International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 06 Issue: 01 | Jan - 2022 ISSN: 2582-3930

CFCC Method for Robust Speaker Identification

Prof.Sonali S.Kumbhar [1], Prof. Parija S.Shaikh [2], Prof.Chaitanya A.Kulkarni [3]

[1]Electronics & Telecommunication Department Nanasaheb Mahadik Poytechnic Institute , Peth

[2]Electronics & Telecommunication Department Nanasaheb Mahadik Poytechnic Institute Peth

[3]Electronics & Telecommunication Department Nanasaheb Mahadik Poytechnic Institute Peth

---***---

1. INTRODUCTION

Spoken language is the most natural way used by humans to communicate information. The speech signal conveys several types of information. In the speech production point of view, the speech signal conveys linguistic information (e.g., message and language) and speaker information (e.g., emotional, regional, and physiological characteristics). In the speech perception point of view, it also conveys information about the environment in which the speech was produced and transmitted. Also there is another method of perception is human hearing system, which gives the information how the speech signal receives into the human ear..Even though this wide range of information is encoded in a complex form into the speech signal, humans can easily decode most of the information. Such human ability has is used to understand speech production and perception for developing systems that automatically extract and process the richness of information in speech. This paper is presenting automatic systems that recognize who is speaking (i.e. speaker identification) using Cochlear Filter Cepstral Coefficients(CFCC).

Key Words: linguistic information, speaker information, CFCC.

1.Basic Structure of a Speaker Recognition System

The main aim of this project is speaker identification, which consists of comparing a speech signal from an unknown speaker to a database of known speaker. The system can recognize the speaker, which has been trained with a number of speakers.

1.1 Acoustic Features

All audio processing techniques start by converting the raw speech signal into a sequence of acoustic feature vectors carrying characteristic information about the signal.[9] This preprocessing module (feature extraction) is also referred to as "front-end" model. The most commonly used acoustic vectors are Cochlear Filter Cepstral Coefficients(CFCC),Mel Frequency Cepstral Coefficients (MFCC), Linear Prediction Cepstral Coefficients (LPCC) and Perceptual Linear Prediction Cepstral (PLPC) Coefficients [1].

In this project the most important thing is to extract the feature from the speech signal. As from the fundamental formation of speaker identification and verification systems, that the number of training and test vector needed for the classification problem grows exponential with the dimension of the given input vector, it needs feature extraction.[4]

But extracted feature should meet some criteria while dealing with the speech signal. Such as:

- Easy to measure extracted Speech features.
- Distinguish between speakers while being lenient of intra speaker variability's.
- It should not be susceptible to mimicry.
- It should show little fluctuation from one speaking environment to another.
- It should occur frequently and naturally in speech.[2]

In this project, using the Cochlear Filter Cepstral Coefficients (CFCC) technique is used to extract features from the speech signal and compare the unknown speaker with the exits speaker in the database[1].

1.2 Speaker Modeling Gaussian Mixture Model (GMM)

A GMM is a mixture of several Gaussian distributions and is used to estimate the

Probability Density Function (PDF)of a sequence of feature vectors.The GMM has several properties that motivate their use for representing a speaker:

• One of the powerful properties of the GMM is its ability to form smooth approximations to arbitrarily shaped density. The GMM can be viewed as a parametric PDF based on a linear combination of Gaussian basis functions capable of representing a large class of arbitrary densities.

• GMM can be considered as an implicit realization of probabilistic modeling of speaker dependent acoustic classes with each Gaussian component corresponding to a broad acoustic class such as vowels, nasals and fricatives etc. [7]

1.3 Training Mode:

In the first phase, a training phase,is for the statistical GMM models based on some training material. It includes speech samples in very clean environment.[2]

1.4 Testing Mode:

In the second phase, a test phase, the identification accuracy of the learned models is evaluated using data that was not included in the model training. It includes speech samples under mismatched conditions like babble noise and white noise.

Both the training and testing phases work on spectral feature vectors extracted frame-wise from speech waveforms.[1]

1.5 Speaker Model Database:

This system uses database which is taken under clean environment i.e. called training mode and other under mismatched conditions like white noise and babble noise with -6db,0db , 6db and also at clean environment conditions. Total 250 voice samples are taken with the help of 25 speakers.

