SPEAKER IDENTIFICATION

Ms. Arundhati S. Mehendale and Mrs. M. R. Dixit

Department of Electronics K.I.T.'s College of Engineering, Kolhapur

ABSTRACT

Speaker recognition is the computing task of validating a user's claimed identity using characteristics extracted from their voices. Voice -recognition is combination of the two where it uses learned aspects of a speaker's voice to determine what is being said - such a system cannot recognize speech from random speakers very accurately, but it can reach high accuracy for individual voices it has been trained with, which gives us various applications in day today life.

KEYWORDS

Speech-recognition, speaker recognition.

1. INTRODUCTION

The task of speaker identification is to determine the identity of a speaker by machine. To recognize voice, the voices must be familiar in case of human beings as well as machines. The second component of speaker identification is testing; namely the task of comparing an unidentified utterance to the training data and making the identification. The speaker of a test utterance is referred to as the target speaker. Recently, there has been some interest in alternative speech parameterizations based on using formant features. To develop speech spectrum formant frequencies are very issential.But formants are very difficult to find from given speech signal and sometimes they may be not found clearly.Thats why instead of estimating the resonant frequencies, formant-like features can be used.

Depending upon the application the area of speaker recognition is divided into two parts. One is identification and other is verification. In speaker identification the aim is to match input voice sample with available voice samples. And in speaker verification, from available voice sample to determine the person who is claiming. Again in case of speaker identification there are two types, one is text –dependent and another is text-independent. The success of speaker identification in both cases in depends upon the various speaker characteristics which differ the one speaker from other [1].The speaker identification is divided into two components: feature extraction and feature

classification. These two components are attached with each other[13].In speech processing, speech is processed on the basis of frame by frame. Where frame may be speech or silence. But in speaker identification, the useful frame is speech frame, not a silence frame, which contains higher information about the speaker. The usable speech frames can be defined as frames of speech that contain higher information content compared to unusable frames with reference to a particular application[2].

Speaker identification and adaption have various applications than speaker verification. In speaker identification for example the speaker can be identified by his voice, where in case of speaker verification the speaker is verified using database[3].Here in this paper pitch is used for speaker identification. Pitch is nothing but the fundamental frequency of that particular person. This is one of the important characteristic of human being, which differ them from each other.

2. THEORETICAL ANALYSIS

Speech can be sampled at 8 KHz, 16Khz, 44.1Khz, 41.41Khz. But here speech will be sampled at 8 KHz.

   • Pitch calculation: Pitch represents the perceived fundamental frequency of a sound.

   • (MFCC) evaluation: This is the best and popular method for speaker identification. MFCC's are based on the known variation of the human ear's critical bandwidths with frequency. The MFCC technique makes use of two types of filter, namely, linearly spaced filters and logarithmically spaced filters[4].
   • Inverted Mel-Frequency Cepstral Coefficients(IMFCC)evaluation: the Inverted Mel-Frequency Cepstral Coefficients (IMFCC) is useful feature set for speaker identification, which is generally ignored by MFCC coefficients.

   • Analysis of Gaussian mixture model: It use as a representation of speaker identity for text independent   speaker identification.
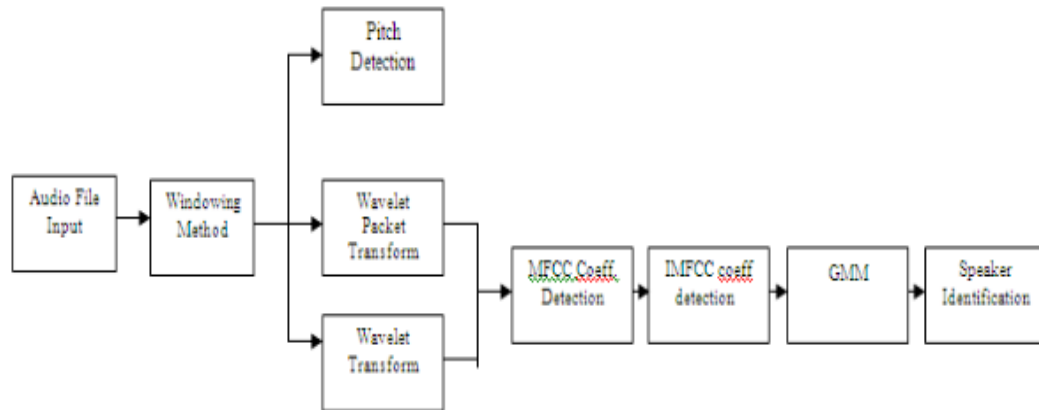
3. BLOCK DIAGRAM

