

## Robust Speaker Identification using CFCC

<sup>1</sup>Shri.R.T.Patil , <sup>2</sup>Ms.Sonali Sambhaji Kumbhar

<sup>1</sup>Professor,T.K.I.E.T.Waranagar Dept.Of Electronics Engg. Maharashtra, India

<sup>2</sup>Lecturer,Adarsh Institute of Technology & Research Centre,Vita  
Dept.Of E&TC Engg.Maharashtra,india

---

**ABSTRACT:** In Auditory Based features for Robust Speaker Identification, an auditory-based(AT) feature extraction algorithm is used. It is modeled on the basic signal processing functions in the ear. The proposed algorithm is also based on our recently developed auditory-based time-frequency transform named Auditory Transform. The features generated from the proposed algorithm are named cochlear filter cepstral coefficients (CFCC).The new features of a CFCC based on a recently developed auditory transform (AT) plus a set of modules to present the signal processing functions in the cochlea. The CFCC features are applied to a speaker identification task to address the acoustic mismatch problem between training and testing environments.[1]

**INDEX TERMS:** AT feature extraction, CFCC, cochlear filter, speaker identification

---

### I. INTRODUCTION

The CFCC features consistently perform better than the baseline MFCC features[1] under all three mismatched testing conditions –white noise, car noise, and babble noise. In clean conditions, both MFCC and CFCC features perform similarly, over 96%, but when the SNR of the input signal is 6 dB, the accuracy of the MFCC features drops to 41.2%, while the CFCC features still achieve an accuracy of 88.3%[1].

### II. SPEAKER RECOGNITION SYSTEM STAGES

Speaker recognition system may be viewed as working in four stages- analysis, feature extraction, modeling and testing.[7]

#### A. Feature Extraction Techniques:

**i)Spectral Features-**It is like band energies, formants, spectrum and cepstral coefficients representing mainly the speaker-specific information due to the vocal tract. The main reasons for the same may be the less intra-speaker variability and also availability of MFCCs and LPCCs rich spectral analysis tools. However, the speaker-specific information due to excitation source and behavioral trait represents different aspects of speaker information.

**ii) Excitation source features-**It is like pitch, variations in pitch, information from LP residual and glottal source parameters. PLP and RASTA-PLP are excitation source methods. However, the speaker-specific information due to excitation source and behavioral trait represents different aspects of speaker information.

**iii)Long-term features-** It is like duration, intonation, energy, AM and FM components representing mainly the speaker-specific information due to the behavioral traits.

#### B. The modeling techniques & Speaker recognition systems:

The objective of modeling technique is to generate speaker models using speaker-specific feature vectors. In case of text-independent speaker recognition, Vector Quantization (VQ) and its variants like fuzzy vector quantization (FVQ), self-organization map (SOM) and learning vector quantization (LVQ) Among these, from the simplicity point of view, VQ is the mostly used one; and from the performance point of view, LVQ is the preferred one[7]. HMM for text-independent speaker recognition under the constraint of limited data and mismatched channel conditions. In MFCC models were built using the broad phonetic category (BPC) and the HMM-based maximum likelihood linear regression (MLLR) adaptation technique. The BPC modeling is based on identification of phonetic categories in an utterance and modeling them separately.The Gaussian mixtures models (GMMs) technique is the mostly used modeling technique from among the Gaussian classifiers[5]. In neural networks, the ones for speaker modeling are the MLP, RBF and auto-associative neural network AANN models.

Most recently, SVM has also been demonstrated to be a potential discriminatory-type classifier for speaker modeling, especially under conditions of limited data. As a final comment, it should be stated that the GMM-SVM (support vector machine). combination has been demonstrated to provide better modeling compared to either GMM or SVM alone[6].

Two types of speaker identification systems:

- a) State-of-the-art speaker recognition systems
- b) Direct template matching between training and testing data

It is recently used method. During testing, the speech signal is analyzed and features are extracted using the same techniques employed during training. The feature vectors are compared with the reference models using some distance measure like Euclidean distance; and based on the comparison result, the speaker in the test speech data will be recognized[4].

### III. PROPOSED CFCC ALGORITHM

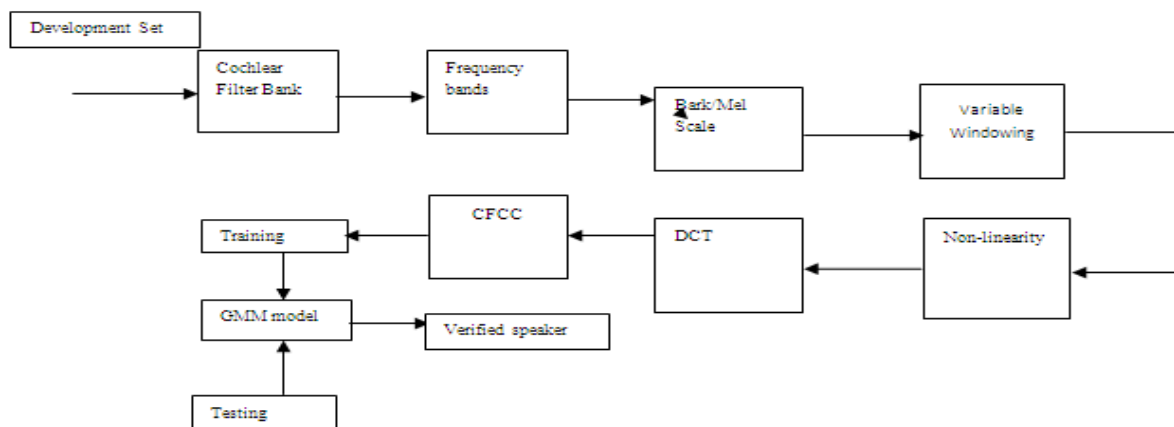
The block diagram of the proposed algorithm is shown in “Fig. 1”. The proposed algorithm replicates the hearing system at a high level and consists of the following modules: auditory transform implemented by a cochlear filter bank, hair-cell function with windowing, cubic-root nonlinearity, and discrete cosine transform(DCT)[2].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. Fig1 shows AT based CFCC algorithm[1][4].

