International Journal of Recent Technology and Engineering (IJRTE)
ISSN: 2277-3878 (Online), Volume-8 Issue-2, July 2019

Speaker Recognition system and it's Applications

Arundhati Y. Niwatkar, Y. K. Kanse

Abstract: ***Speaker recognition is the task in which the speaker is identified based on various features from his speech. Speaker recognition is combination of various mathematical operations in which training and testing is the major part. For speaker recognition its very important to extract the features. So far many researches are going on about feature extraction techniques like MFCC, IMFCC etc. In which features can be extracted, but for the exact speaker recognition its very important to get the exact and accurate features so that we can increase the success rate of speaker recognition. For any speaker recognition system feature extraction is the primary and very important step. So the precise result depends on the accurate result of feature extraction technique. In this paper we are proposing a modified feature extraction system.***

Index Terms: ***IMFCC,MFCC***

I. INTRODUCTION

The task of speaker recognition is to determine the identity of a speaker. In order for humans to recognize voices, the voices must be familiar similarly for the tool using for it. For speaker identification there are two parts, one is training and other is testing. In training step. Training of the system takes place with different voices and in testing phase we are going to test all the voices with trained voices. For that purpose data has to be created which is also called as database for different voices.

Speaker recognition system has two major stages. First is feature extraction and second is to classify the speaker based upon different characteristics of speech of a speaker. Every speaker recognition system uses these features as the primary input to recognize the correct speaker[1].

Voice of every human is unique and hence each speaker is having different speech features. We can take advantage of the nature's gift and we can use human voice as his unique identity in some applications[3-4].Many researches are going on but still some challenge problems are there who are attracting the researchers to do more study about the system. The main aim of this paper is to propose a modified MFCC feature system to increase the success rate of speaker recognition system[2-3].

Revised Manuscript Received on 30 July 2019.

* Correspondence Author

Arundhati Y. Niwatkar*,Assistant Professor in Department of Electronics, in UMIT, SNDT Women's University, Mumbai.

Dr. Y. K. Kanse, Associate Professor in Department of Electronics,, Karmaveer Bhaurao Patil College of Engg, Satara,India.

II. FEATURE EXTRACTION

A. MFCC processor

Speech contains significant energy from a frequency up to around 5 KHz. to study speech signal the concept of time varying Fourier representation is used. However temporal characteristics of speech signal such as energy,zero crossing ,correlation etc are assumed constant over a short period. i.e. its characteristics are short-time stationary. To transform speech signal in frequency domain its very important to make short duration blocks of speech signal of short duration. MFCC is based on the human auditory system[4-5]. The human perception of the frequency contents of sound for speech signals follows a logarithmic scale,called as 'Mel scale' .

B. Linear Predictive Coefficients (LPC)

For producing speech signal vocal folds, which are situated in vocal tract plays very important role. Using these biological structure we can determine the coefficients. LPC is the method in which prediction is the key word. All coefficients are predicted from the past values. So this LPC method represents the vocal tract just by varying the diameter.

C. Linear Predictive Cepstral Coefficients (LPCC)

LPCC find out the difference between characteristics of vocal tracts and computes the values using LPCC algorithm. Human vocal tract system can be represented by filters because of its biological structure. This model assumes two important signals i.e. glottal pulse and random noise. Glottal pulse generator generates voiced signal where random noise generates unvoiced signals.

III. MODELING OF THE SPEAKER

In this process first feature vectors are extracted using feature extraction techniques from the speech signal and with the help of these feature vectors speaker model can be created. Many modeling techniques can be created. Different types of modeling techniques are used by different people. To select the method of modeling some factors should be considered. That are, type of speech signal, expected result, computation techniques to be used etc. Some speaker modeling techniques are given below.

A. Hidden Markov Models

It is the model based representation of the speaker based on statistical properties of the speaker. It shows how speaker generates sound. To form the model, speaker parameters from the speech signal should be found and these parameters should enroll the model, which will represent that particular speaker only.

At the time of speaker recognition different likelihood algorithms can be used. This is used for both text-independent and Tex-dependant applications.

B. Gaussian Mixture Model

This is the probabilistic modeling technique which is widely used. This can be approximated by using basic probability density function(Pdf).

It is given by,

$$P(x/\lambda)=\sum_{i=1}^{M} p_i b_i(x)$$

where M is the number of component densities, the mixture weights for i = 1, .., M. x is a D dimensional observed data. with mean vector μ and covariance matrix Σ. Every speaker can be represented by a GMM and model denoted by λ . All GMM parameters should be calculated precisely to obtain good score in speaker recognition system. For the process,ML algorithm is used for estimation[7].

