Improving Performance of Speaker Recognition Using Prosodic Features

SUMMARY
of
THESIS

Submitted to
Babasaheb Bhimrao Ambedkar University
(A Central University)
Lucknow

BABASAHEB
BHIMRAO
AMBEDKAR
UNIVERSITY
LUCKNOW
प्रज्ञा शील करुणा
ESTABLISHED 1996

For the Degree of

Doctor of Philosophy

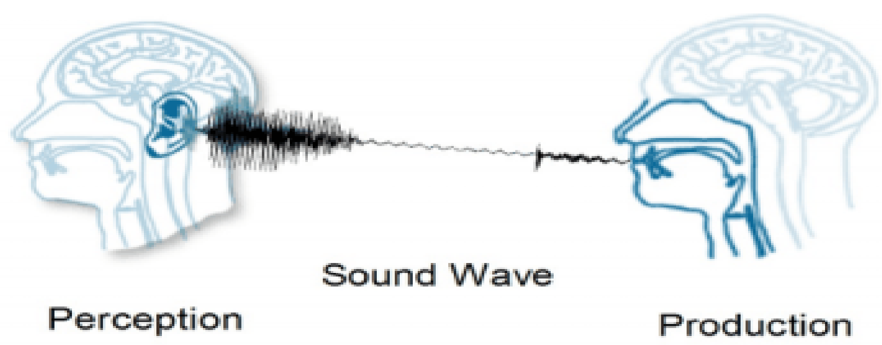
In

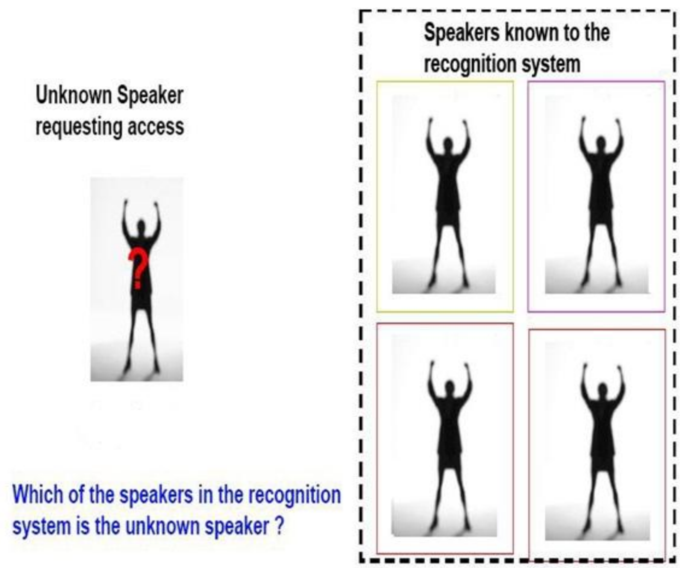
INFORMATION TECHNOLOGY

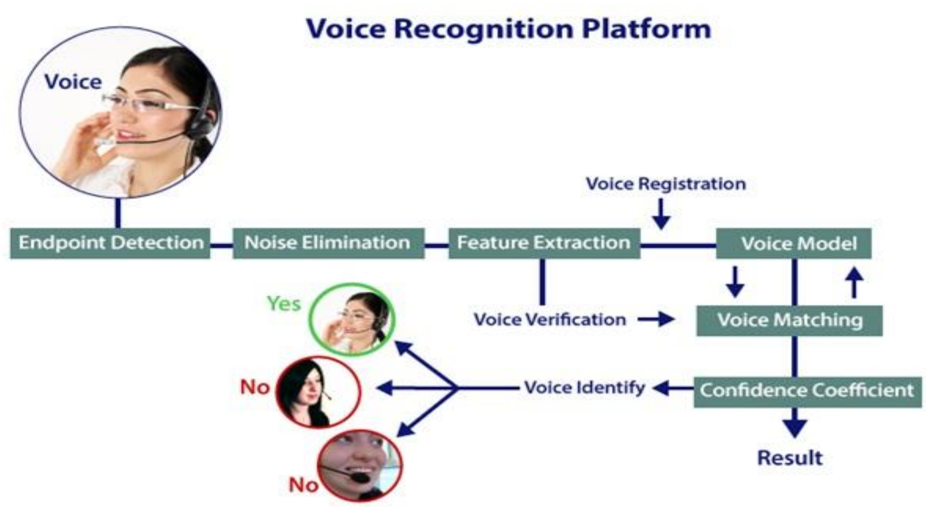
By:
Nilu Singh
(Enrollment No. 1293/15)

DEPARTMENT OF INFORMATION TECHNOLOGY
BABASAHEB BHIMRAO AMBEDKAR UNIVERSITY
(A CENTRAL UNIVERSITY)
VIDYA VIHAR, RAEBARELI ROAD, LUCKNOW-226 025 (U.P.), INDIA

2018


Sound Wave
Perception
Production


Unknown Speaker
requesting access
Speakers known to the
recognition system
Which of the speakers in the recognition
system is the unknown speaker ?


