Comparative Study of Different Techniques in Speaker Recognition: Review

Sonali T. Saste[1], Prof. S. M. Jagdale[2]

[1]Department of Electronics and Tele-Communication Engineering, Bharati Vidyapeeth College of Engg. For Women, Pune, Maharashtra, India
[2] Department of Electronics and Tele-Communication Engineering, Bharati Vidyapeeth College of Engg. For Women, Pune, Maharashtra, India

Abstract— *The speech is most basic and essential method of communication used by person.On the basis of individual information included in speech signals the speaker is recognized. Speaker recognition (SR) is useful to identify the person who is speaking. In recent years speaker recognition is used for security system. In this paper we have discussed the feature extraction techniques like Mel frequency cepstral coefficient (MFCC), Linear predictive coding (LPC), Dynamic time wrapping (DTW), and for classification Gaussian Mixture Models (GMM), Artificial neural network (ANN)& Support vector machine (SVM).*

Keywords— ANN,DTW,GMM, LPC, MFCC.

I. INTRODUCTION

Speech is one of the most important ways to communication. It gives many levels of information to the listeners. It conveys the message; also give information about gender, emotion and identity of speaker. There are many situations where the correct recognition of speaker is required. In biometric there are many ways to identify the person like finger print, palm, iris, face, voice recognition. The objective of the speaker acknowledgment is to describe and separate the data from discourse or speaker voice. Text independent and text dependent speaker recognition are the types of speaker recognition. [2]The behavioral aspect of human voice is utilized for distinguishing proof by changing over a talked expression from simple to computerized design, and extracting unique vocal characteristics, such as pitch, frequency, tone and cadenceto set up a speaker model or voice test. In voice acknowledgment, enlistment and confirmation procedures are included.Enrollment process describes the registration of speaker by training his voice features.[3]

II. RELATED WORK

There are many methodologies have been proposed for speaker recognition. This is a system that can recognize a person based on his/her voice. This is accomplished by actualizing complex flag handling calculations that keep running on an advanced PC or a processor.The speaker recognition system can be classified as speaker identification or speaker verification. [7]

1. Mel frequency cepstral coefficient (MFCC): This is the most powerful feature extraction technique, used in speaker recognition which works on human auditory system. Seiichi Nakagawa used the MFCC for speaker identification and verification.[1] Abdelmajid H. Mansour also used MFCC for voice recognition.[3]

2. Linear predictive coding (LPC): LPC is the simplified model for the speech production. It is a technique for determining the basic parameters of speech and provides precise estimation of speech parameters and computational model of speech.Speech test can be approximated as a straight mix of past speech tests is the fundamental thought behind LPC. [11] KinnalDhameliya used the LPC method for feature extraction in speaker recognition.[2]

3. Dynamic time wrapping (DTW): DTW mainly focuses on matching of two sequences of feature vectors. It is used to calculate the distance between two time series that vary in time.[3] Abdelmajid H. Mansour uses DTW for feature extraction in voice recognition system.

III. FEATURE EXTRACTION

Feature extraction is the way toward distinguishing distinctive elements from the information flag. After the pre-processing this feature extraction is finished. For speech there are many elements like MFCC, pitch, energy, formant and so forth are extracted.

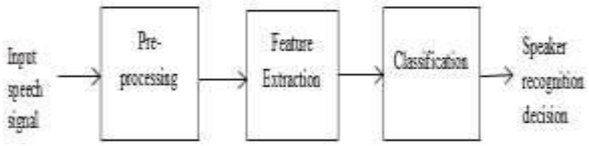

Fig.1: Basic speech recognition system

1. MFCC: MFCC is classical approach to analyze speech signal, which represents the short-term power spectrum of sound, based on linear cosine transform of a log power spectrum on a nonlinear Mel scale of frequency. It is most popular system because it approximates the human auditory system response more closely than other system.[1][8]

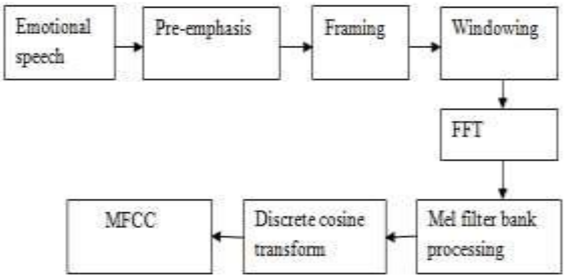

Fig.2: Steps involved in MFCC algorithm.