In first database, each speaker speaks two same sentences for 2 times and in second 10 different sentences are taken by different 25 speakers. While in third database 10 voice samples of Amitabh Bachchan in original voice and his mimicrians 10 voice samples are taken.

Speaker identification system is the process of selecting the best matched speaker among the enrolled speakers, with features extracted from speech signals.[3]

Many techniques involving statistical or probabilistic approaches have been applied to speaker specific speech patterns (Leena Mary and

Yegnanarayana (2008), Jyoti et al (2011)) . Several methods were employed to separate mixed signals known as 'Blind Source Signals' (BSS). The term blind refers to the fact that the method of combination and source signal characteristics are unknown, so BSS permits a wide range of signals as input.[3]

Text independent speaker identification system has many potential applications like security control, telephone banking, information retrieval systems, speech and gender recognition systems, etc. Speaker identification system involves two parts: front-end (feature extractions) and back-end (actual recognition). These system use processed form of speech signals instead of using raw speech signals as it is obtained. This is to reduce the time consumed in identifying the speaker and to make the process easy, by reducing the data stream and exploiting its advantage of being redundant. Computation of cepstral coefficients using preprocessing and feature extraction phases plays a major role in text independent speaker identification systems Ning Wang et al (2010) [6].

2. Proposed Experimental Work:

The proposed algorithm replicates the hearing system at a high level and consists of the following modules: auditory transform implemented by a cochlear filter bank, hair-cell function with windowing, cubic-root nonlinearity, and discrete cosine transform(DCT).

CFCC features are extracted from development dataset. In this experiment speaker models were first trained using the clean training set and then tested on noisy speech at four SNR levels. It is created using three disjoint subsets from the database as the training set, development set, and testing set. Fig shows AT based CFCC algorithm.

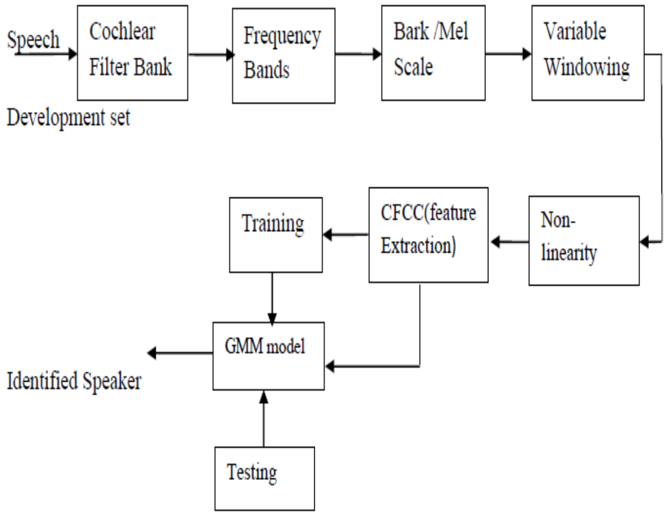


Fig Block Diagram of Proposed Work

2.1 Feature Extraction Using CFCC

The structure of the proposed auditory-based feature extraction algorithm and provides details of its computation. To emulate the human peripheral hearing system, the computational aspects must meet the requirements of real-time applications; therefore, we will simulate only the most important features of the human peripheral hearing system. An illustrative block diagram of the proposed algorithm is shown in Fig. The proposed algorithm is intended to conceptually replicate the hearing system at a high level and consists of the following modules: auditory transform implemented by a cochlear filter bank, hair-cell function with windowing, cubic-root nonlinearity, and discrete cosine transform (DCT). A detailed description of each module follows.