Voice Recognition Platform
Voice
Endpoint Detection
Noise Elimination
Feature Extraction
Voice Registration
Voice Model
Voice Verification
Voice Matching
Yes
No
No
Voice Identify
Confidence Coefficient
Result

MY VOICE IS MY PASSWORD

TABLE OF CONTENT

1 Introduction

Speech is a natural way to convey information by humans. Speech signal is enriched with information of the individual. Recognizing a person's individuality by his/her voice is known as Automatic Speaker Recognition (ASR). Speaker recognition falls in the category of biometric security systems. Biometric is related to human characteristics or individuality. Biometric verification or realistic authentication is used to recognize an individual through his/her voice's individual characteristic. Voice biometric includes behavioral or physiological measurements of individual. Behavioral biometric is performed by Voice, Signature, Keystrokes, and Typing etc. whereas physiological biometric includes iris, face, retina, fingerprints, ear, DNA etc. Now a days voice biometric is an emerging research area [1-2].

Human speech is a medium for expressing their thoughts during communication. Spoken language is the most natural way for human to transfer information. A speech signal is a complex signal which is packed with several knowledge resources such as acoustic, articulatory, semantics, linguistic and many more [3-4]. During communication, human easily understand information such as emotion, language, and mental status etc. This ability of human to decode information motivated many researchers to understand speech signal production and perception. This idea helps to developing a system which automatically extract and process the built in information in a speech signal. A person's voice is different from another due to the acoustic properties of speech signal. Speaker's voice is unique to an individual due to differences which occur as anatomical differences inherent in the vocal tract and the cultured speaking behaviors of different individuals [3-5].

Speaker recognition is a process of recognizing who is speaking on the basis of information included his/her speech signal/waves. In this digital era speaker recognition is the most useful biometric recognition technique [5]. Now days many organizations like bank, industries, access control systems etc. are using this technology for providing greater security to their vast databases [5-6]. Speaker recognition is broadly classified into speaker identification (1: N matching) and speaker verification (1:1 matching). Identification is considered as more difficult than verification [7]. This is intuitive that performance of speaker identification system affected by the number of registered speakers increases (the probability of incorrect decision increases). While the performance of speaker verification system is not affected by increase in voice database size since only two speakers are compared.

In last few years, requirement for authentication has been increased with the increasing digital world of information. It has already been proved that a biometrics authentication technique increases security levels. Speaker identification is the process of identifying an utterance from the known set of speakers while speaker verification is the process of accepting or rejecting the claimed identity. Speaker verification systems are the real example of biometric authentication systems. Further, it can be classified as text-dependent and text-independent [8]. The text-dependent systems are based on same utterance spoken by speaker in both cases i.e. training and testing while in text-independent systems it is not required to utter the same sentence/words during training and testing [7]. It is accepted that text-dependent systems provide more accurate results as both the content and voice can be compared that is speaker utters exact the sentence which he/she uttered during training. While text-independent recognition systems, may use either the same utterance or different for every verification/identification session.

Gaussian mixture model (GMM) is one of the popular approaches used to speaker modeling for speaker identification. GMM is used as two distinct ways for identification system; firstly, when training database principle is the maximum likelihood (ML) and parameter estimation is performed by using expectation maximization (EM) algorithm; and secondly when the training database principle is maximum a posteriori (MAP). In this case, GMM as a universal background model (UBM), is created for training database of speakers and these models are trained by the UBM (using registered speakers specific data) [9-11].

Speaker recognition basically has two categories; speaker identification and speaker verification. Speaker verification is used for those applications where speech is used as the key to authorize the identity claim of a speaker. Speaker identification is used to decide that a given utterance comes from a certain registered speaker. Speaker verification has larger usability than speaker identification. The basic purpose of speaker identification is crime investigation. It is used to decide which of the suspected speaker's voice match with the registered speakers. With the increase in voice database, difficulty of speaker identification increases. Speaker verification is independent of voice database population since it works only on binary decision that is acceptance or rejection. Speaker recognition system performance (recognition accuracy) is most affected by intersession variability (variability over time) and spectra of a speakers speech signal [1] [12].