Figure.1. Block Diagram for Speaker Identification

4. PITCH CALCULATION

Pitch represents the perceived fundamental frequency of a sound. Pitch is actually related to the period. The American National Standard Institute defines pitch as that attribute of auditory sensation in terms of which sounds may be ordered on a scale extending from low to high (ANSI 1973).For detection of pitch, the autocorrelation pitch detector is best and reliable pitch detector. The autocorrelation computation is made directly on waveform and is fairly straightforward (albeit time consuming) computation [5].

Recording of 4 male and 4 female

Evaluation of pitch:

1) Output Window:
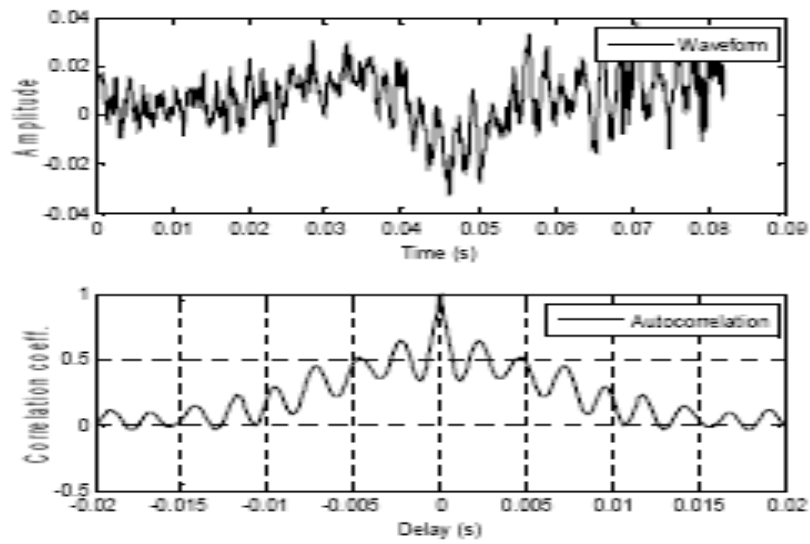
Female Voice Output Pitch Fx=431.507 Hz.

Figure.2.correlation coefficients for female voice

2) Output Window:

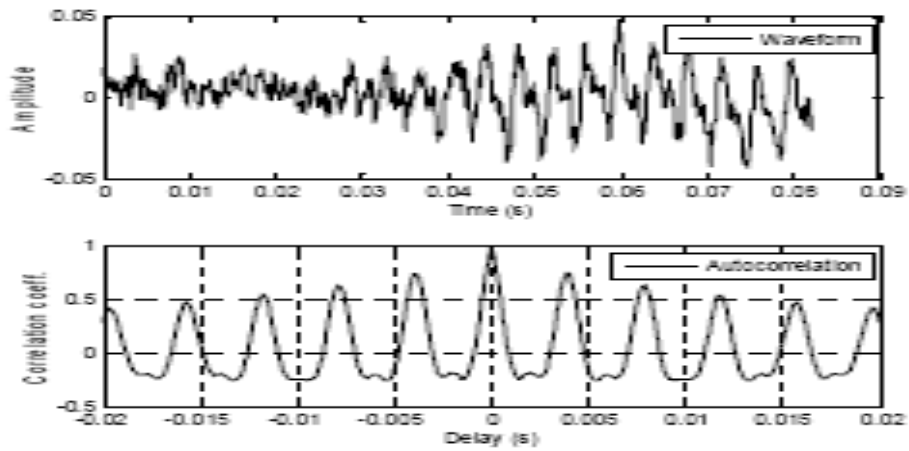**Male voice Output: Pitch== Fx=250.284Hz**