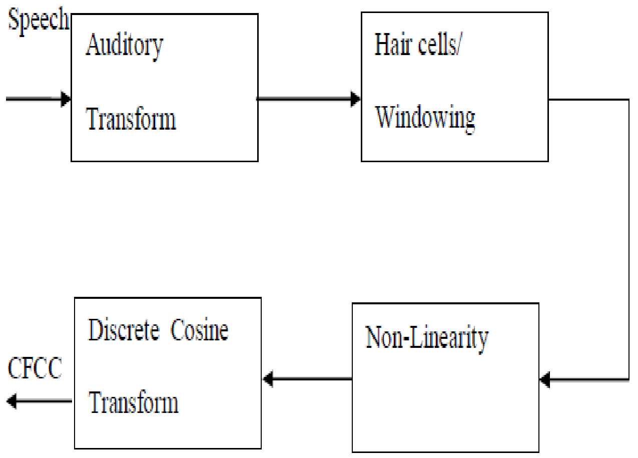


Fig: Schematic diagram of proposed Auditory based CFCC algorithm

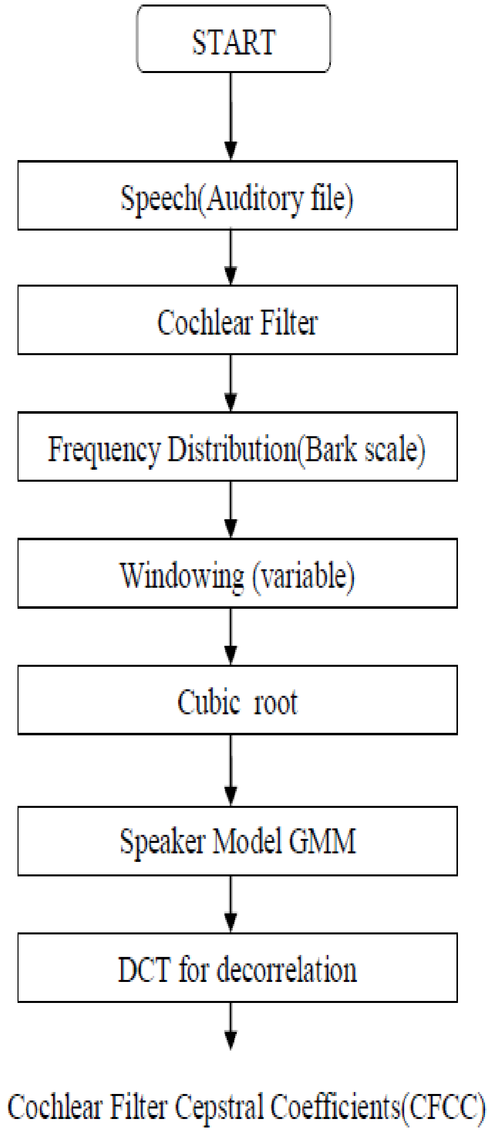


Fig. Flowchart for AT based CFCC algorihm

The aim of studying human hearing system is to simulate the processes present within the human auditory system, rather than attempting to simulate its effects. These processes are simulated on a large scale rather than on the neural level. By doing this, the major effects of the auditory system arise inherently while the unwieldy computational requirements of a neural level model are avoided.

Human auditory system includes important parts like ear drum, auditory nerve, middle ear bone, basilar membrane and cochlea in the ear. Fig. is a cross section illustrating the main components of the human ear. The following, adapted from [8], briefly describes the function of each section of the ear. Incoming sounds are funneled into the ear canal by the *pinna*, the only external part of the ear. It directionally filters sound, enabling humans to localize sound in 3- dimensions. The *ear canal* filters the sound, attenuating low and high frequencies, giving resonance at about 5kHz.

