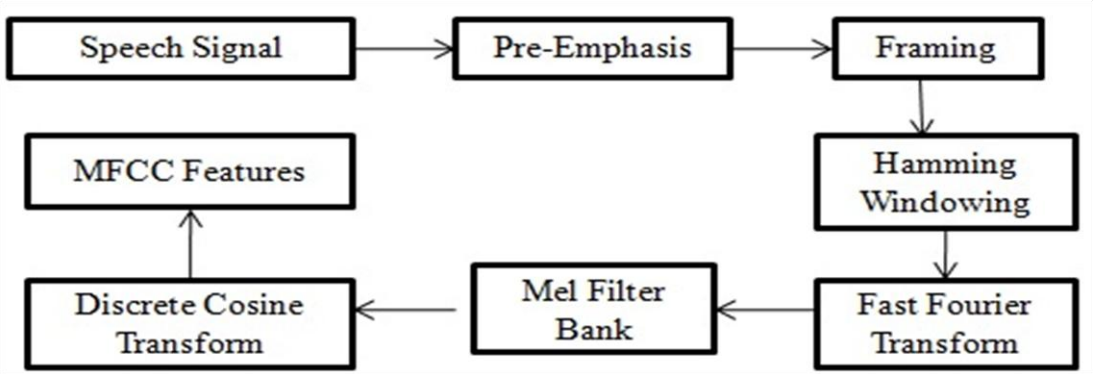

Fig.5. Block Diagram of MFCC Feature Extraction Techniques

We used Mel- Frequency Cepstral Coefficient (MFCC) for extracting features from recorded speech sample. The Mel-frequency Cepstrum Coefficient (MFCC) technique is often used to create the fingerprint of the sound files. The MFCC are based on the known variation of the human ear's critical bandwidth frequencies with filters spaced linearly at low frequencies and logarithmically at high frequencies used to capture the important characteristics of speech [12]. The signal is divided into overlapping frames to compute MFCC coefficients. Let each frame consist of N samples and let adjacent frames be separated by M samples where M<N. Each frame is multiplied by a Hamming window.

A block diagram of the MFCC processes is shown in figure5. Block diagram of MFCC The speech waveform is cropped to remove silence or acoustical interference that may be present in the beginning or end of the sound file. The windowing block minimizes the discontinuities of the signal by tapering the beginning and end of each frame to zero. The FFT block converts each frame from the time domain to the frequency domain. In the Mel-frequency wrapping block, the signal is plotted against the Mel spectrum to mimic human hearing. In the final step, the Cepstrum, the Mel-spectrum scale is converted back to standard frequency scale. This spectrum provides a good representation of the spectral properties of the signal which is key for representing and recognizing characteristics of the speaker.

Table 4. Technical parameter for MFCC feature extraction

Parameter	**Values**
Sampling Frequency	16kHz
Window Type	Hamming
Number of Coefficients (in each frame)	12
Filters in filter bank	29
Length of the frame	256
Frame increment	128

Results of the speaker recognition performance by the number of filters of MFCC to 12, are given. The recognizer reaches the maximal performance at the filter number K = 12. Too few or two many filters do not result in better accuracy. Hereafter, if not specifically stated, the number of filters is chosen to be K = 12. The step by step implementation of MFCC is described in figure. The detail parameter of MFCC feature extraction is described in table 4.

IV. CLASSIFICATIONS & RESULTS

The classification we used the Euclidian distance for measure of recognition performance. The extracted MFCC feature is trained and tested using Euclidian distance measure. Very often, especially when measuring the distance in the plane, we use the formula for the Euclidean distance. According to the Euclidean distance formula, the distance between two points in the plane with coordinates (x, y) and (a, b) is given by

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 8, August 2015

$$dist((x, y), (a, b)) = \sqrt{(x - a)^2 + (y - b)^2}$$

For the testing of the system we created five module, the module first is ten Male subject, group two is 05 female subject and group 3 is random selected subject. The Table 5 is describing the performance of the speaker verification system for Male selected 05 speaker group. The Female speaker group performance is described in Table 6.Random selected group performance in table 7.