IV. PROPOSED METHODOLOGY

In this system GMM is used with MFCC feature extraction technique. Different samples are collected from different people(male+ female). Following flow chart explains the steps:

Step1:

Step 2:

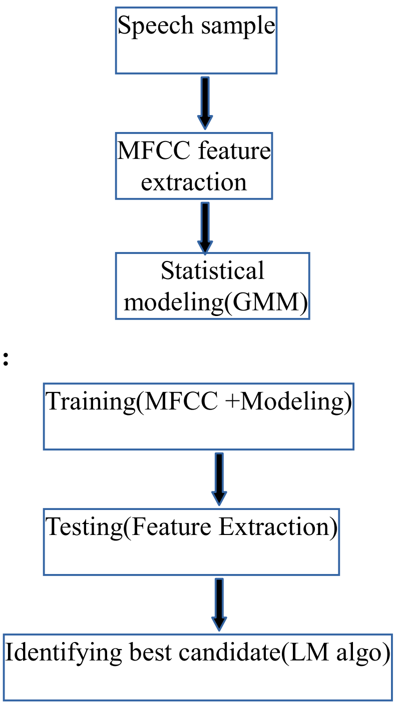

Fig. 1. **Steps in speaker recognition using GMM**

In training phase statistical models of each speaker is formed by calculating it's means and co-variance matrix.which is given in following table.

Table- I: Results of GMM

	μ	σ^2	σ
Speaker 1	-0.47	62.41	7.921
Speaker 2	0.187	25.60	5.052
Speaker 3	-0.61	37.08	6.089
Speaker 4	8.264	82.44	9.079
Speaker 5	5.293	35.05	5.920

Table- II: Score of GMM

	Speaker 1	Speaker 2	Speaker 3	Speaker 4	Speaker 5
Speaker 1	**-19.826**	-20.430	-34.575	-20.801	-20.444
Speaker 2	-21.980	**-18.870**	-41.090	-20.114	-21.331
Speaker 3	-25.112	-23.717	**-18.569**	-22.648	-20.623
Speaker 4	-21.987	-19.906	-39.913	**-19.151**	-19.985
Speaker 5	-23.033	-22.308	-26.162	-21.048	**-18.197**

Each column i represents the test recording of Speaker i
Each row j represents the training recording of Speaker j
The diagonal elements (corresponding to same speaker comparison)

V.APPLICATIONS

Speaker recognition has many applications which helps to improve human life. Some techniques are as follows.

A. Authentication for money transactions

For high value of money this application is used by using telephonic application. i.e. by setting particular password of a speaker we can control the access.

B. In legal cases

In some legal cases verification of the exact speaker from a speech sample can be considered as a valid proof.

C. Access control

Access to many applications e.g. certain device, documents, websites, digital diary, door lock can be controlled by setting a password in a speech format. Only that speaker can access these applications.

D. Customization

Many devices or applications can be customized using user voice. Many commands can be set for a device to operate it from a distance.

VI. FACTORS AFFECTING ON SPEAKER RECOGNITION SYSTEM

Following are some factors which affects the performance of speaker recognition system.

A. Microphones

In any speaker recognition system microphone plays very important role.

Retrieval Number: B3719078219/19©BEIESP
DOI: 10.35940/ijrte.B3719.078219
Journal Website: www.ijrte.org

Published By:
Blue Eyes Intelligence Engineering
Sciences Publication

It is used to take voice samples from a speaker. To get accurate results from the system, it is essential that the microphone should be of a good quality.

B. Noise

It can affect the entire speaker recognition system. Its very important to record the samples with good quality of microphones in noiseless environment. Also during training and testing of a speaker noise should be avoided. Some channel noise also can affect the system.

C. Length of a voice sample

It depends on speaker recognition system applications. i.e. generally voice samples with greater length gives more precise and accurate results. But if the background is noisy, shorter voice samples gives better results.

D. Text

While designing the speaker recognition system, its very important to make it clear that whether the system is Text-independent or Text-dependent. Comparatively Tex-independent system is complicated to design and implement.

E. Speaker database

There are two important phases in speaker recognition system. i. e. training and testing of a speaker. By this process number of speakers are enrolled in the system database. To get accurate result large number of speakers should be enrolled.

F. Health of a speaker

While taking samples this can affect the accuracy of the system. e.g. if speaker is having running nose, fever or throat infection, this may give wrong result.

VII. CHALLENGES AND STRENGTHS

Every system has its strengths and challenges. Following are the same for speaker recognition system. Here the challenges can be treated as the future scope of that system.

A. Hardware and software used

For speaker recognition Hardware required is not that much expensive. e.g. microphone. For coding and pattern matching algorithms many software platforms are available. Among which popular tool is MATLAB.

B. User friendly

Speaker recognition system is user friendly once it is implemented. e.g. it can be used for biometric application[8]. Even a non-technical person can use it for many application. Also it is portable.

C. Spoofing Attack

In speaker recognition system,spoofing attack is possible. e.g. one can use the recorded voice of the authenticated speaker. Also data can be hacked from the system database[9].

D. Precise

Speaker recognition system is not precise. For training and testing of a speaker it is very important to follow the rules of acoustics. Also mispronunciation, MTI(mother tongue influence) can give wrong results[10].

E. Unavoidable health issues

In speaker recognition system speech signal of a speaker plays very important role. But this system is not an universal system. The deaf people who are mute cannot use this system. Ageing is a big issue, where voice of a person changes with age. Emotional state of a speaker while taking sample is also important[11].