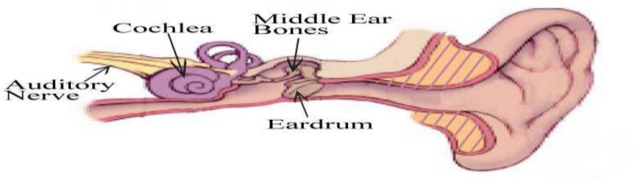


Fig.4.6 Human Auditory System

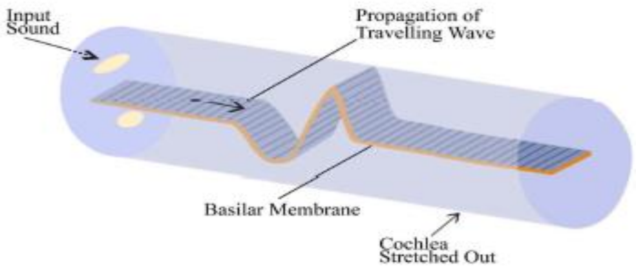


Fig The Cochlea and travelling wave

- The fluid-filled *cochlea* is a coil within the ear, partially protected by bone. It contains the *basilar membrane*, and *hair cells*, responsible for the transduction of the sound pressure wave into neural signals.
- The *basilar membrane* semi-partitions the *cochlea*, and acts as a spectrum analyzer, spatially decomposing the signal into frequency components. Each point on the basilar membrane resonates at a different frequency, and the spacing of the resonant frequencies along the basilar membrane, as seen in fig. , is nearly logarithmic.
- The *outer hair cells* are distributed along the length of the basilar membrane. They react to feedback from the brainstem. This causes the frequency response of the basilar membrane to be amplitude dependent.
- The *inner hair cells* fire when the basilar membrane moves upwards, so transducing the sound wave at each point into a signal on the auditory nerve. In this way, the signal is effectively half wave rectified. Thus, the inner hair

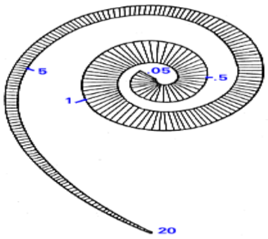


Figure 4.8 Distribution of frequencies along the basilar membrane (kHz).

3.Software development using Matlab

The main notion behind these techniques is to extract relevant features and characteristics of a speech signal. MATLAB is chosen for this programming environment because it offers several advantages.

3.1 Database used for this Project

The set use in this work is database created by 25 speakers under clean conditions each talks same sentences.

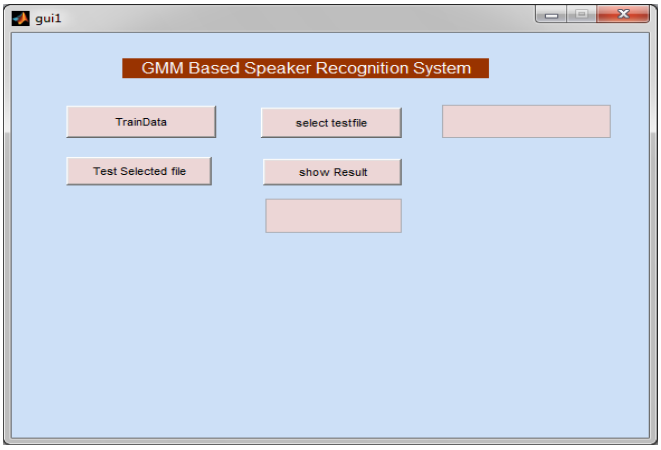
The second data set is database created by 25 speakers under 3 different SNR levels like -6dB,0dB and 6dB using babble noise and each speaker talks two different sentences 5 times each.

The next set in this work is database created by 25 speakers under clean conditions each talks different sentences. The second data set is database created by 25 speakers under 3 different SNR levels like -6dB,0dB and 6dB using babble noise and each speaker talks with different sentences.

Also the third database created for 1 speakers with his mimicrian and his original voice 10 voice samples under clean conditions.

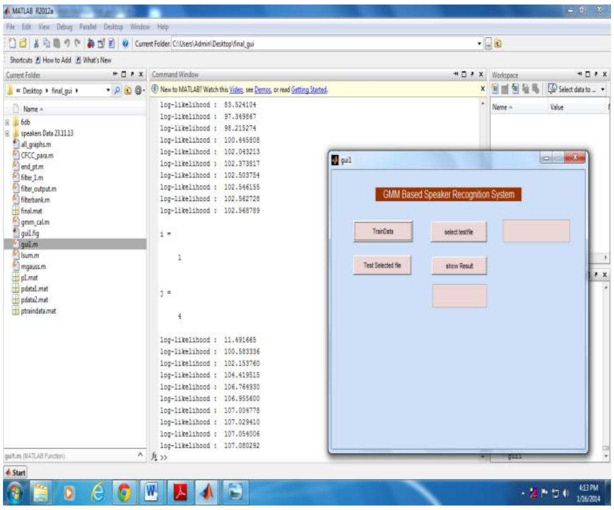
3.3 GUI for Robust Speaker Identification using CFCC