Table 5. The performance of Speaker recognition system

Speaker	Number of Attempt	False Acceptance	False Rejection	Accuracy
S1	10	2	1	80
S2	10	1	3	90
S3	10	2	1	80
S4	10	1	2	90
S5	10	2	4	80
Overall Accuracy				84%

Table 6. The performance of Speaker recognition system for Female subject group

Speaker	Number of Attempt	False Acceptance	False Rejection	Accuracy
S1	10	2	1	80
S2	10	1	2	90
S3	10	3	1	80
S4	10	2	1	80
S5	10	2	1	80
Overall Accuracy				82%

Table 7. The performance of Speaker recognition system for random selected subject group

Speaker	Number of Attempt	False Acceptance	False Rejection	Accuracy
S1	10	1	2	80
S2	10	1	2	80
S3	10	3	4	70
S4	10	2	2	80
S5	10	2	1	80
Overall Accuracy				78%

$$FAR = \frac{\#\text{ accepted imposter claims}}{\#\text{imposter accesses}} X100\%$$

$$FRR = \frac{\#\text{ rejected genuine claims}}{\#\text{genuine accesses}} X100\%$$

In this paper we are implementing the database development for Marathi speaker recognition. The feature is extracted from MFCC techniques. As gender varies the performance of the system. Using this factor we divided the test dataset is Male group, female subject group, and random selected subject group. The performance of system is calculated using FAR and FRR. The accuracy of the Male subject group is more than the accuracy of female subject group. The overall performance of the random selected subject group was 81.33%.

V. CONCLUSION

The speaker recognition is having number of real time application. The challenging task is to improve the performance of system. Current era of speaker recognition is focused towards improving the performance of the system. The original objective of this research was to build a Marathi speaker recognition system. The system is tested with the help of Mel Frequency Cepstral coefficient and classification is done with Euclidian Distance. The gender and age group of the speaker play an important role in speaker recognition. The performance of the system is tested on the basis of different three groups of speaker as Male speaker, Female speaker and randomly selected speaker. The performance of system is calculated using FAR and FRR. The accuracy of the Male subject group is more than the accuracy of female subject group. The overall performance of the selected subject group is 81.33%.

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 8, August 2015

REFERENCES

1. S. Furui, “Speaker independent isolated word recognition using dynamic features of speech spectrum”, IEEE Transactions on Acoustic, Speech, Signal Processing, Vol. ASSP-34, No. 1, pp. 52-59,1986.
2. Chao Huang, Eric Chang, Tao Chen “Accent Issues in Large Vocabulary Continuous Speech Recognition (LVCSR)”, Microsoft Research China, MSR-TR-2001-69, pp.1-27,2001.
3. http://en.wikipedia.org/wiki/List_of_languages_by_number_of_native_speakers.html,Viewed 1st June, 2014.
4. Vaibhav V.Kamble, Bharatratna P.Gaikwad, Deepak M.Rana.: Spontaneous Emotion Recognition for Marathi Spoken Words, In: International Conference on Communication and Signal Processing, 978-1-4799-3356-3/14,pp.1884-1889, IEEE Xplore,2014.
5. Urmila Shrawankar, and Dr. Vilas Thakare, "Techniques for Feature Extraction in Speech Recognition System: A Comparative Study", International Journal of Computer Applications in Engineering, Technology and Sciences (IJCAETS), ISSN 0974-3596, pp 412-418, 2010.
6. S. Furui, “An overview of speaker recognition technology”, ESCA Workshop on Automatic Speaker Recognition, Identification and Verification, pp. 1-9, 1994.
7. M. A. Anusuya, S. K. Katti, “Speech Recognition by Machine: A Review”, International Journal of Computer Science and Information Security (IJCSIS), Vol. 6, No. 3, pp. 181-205, 2009.
8. X. D. Huang, "A Study on Speaker - Adaptive Speech Recognition", Proc. DARPA Workshop on Speech and Natural Language, pp. 278-283, 1991.
9. Lawrence R. Rabiner and Ronald W. Schafer, “Digital Processing of Speech Signals, Signal Processing”, Prentice-Hall, Englewood Cliffs, New Jersey, 1978.
10. Bhupinder Singh, Rupinder Kaur, Nidhi Devgun, Ramandeep Kaur, "The process of feature extraction in automatic speech recognition system for computer Machine interaction with humans: A Review", International Journal of Advanced Research in Computer Science and Software Engineering, Vol. 2, No. 2,2012.
11. Vibha Tiwari, "MFCC and its application in speaker recognition", International Journal on Emerging Technologies, Vol. 1, No. 1, pp. 19-22 ,2010.
12. S. E. Levinson, “Structural methods in automatic speech recognition,” Proc. IEEE,vol. 73, no. 11, pp. 1625-1650, 1985

BIOGRAPHY

Dr. Bharatratna Pralhadrao Gaikwad

Completed M.Sc., NET, Ph. D. (Computer Science) in University Department of Computer Science and Information Technology, Dr. Babasaheb Ambedkar Marathwada University Aurangabad under Guidance of respected Dr.R.R.Manza. Paper published in IEEE Xplore, Springer, IJCA, CIIT and Area of specialization in Network Security, Video Processing, Cyber Law, E- Commerce, and Pattern Recognition. Member of IEEE, Life Member of ISCA, ICSA, IAENG, IACSIT