**The CFCC algorithm can be explained as follows:**

- i) the speech audio file is passed through the band-pass filter bank. The filter width parameter  $\beta$  was set to 0.035. The Bark scale is used for the filter bank distribution and equal-loudness weighting is applied at different frequency bands.
- ii) the travelling waves generated from the cochlear filters are windowed and averaged by the hair cell function.
- iii) A cubic root is applied.
- iv) Back-end systems adopt diagonal covariance based GMM or HMM models, the discrete cosine transform (DCT) is used to decorrelate the features. The 0th component, related to the energy, is removed from DCT output. “Fig 1” shows the speaker identification algorithm using CFCC.

### IV. FEATURE EXTRACTION USING CFCC

The goal of the development set is to determine the effects of each component on the overall performance and ultimately optimize the feature extraction .Following features are extracted. Fig 2 shows feature extraction algorithm.



▲ Fig1. Schematic block diagram of the proposed algorithm

- [1] Effect of Filter Bandwidth ( $\beta$ ): Speaker identification accuracies of the auditory-based cochlear
- [2] features (CFCC) with different filter bandwidth adjusted by parameter ( $\beta$ ).
- [3] Effect of Equal-Loudness: Speaker identification accuracy results of the auditory-based cochlear features (CFCC) with/without equal loudness.
- [4] Effect of Various Windowing Schemes: Speaker identification accuracy results of the auditory-based cochlear features (CFCC) with a fixed-length window, fixed-epoch window, or a combination of the fixed-length and fixed-epoch window.
- [5] Effect of nonlinearity: Speaker identification accuracy results of the auditory-based cochlear features (CFCC) with logarithm and cubic nonlinearity.

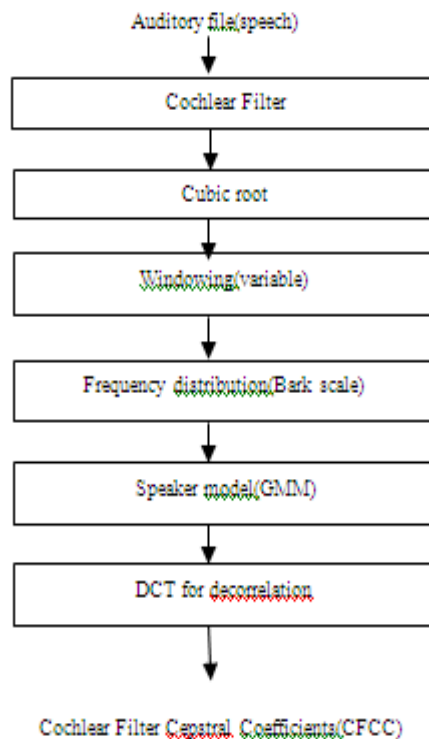
## V. APPLICATIONS

The speaker recognition has following applications.

- Telephone banking
- Voice mailing
- Accent recognition
- Forensic purpose

## VI. CONCLUSION

Thus CFCC algorithm is used for the speech feature extraction as well as for speaker identification.



**Fig2. Flowchart for AT based CFCC algorithm for feature Extraction**

## 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, August 2011
- [2] Q. Li, "Solution for pervasive speaker recognition," SBIR Phase I Proposal, Submitted to NSF IT.F4, Li Creative Technologies, Inc., NJ, June 2003.
- [3] Q. Li, "An auditory-based transform for audio signal processing," in Proceedings of IEEE Workshop on Applications of Signal Processing to Audio and Acoustics, (New Paltz, NY), Oct. 2009, pp 181-184
- [4] Y. Shao and D.Wang, "Robust speaker identification using auditory features and computational auditory scene analysis," in Proceedings of IEEE ICASSP, pp. 1589-1592, 2008.
- [5] D. Reynolds and R. C. Rose, "Robust text-independent speaker identification using Gaussian mixture speaker models," IEEE Trans. on Speech and Audio Processing, vol. 3, no. 1, pp. 72-83, 1995.

- [6] Marc Ferras, Cheung-Chi Leung, Claude Barras, and Jean-Luc Gauvain, "Comparison of speaker adaptation methods as feature extraction for svm-based speaker recognition" Member, IEEE Transactions On Audio, Speech, And Language Processing, Vol. 18, No. 6, August 2010, pp 1558-7916
- [7] Prof. Rupali Pawar and Miss. Hemangi Kulkarni "Analysis of FFSR, VFSR, MFSR techniques for feature extraction in speaker recognition: a review" -1. Department of Computer Engineering, Viswvakarma Institute of Technology, Pune University Pune, Maharashtra, 411037, India 2. Department of Computer Engineering, R.H. Raisonni College of Engineering and Management, Pune University Pune, Maharashtra, 411037, India IJCSI International Journal of Computer Science Issues, Vol. 7, Issue 4, No 1, July 2010
- [8] Qi (Peter) Li and Yan Huang Li Creative Technologies (LcT), Inc. "Robust speaker identification using an auditory-based feature" " 25 B Hanover Road, Suite 140, Florham Park, NJ 07932 USA, Biometric Consortium Conference, 9:40 AM, September 21, 2010