Fig.GUI for Robust Speaker Identification using CFCC



3.4 GMM values for Training Data:

When clicked on Train data it calculates log likelihood values of speakers data.



```
log-likelihood :  83.524104
log-likelihood :  97.349867
log-likelihood :  98.215274
log-likelihood :  100.445808
log-likelihood :  102.043213
log-likelihood :  102.373817
log-likelihood :  102.503754
log-likelihood :  102.546155
log-likelihood :  102.562728
log-likelihood :  102.568789

i =

     1


j =

     4

log-likelihood :  11.491665
log-likelihood :  100.583336
log-likelihood :  102.153760
log-likelihood :  104.419515
log-likelihood :  106.764930
log-likelihood :  106.955600
log-likelihood :  107.004778
log-likelihood :  107.029410
log-likelihood :  107.054006
log-likelihood :  107.080292
>>
```

Fig Calculation of GMM values for training data

3.5 Result when the mimicrian's voice compared with original voice (under noisy conditions)

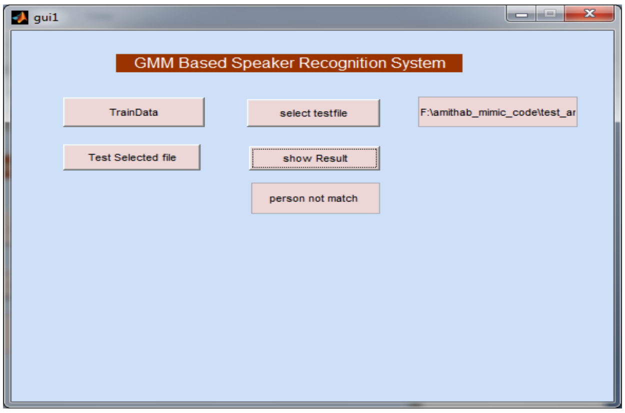
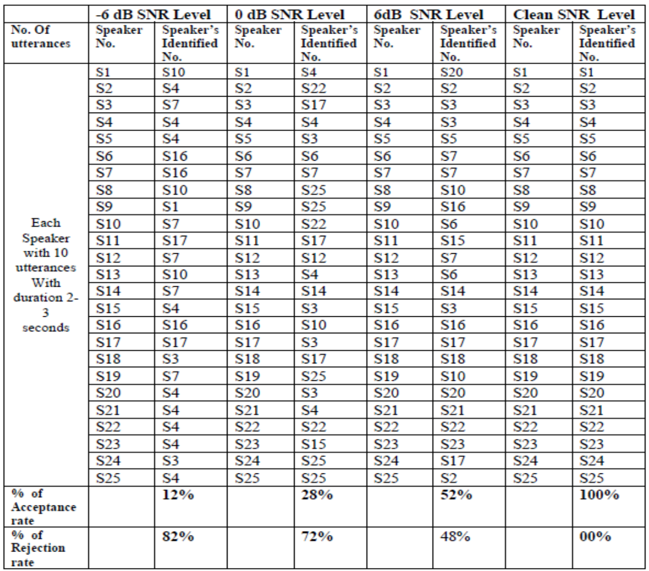


Fig Result when the mimicrian's voice compared with original voice(under noisy conditions)

3.6 For the different sentences CFCC concludes that: (For Babble Noise)

From the table it can conclude that Speaker Identification results for epoch window gives 8%, 20%,28%,88% acceptance rates for -6dB,0dB,6dB SNR levels and Clean speech respectively.