MFCC can be computed by using the formula,F(Mel) = [2595* log 10 [1+F] /700]…………(1)The recognition accuracy is high. That means the performance rate of MFCC is high. MFCC captures main characteristics of phones in speech. Low Complexity when using MFCC.[18]

2. LPC: To decide the fundamental parameters of speech and gives exact estimation of speech parameters and computational model of discourse this LPC system is utilized. Speech test can be approximated as a direct blend of past speech tests is the fundamental thought behind LPC.[11]The following fig. shows the steps involved in the LPC feature extraction.

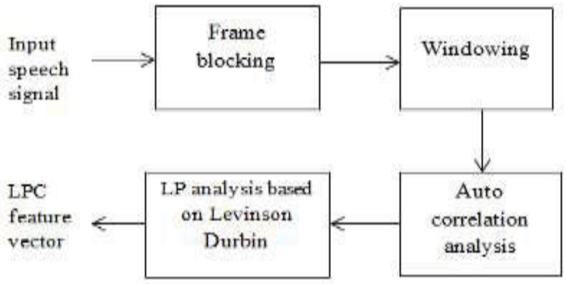
Input speech signal
Frame blocking
Windowing
Auto correlation analysis
LP analysis based on Levinson Durbin
LPC feature vector

Fig.3: Steps involve in LPC feature extraction

Is a reliable, accurate and robust technique for providing parameters which describe the time varying linear system which represents the vocal tract.Computation speed of LPC is good and provides with accurate parameters of speech. LPC is useful for encoding speech at low bit rate.[18]

3. DTW: This algorithm depends on element programming which utilized for measuring similarity between two time arrangement which may differ in time or speed.This warping between two time series can then be used to find corresponding regions between the two time series or to determine the similarity between the two time series.[6]The groupings are "warped" non-directly in the time measurement to decide a measure of their likeness free of certain non-straight varieties in the time measurement. This succession arrangement technique is frequently utilized as a part of time arrangement grouping. In spite of the fact that DTW measures a separation like amount between two given groupings, it doesn't ensure the triangle disparity to hold. Notwithstanding a closeness measure between the two groupings, a purported "warping path" is delivered,by warping according to this path the two signals may be aligned in time.

DTW that is ordered into highlight based example acknowledgment procedures does not have to set up (or prepare) an arrangement display in progress; thusly, it is for the most part seen as an adroitly straightforward and coordinate acknowledgment strategy. Ordinary DTW is by and large utilized for performing discourse acknowledgment, and few reviews are believed to utilize the DTW method for execute speaker acknowledgment.[19]

4. Other feature extraction techniques: Various other feature extraction strategies are accessible by basically altering the above element extraction procedures, including the following:

- Mean and Variance of the residual phase
- Delta and double delta of MFCC features

Along these many feature extraction strategies are accessible. One can utilize any method as suitable by application and can show signs of improvement acknowledgment exactness. Presently these elements we need to store in a database for N distinctive speakers and this database will be utilized for classification.[2]

IV. CLASSIFICATION

1. GMM: A Gaussian mixture model is a probabilistic model that accepts every one of the information focuses are produced from a blend of a limited number of Gaussian circulations with obscure parameters. One can consider blend models as summing up k-cluster grouping to join data about the covariance structure of the information and additionally the focuses of the inactive Gaussians.[10]