VIII. CONCLUSION AND FUTURE SCOPE

In this paper we have presented the different aspects of speaker recognition system. Also we have presented some techniques used in this system. Here, in this case different algorithms are used for pattern matching. Also speaker modeling can be done by using various methods. But now a days the latest trend is about Artificial intelligence[12]. In speaker recognition system also one can use this technique. i.e. by using Deep neural network[13] can develop a speaker recognition system. Wavelets also can be used for speaker recognition system, as they have more advantages than DFT. Also in speaker recognition system, feature extraction is very important. These feature are used for speaker recognition. In this paper we have given some feature extraction techniques like MFCC,LPC etc. similarly one can try different feature extraction techniques e.g. IMFCC[14].
Also in this paper we have discussed about the weaknesses and strengths of the system In todays world, to recognize a person DNA is used. And no one can challenge that result. In many legal cases DNA samples are used and considered as a strong proof. But speaker recognition system is not precise. One cannot rely on this system because this system is having some disadvantages as discussed in the paper. So its very important to work on these disadvantages to make traditional system better.

REFERENCES

1. Douglas A. Reynolds, "An Overview of Automatic speaker Recognition Technology", IEEE, pp. IV – 4072 - IV– 4075, 2002.
2. Tomi kinnunen, Haizhou Li, "An overview of text- independent speaker recognition: From features to super vectors speech Communication 52, pp. 12-40, 2010.
3. Amin Fazel, Shantanu Chakrabartty, "An Overview of statistical Pattern Recognition Techniques for Speaker verification", IEEE Circuit and System Magazine, pp. 62-81, 2011.
4. Ajgou, R., Sbaa, S., Ghendir, S., Chamsa, A. and Taleb-Ahmed, A., "Robust remote speaker recognition system based on AR-MFCC features and efficient speech activity detection algorithm", In Wireless Communications Systems (ISWCS), 2014 11th International Symposium on (pp. 722-727). IEEE,August,2014.
5. Amit Kumar Singh, Rohit Singh, AshutoshDwivedi, "Mel Frequency Cepstral Coefficients Based Text Independent Automatic Speaker Recognition Using Matlab", International Conference on Reliability, Optimization and Information Technology (ICROIT), pp.524-527, February 2014.
6. P. Chakraborty1, F. Ahmed , Md. Monirul Kabir , Md. Shahjahan1, and Kazuyuki Murase "An Automatic Speaker Recognition System", M. Ishikawa et al. (Eds.): ICONIP 2007, Part I, LNCS 4984, pp. 517–526, 2008. Springer- Verlag Berlin Heidelberg 2008.
7. Reynolds D. A., & Campbell W. M.,"Text-independent speaker recognition", In Springer Handbook of Speech Processing (pp.763-782),Springer,Berlin,Heidelberg,2008.
8. Yan, Z. and Zhao, S., "A usable authentication system based on personal voice challenge", In Advanced Cloud and Big Data (CBD), 2016 International Conference on(pp. 194-199),IEEE, August 2016.
9. Stuttard D, & Pinto M,"The web application hacker's handbook: Finding and exploiting security flaws",John Wiley & Sons,2011.
10. Togneri, R., &Pullella, D."An overview of speaker identification: Accuracy and robustness issues",IEEE circuits and systems magazine, 11(2), 23-61,201.

11. Haq, S., Asif, M., Ali, A., Jan, T., Ahmad, N. and Khan, Y., "Audio-visual emotion classification using filter and wrapper feature selection approaches",Sindh University Research Journal-SURGE (Science Series), 47(1),2016.
12. Suryanarayana, L., Grannan, M.F. and Vishik, C.S., "AT&T Intellectual Property Methods of using biometric data in a phone system and apparatuses to perform the methods".,U.S. Patent 8,374,324,2013.
13. Tzirakis, P., Trigeorgis, G., Nicolaou, M.A., Schuller, B.W. and Zafeiriou, S.,"End-to-end multimodal emotion recognition using deep neural networks"IEEE Journal of Selected Topics in Signal Processing, 11(8), pp.1301-1309,2017.
14. Jin, Q. and Zheng, T.F.,"Overview of front-end features for robust speaker recognition",Proc. APSIPA.,2011.
15. A.E. Rosenberg, F.K. Soong, "Recent research inautomatic speaker recognition", Advances in Speech Signal Processing, Marcel Dekker, New York, pp. 701–738, 1991.
16. J.P. Campbell, "Speaker recognition: A tutorial", Proc.IEEE 85, pp. 1437–1462, 1997.

AUTHORS PROFILE

Arundhati Y. Niwatkar was awarded B.E. degree from TKITE, Warananagar, SUK. And awarded M.E. Degree from KIT's college of engineering ,Kolhapur.

Dr. Y. K. Kanse was awarded by Ph.D. degree from Shivaji University,Kolhapur. He is recognized guide in Shivaji University, Kolhapur

Retrieval Number: B3719078219/19©BEIESP
DOI: 10.35940/ijrte.B3719.078219
Journal Website: www.ijrte.org

Published By:
Blue Eyes Intelligence Engineering
Sciences Publication