3.7 Speaker Identification Accuracy Results of MFCC for Babble Noise

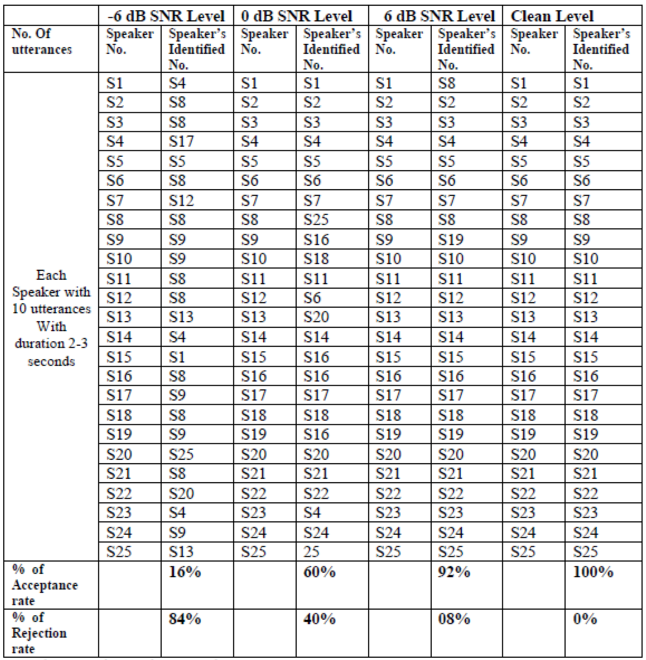


	-6 dB SNR Level		0 dB SNR Level		6dB SNR Level		Clean SNR Level	
No. Of utterances	Speaker No.	Speaker's Identified No.	Speaker No.	Speaker's Identified No.	Speaker No.	Speaker's Identified No.	Speaker No.	Speaker's Identified No.
Each Speaker with 10 utterances With duration 2-3 seconds	S1	S10	S1	S4	S1	S20	S1	S1
	S2	S4	S2	S22	S2	S2	S2	S2
	S3	S7	S3	S17	S3	S3	S3	S3
	S4	S4	S4	S4	S4	S3	S4	S4
	S5	S4	S5	S3	S5	S5	S5	S5
	S6	S16	S6	S6	S6	S7	S6	S6
	S7	S16	S7	S7	S7	S7	S7	S7
	S8	S10	S8	S25	S8	S10	S8	S8
	S9	S1	S9	S25	S9	S16	S9	S9
	S10	S7	S10	S22	S10	S6	S10	S10
	S11	S17	S11	S17	S11	S15	S11	S11
	S12	S7	S12	S12	S12	S7	S12	S12
	S13	S10	S13	S4	S13	S6	S13	S13
	S14	S7	S14	S14	S14	S14	S14	S14
	S15	S4	S15	S3	S15	S3	S15	S15
	S16	S16	S16	S10	S16	S16	S16	S16
	S17	S17	S17	S3	S17	S17	S17	S17
	S18	S3	S18	S17	S18	S18	S18	S18
	S19	S7	S19	S25	S19	S10	S19	S19
	S20	S4	S20	S3	S20	S20	S20	S20
	S21	S4	S21	S4	S21	S21	S21	S21
	S22	S4	S22	S22	S22	S22	S22	S22
	S23	S4	S23	S15	S23	S23	S23	S23
	S24	S3	S24	S25	S24	S17	S24	S24
	S25	S4	S25	S25	S25	S2	S25	S25
% of Acceptance rate		12%		28%		52%		100%
% of Rejection rate		82%		72%		48%		00%

Note: Where "S" is Speakers Number

Table:1 Speaker Identification Accuracy Results of MFCC for Babble Noise

3.8 Speaker Identification Accuracy Results of Equal Loudness for 25 Speakers with CFCC for Babble Noise

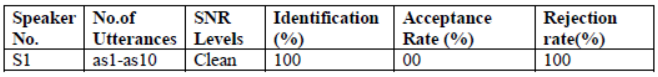


	-6 dB SNR Level		0 dB SNR Level		6 dB SNR Level		Clean Level	
No. Of utterances	Speaker No.	Speaker's Identified No.	Speaker No.	Speaker's Identified No.	Speaker No.	Speaker's Identified No.	Speaker No.	Speaker's Identified No.
Each Speaker with 10 utterances With duration 2-3 seconds	S1	S4	S1	S1	S1	S8	S1	S1
	S2	S8	S2	S2	S2	S2	S2	S2
	S3	S8	S3	S3	S3	S3	S3	S3
	S4	S17	S4	S4	S4	S4	S4	S4
	S5	S5	S5	S5	S5	S5	S5	S5
	S6	S8	S6	S6	S6	S6	S6	S6
	S7	S12	S7	S7	S7	S7	S7	S7
	S8	S8	S8	S25	S8	S8	S8	S8
	S9	S9	S9	S16	S9	S19	S9	S9
	S10	S9	S10	S18	S10	S10	S10	S10
	S11	S8	S11	S11	S11	S11	S11	S11
	S12	S8	S12	S6	S12	S12	S12	S12
	S13	S13	S13	S20	S13	S13	S13	S13
	S14	S4	S14	S14	S14	S14	S14	S14
	S15	S1	S15	S16	S15	S15	S15	S15
	S16	S8	S16	S16	S16	S16	S16	S16
	S17	S9	S17	S17	S17	S17	S17	S17
	S18	S8	S18	S18	S18	S18	S18	S18
	S19	S9	S19	S16	S19	S19	S19	S19
	S20	S25	S20	S20	S20	S20	S20	S20
	S21	S8	S21	S21	S21	S21	S21	S21
	S22	S20	S22	S22	S22	S22	S22	S22
	S23	S4	S23	S4	S23	S23	S23	S23
	S24	S9	S24	S24	S24	S24	S24	S24
	S25	S13	S25	25	S25	S25	S25	S25
% of Acceptance rate		16%		60%		92%		100%
% of Rejection rate		84%		40%		08%		0%