Figure.3.correlation coefficients for male voice

5. MFCC & IMFCC EVALUATION

The speaker-specific vocal tract information is mainly represented by spectral features like mel-frequency cepstral coefficients (MFCCs) and linear prediction (LP) cepstral coefficients. MFCCs are widely used spectral features for speaker recognition. Computation of the MFCCs differs from the basic procedure described earlier, where the log-magnitude spectrum is replaced with the logarithm of the mel-scale warped spectrum, prior to the inverse Fourier transform operation. Hence, the MFCCs represent only the gross characteristics of the vocal tract system[6]. IMFCC contains complementary information present in high frequency region,which is generally neglected sometimes.
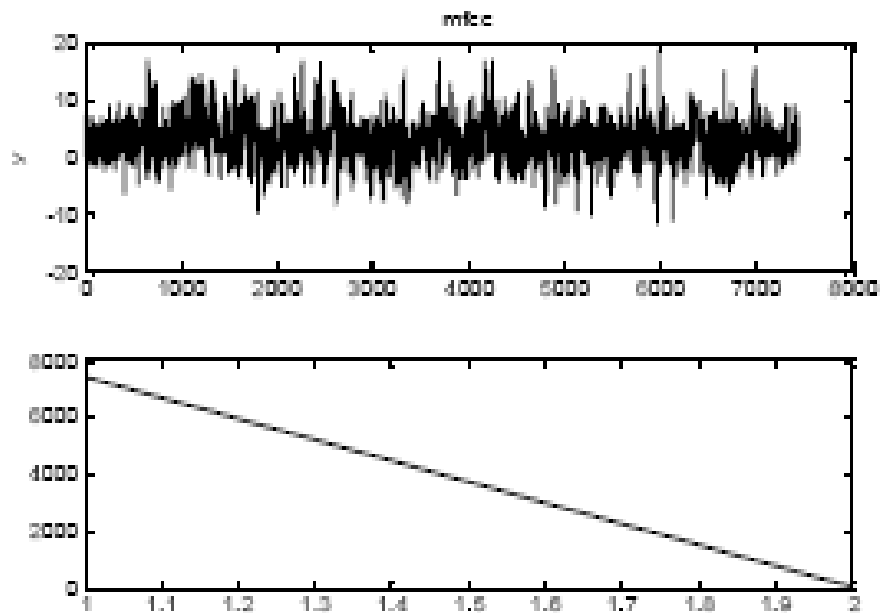

Figure. 4. MFCC coefficients

6. ANALYSIS OF GAUSSIAN MIXTURE MODEL

GMMs are commonly used as a parametric model of the probability distribution continuous measurements or features in a biometric system, such as vocal-tract related spectral features in a speaker recognition system.

GMMs are commonly used as a parametric model of the probability distribution of continuous measurements or features in a biometric system, such as vocal-tract related spectral features in a

speaker recognition system. Various forms of GMM feature extraction are outlined, including methods to enforce temporal smoothing and a technique to incorporate a prior distribution toconstrain the extracted parameters. Gaussian mixture models have proven to be a powerful tool for distinguishing acoustic sources with different general properties. This ability is commonly exploited in tasks like speaker identification and verification, where each speaker or group is modeled by GMM.The major advantage lies in the fact that they do not rely on any segmentation of the speech signal. A fact that makes them ideal for on-line application. However this advantage means at the same time, that they are not suitable for modeling temporal dependencies but this disadvantage is of minor importance, if the focus lies on the representation of global spectral properties [7].

7. FUTURE WORK

 In particular, the incorporation of some form of log- spectral normalization prior to estimating the GMM features could be investigated, as these yield significant improvements when applied to MFCC and PLP features on larger tasks. Work with formant estimation techniques has achieved smoother and more consistent trajectories using continuity constraints. Since the EM algorithm is a statistical approach, it could be possible to apply similar techniques using cost functions to the estimation of the GMM components. A subset of the Gaussian components estimated from the spectrum could be selected using a DP alignment and a cost function based on the continuity and reliability of estimate. Further investigation could be performed into other methods for estimating the GMM parameters as well, using other forms of trajectory constraint or implementing class-dependant priors combining the two features together. It could also be possible to investigate alternative schemes to use the condensed metric when combining the features. on the estimated features. The technique for combining the GMM and MFCC features which yielded the lowest WER was the concatenative approach. However, it may be interesting to investigate other methods for The use of constrained MLLR schemes suggests that these transforms are not appropriate for the GMM features. Further research could be performed into alternative transformations using non-linear adaptation schemes for the GMM features. Other transforms of the GMM features may also be possible.