The Gaussian Mixture Model executes the desire expansion (EM) calculation for fitting blend of-Gaussian models. It can likewise draw certainty ellipsoids for multivariate models, and register the Bayesian Information Criterion to evaluate the quantity of bunches in the information. A Gaussian Mixture. fit strategy is given that takes in a Gaussian Mixture Model from prepare information. Given test information, it can dole out to each example the Gaussian it generally most likely has a place with utilizing the Gaussian Mixture. predict technique.[1][10]

GMM require less preparing and test data. It performs better as it requires less measures of information to prepare the classifier consequently memory prerequisite is less.[19]

2. ANN: An artificial neural system (ANN) is a computational model in light of the structure and elements of organic neural systems. Data that moves through the system influences the structure of the ANN in light of the fact that a neural system changes - or learns, it could be said - in view of that info and yield. ANNs are viewed as nonlinear factual information demonstrating apparatuses where the mind boggling connections amongst data sources and yields are displayed or examples are found.[5][8]ANNs is that element extraction and speaker demonstrating can be consolidated into a solitary system, empowering joint improvement of the (speaker-dependent) include extractor and the speaker model.[20]

3. SVM: The support vector machine is mainly used for the speaker verification or speaker identification.SVM is a twofold classifier which models the choice limit between two classes as an isolating hyperplane. In speaker check, one class comprises of the objective speaker preparing vectors, and alternate class comprises of the preparation vectors from an background data. Utilizing the marked preparing vectors, SVM streamlining agent finds an isolating hyperplane that amplifies the edge of division between these two classes.[19] In a high-dimensional space, the two classes are less demanding to isolate with a hyperplane. Instinctively, linear hyperplane in the high dimensional portion include space compares to a nonlinear choice limit in the first information space.[20]

V. CONCLUSION

In this paper we discuss feature extraction techniques and classification techniques forspeaker recognition. For high performance and low complexity MFCC is preferred. With low bit rate for encoding LPC is useful.In DTW not need to establish (or train) a classification model in advance. In classification for less memory and less preparing and test data GMM is useful.When there is combination of feature extraction and speaker modeling in to a single network, ANN is useful. In binary classification SVM is simple and effective algorithm. Depending upon the application one maychoose any technique or it is also possible to use combination of one or more techniques out of this described techniques for better speaker recognition system.

ACKNOWLEDGMENT

I would like to express my gratitude towards Prof. Dr. H. V. Vankudre, Principal, for encouraging us, also my gratitude to all those who supported me for this paper work.

REFERENCES

[1] Seiichi Nakagawa, Member, IEEE, Longbiao Wang, Member, IEEE, and Shinji Ohtsuka “Speaker Identification and Verification by Combining MFCC and Phase Information”*IEEE transactions on audio, speech, and language processing, vol. 20, no. 4, may 2012.*

[2] KinnalDhameliya, Ninad Bhatt “Feature Extraction and Classification Techniques for Speaker Recognition: A Review” *978-1-4799-7678-2/15/$31.00 ©2015 IEEE*

[3] Abdelmajid H. Mansour , Gafar Zen AlabdeenSalh, Khalid A. Mohammed “Voice Recognition using Dynamic Time Warping and Mel-Frequency Cepstral Coefficients Algorithms”*International Journal of Computer Applications (0975 – 8887) Volume 116 – No. 2, April 2015*

[4] B. P. Bogert, J. R. Healy and J. W Tukey, “The Quefrency Analysis of Time Series for Echoes: Cepstrum, Pseudo-Autocovariance, Cross Cepstrum, and Saphe Cracking,” *Proceedings of the Symposium on Time Series Analysis, 1963, pp. 209-243.*

[5] Alfredo Maesa, Fabio Garzia, Michele Scarpiniti and Roberto Cusani, “Text Independent Automatic Speaker Recognition System Using Mel- Frequency Cepstrum Coefficient and Gaussian Mixture Models,” *Journal of Information Security, Vol. 3, No. 4, October 2012, pp.335-3*