Note: Where "S" is Speakers Number

Table:2 Speaker Identification Accuracy Results of Equal Loudness for 25 Speakers with CFCC for Babble Noise

3.9 Results of Speaker Identification System for Mimicrian



Speaker No.	No.of Utterances	SNR Levels	Identification (%)	Acceptance Rate (%)	Rejection rate(%)
S1	as1-as10	Clean	100	00	100

Table3: Acceptance and Rejection rate of Speaker Identification Accuracy(Clean Conditions)

CONCLUSIONS

i) This project presented a new auditory-based algorithm for speech feature extraction and applied it to robust speaker identification under mismatched conditions.

ii)The algorithm was developed based on a recently presented invertible auditory transform plus several components motivated by the human periphery hearing system.

iii)Table 1 & 2 shows that CFCC gives better performance as compared to MFCC.

iv)Table3 shows that when original voice compared to mimicrians voice it does not fund the match.

4. APPLICATIONS

Research in the area of speaker recognition has significantly grown over the last few years due to a vast area of applications where the recognition can be used such as

- Access control: to devices, networks, and data in general;
- Authentication for business transactions as a tool to prevent fraud in: shopping over telephone, credit card validation, transactions over Internet, bank operations, etc.
- Law enforcement: penitentiary monitoring, forensic applications, etc.
- Military use: classified information requiring speaker identification.

ACKNOWLEDGEMENT

We gratefully acknowledge the help & cooperation offered By Professor M.B.Joshi, Principal of our college and Management of NMPI,Peth for providing help and support to carry out the project.

REFERENCES

1] Q Li and Yan Huang "An auditory-based feature extraction algorithm for robust speaker identification under mismatched conditions"- *Senior Member, IEEE Transactions on Audio,Speech and language processing,Vol.19,No.6,August2011*

[2]J. P. Campbell. Speaker Recognition: A Tutorial. *Proceeding of the IEEE*, 85:1437-1462, September 1997.

[3]N.M.Ramaligeswararao, Dr. V Sailaja Department of Electronics & Communication Engineering, GIET affiliated to JNTUK, Rajahmundry, India and Dr.K. Srinivasa Rao Department of statistics, Andhra University Visakhapatnam, India 'Text Independent Speaker Identification using Integrated Independent Component Analysis with Generalized Gaussian Mixture Model',*(IJACSA) International Journal of Advanced Computer Science and Applications, Vol. 2, No.12,2011.*

[4] "MFCC and its applications in speaker recognition" *Vibha Tiwari Deptt. of Electronics Engg., Gyan Ganga Institute of Technology and Management, Bhopal, (MP) INDIA (Received 5 Nov., 2009, Accepted 10 Feb., 2010)*,M.Tech.

Credits seminar report,Electronic Systems Group,EE Dept,IIT Bombay submitted in Nov.03

[5]Ning Wang,,P. C. Ching,,Nengheng Zheng, and Tan Lee, *Member*, *IEEE* 'Robust Speaker Recognition Using Denoised Vocal Source and Vocal Tract Features', IEEE *Transactions On Audio*, *Speech*, *And Language Processing*,Vol.19,No.1 January2011.

[7] D. A. Reynolds. An Overview of Automatic Speaker Recognition Technology. In *Proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing*, Orlando, USA, 2002.

[8]Alex Park and Timothy J. Hazen "ASR Dependent Techniques For Speaker Identification" Language Systems Group MIT Laboratory for Computer Science Cambridge, Massachusetts 02139 USA.. *Proceedings of the 7th International Conference on Spoken Language Processing*, Sep. 16-20, 2002, Denver, Colorado, pp. 1337-1340.