8. CONCLUSION

This paper has evaluated the use of pitch for Robust Speaker Identification. This is how we can evaluate the pitch and Mel Frequency Cepstrum Coefficients. There are various methods to evaluate pitch and MFCC. These parameters help us to identify the speaker. There are various applications of speaker identification, e.g. authentication and all, which can be helpful in day today life.

9. REFERENCES

[1] D.A. Reynolds,R.C.rose,"Robust text-independent speaker identification using Gaussian mixture speaker model",IEEE,1995

[2] R.V Pawar, P.P.Kajave, and S.N.Mali," Speaker Identification using Neural Networks", World Academy of Science, Engineering and Technology 12 2005.

[3] Tomi Kinnunen, Evgeny Karpov, and Pasi Fr¨anti,"Real time speaker identification and verification",IEEE.

[4] Md. Rashidul Hasan, Mustafa Jamil, Md. Golam Rabbani Md. Saifur Rahman,"Speaker identification using Mel Frequency Cepstral coefficients".

[5] L. R. Rabiner," On the Use of Autocorrelation Analysis for Pitch ",IEEE ,Feb 1977

[6] K. Sri Rama Murty and B. Yegnanarayana," Combining Evidence From Residual Phase and

MFCC Features for Speaker Recognition",Jan,2006

[7] R.Falthauser,T.Pfau,G.Ruske,"On-line speaking rate estimation using gaussian mixture models".

[8] Ruhi sarakaya ,bryan pellon and John Hansen," Wavelet packet transform features with application to speaker identification" ,IEEE,JUNE 1998

[9] Bryan L. Pellon and John Hansen," An efficient scoring algorithm for Gaussian mixture model based speaker identification", student member , IEEE and John Hansen, senior member, IEEE, NOV 1998.

[10] Herbert Gish and Michel Schmidt," text- independent speaker recognition", IEEE signal processing magezine, OCT 1994. [4] Douglas A.Reynolds and Richard C. Rose," Robust text – independent Speaker identification using Gaussian.

[11] O. Farooq and S. Datta," Mel Filter-Like Admissible Wavelet Packet Structure for Speech Recognition " IEEE JULY 2001.

[12] Tianhorng Chang and C.C.Jay Kuo," texture analysis transfer" IEEE OCT 1993

[13] Douglas A. Reynolds," experimental evaluation of features for robust speaker identification",IEEE,OCT1994.

[14] M.S.Sinith, Anoop Salim, Gowri Sankar K, Sandeep Narayanan K V, Vishnu Soman," A Novel Method for Text-Independent Speaker Identification Using MFCC and GMM",IEEE,2010

[15] Ozlem Kalinli, Michael L. Seltzer, Jasha Droppo, and Alex Acero," Noise Adaptive Training for Robust Automatic Speech Recognition",IEEE NOV 2010.

[16] Longbiao Wang, Kazue Minami, Kazumasa Yamamoto, Seiichi Nakagawa," Speaker identification by combining MFCC and phase information in noisy environment",IEEE,OCT 2010.

[17] Md Fozur Rahman Chowdhury, Sid-Ahmed Selouani, Douglas O'Shaughnessy," Text-independent distributed speaker identification and verification using GMM -UBM speaker models for mobile communication",IEEE,2020.

[18] Gyanendra K Verma and U.S.Tiwary," Text Independent Speaker Identification using Wavelet Transform",IEEE,2010.

[19] James G. Lyons, James G. O'Connell, Kuldip K. Paliwal," Using Long-Term Information to Improve Robustness in Speaker identification",IEEE,2010.