[6] Pratik K. Kurzekar , Ratnadeep R. Deshmukh , Vishal B. Waghmare , Pukhraj P. Shrishrimal “A Comparative Study of Feature Extraction Techniques for Speech Recognition System”*International Journal of Innovative Research in Science, Engineering and Technology (An ISO 3297: 2007 Certified Organization) Vol. 3, Issue 12, December 2014*

[7] Miss. Sarika S. Admuthe, Mrs. ShubhadaGhugardare “Survey Paper on Automatic Speaker Recognition System” *IJECS Volume 4 Issue 3 March, 2015 Page No.10895-10898*

[8] Sundeep Sivan and Gopakumar C, “An MFCC Based Speaker Recognition using ANN with Improved Recognition Rate,” *International Journal of Emerging Technologies in Computational and Applied Sciences (IJETCAS), Vol. 8, Issue 4, March-May 2014, pp. 365-369.*

[9] Jianglin Wang, AnJi and Michael T. Johnson, “Features for Phoneme Independent Speaker Identification,” *IEEEInternational Conference on Audio Language and Image Processing (ICALIP), Shanghai, July 2012, pp. 1141-1145.*

[10] Nidhi Desai, KinnalDhameliyaand Vijayendra Desai, " Feature Extraction and Classification Techniques for Speech Recognition: A Review," *International Journal of Emerging Technology and Advanced Engineering, Volume 3, Issue 12, December 2013, pp. 367-371.*

[11] Seiichi Nakagawa, Longbiao Wang and Shinji Ohtsuka, "Speaker Identification and Verification by Combining MFCC and Phase Information," *IEEE transaction on audio, speech and language processing, Vol. 20, No. 4, May 2012, pp. 1085-1095.* 13.

[12] Jianglin Wang, AnJi and Michael T. Johnson, "Features for PhonemeIndependent Speaker Identification," *IEEE International Conference on Audio Language and Image Processing (ICALIP), Shanghai, July 2012, pp. 1141-1145.*

[13] Srikanth R Madikeri and Hema A Murthy, "Mel Filter Bank Energy- Based Slope Feature and Its Application to Speaker Recognition," *IEEE National Conference on communication (NCC), Bangalore, January 2011, pp. 1-4.*

[14] Hemant A. Patil, Purushotam G. Radadia and T. K. Basu, "Combining Evidences from Mel Cepstral Features and Cepstral Mean Subtracted Features for Singer Identification," *IEEE International Conference on Asian Language Processing, Hanoi, November 2012, pp. 145-148.*

[15] S. Rajasekaran and G.A. VijayalakshmiPai, "Neural Networks, Fuzzy Logic, and Genetic Algorithms Synthesis and Applications", *PHI, 2003.*

[16] Shahzadi Farah and AzraShamim, "Speaker Recognition System Using Mel-Frequency Cepstrum Coefficients, Linear Prediction Coding and Vector Quantization," *3rd IEEE International Conference on Computer, Control &Communication (IC4), Karachi, September 2013, pp. 1-5.*

[17] Nidhi Desai, KinnalDhameliya and Vijayendra Desai, "Recognizing voice commands for robot using MFCC and DTW," *International Journal of Advanced Research in Computer and Communication Engineering, Vol. 3, Issue 5, May 2014.*

[18] ShreyaNarang, Ms. Divya Gupta "Speech Feature Extraction Techniques: A Review"*IJCSMC, Vol. 4, Issue. 3, March 2015, pg.107 – 114*

[19] Ing-Jr Ding, Chih-Ta Yen and Da-Cheng Ou "A Method to Integrate GMM, SVM and DTW for Speaker Recognition"*International Journal of Engineering and Technology Innovation, vol. 4, no. 1, 2014, pp. 38-47*

[20] TomiKinnunen, Haizhou Li "An overview of text-independent speaker recognition:From features to supervectors"*T. Kinnunen, H. Li / Speech Communication 52 (2010) 12–40*