2 Motivation for the Research

Research in speaker recognition systems have been continued for many years. This technology nowadays widely used for secure authentication. Speaker recognition is defined as the process of recognizing a person by his/her voice. This methodology allows user identity by their voice. The goal of speaker recognition system is to provide secure authentication in daily life such as telephone banking, access control, information services, security check for confidential areas etc. In the current scenario, it becomes a strong security feature for many confidential areas [4]. The performance of speaker identification system is affected by many factors. Figure- 1 shows the characteristics of different types of speech features.

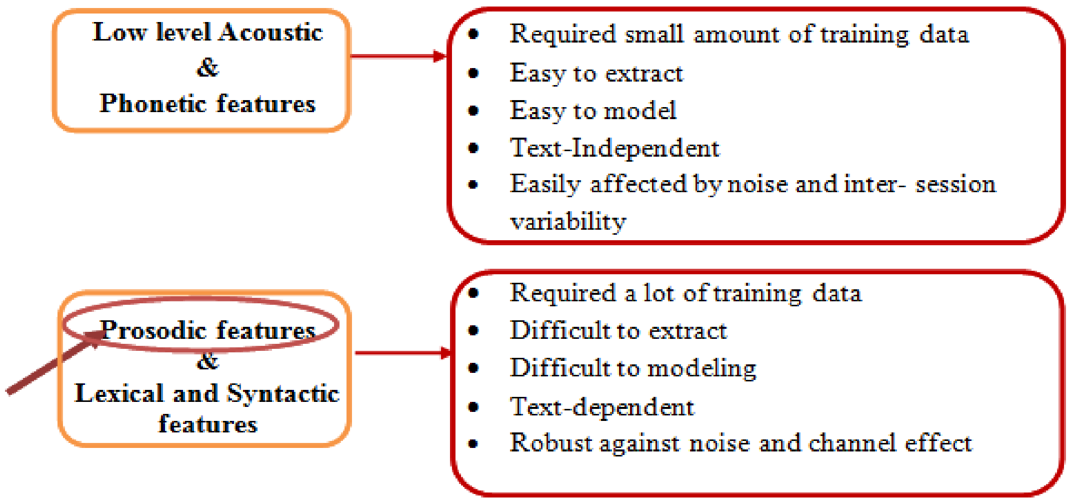


Figure- 1: Characteristics of Different Types of Speech Features

Research in speaker recognition is mainly concerned on the development of fast and robust system, which can work in noisy environment as well as channel mismatch. The time required by the traditional classifier system has to be reduced to make the identification system more useful. If classifier GMM is used, then computational complexity can be reduced by two ways; by reducing the number of mixtures or by reducing the dimension of feature vectors. But in case of reducing the dimension of feature vectors speaker identification rate decreases and this results in making the identification system impractical for using. Therefore, another option is to reduce dimension of feature vector, without losing important information contained in the speech features.

System performance depends on feature extraction technique such as MFCC, LPCC, LPC, Prosodic etc. Commonly used feature extraction technique is MFCC but it is not suitable in the case of noisy data while prosodic features are robust against the noisy data [7] [12-13]. Many researchers have come up with different prosodic features for improving system performance. Hence, the main motivation behind the proposed research is to analyse the different prosodic features available and to find out the way of improving system performance.

3. Identified Issues in the Speaker Recognition System

During the implementation of the speaker recognition system, many problems occur. Some specific one is discussed here such as:

- **Scalability**

Time of identification in speaker recognition system increases with the increase in the number of speakers in the voice database. Hence performance of recognition decreases with respect to increase in speaker models [14].

- **Channel divergence problem**

Channel mismatch problem arises due to differences occur during acquisition of training and testing data.

- **Time complexity:**

The efficiency of automatic speaker recognition system is decided by the time taken by the system during testing process. To achieve better accuracy, higher dimensional feature vectors are needed which again add higher computational complexity and increase computation time. Therefore there is a trade-off between computational time complexity and speaker identification rate [4] [12] [15].

- **Performance against noisy speech signal**

Practically it is next to impossible to capture a noise less speech, so a system which is robust against noise is required to develop.

4. Research Problem

From the foregoing discussion, it is pertinent that security is a big challenge in the current digital era where insecurity is everywhere. The potential of speaker recognition technology is that it relies on a signal (voice) which is natural and available unobtrusively to acquire without any special equipment or training. The primary use of this technology is for remote system accessibility and forensics. Also it is easy to use and portable (portable as handhelds device) and the leading factor is high accuracy. Keeping this in mind, the researcher has formulated a problem as under in order to improve the accuracy of speaker recognition system;

Improving Performance of Speaker Recognition Using Prosodic Features

5. Objective of the Research

In order to achieve the most general goal to improve the performance of speaker recognition system using Prosodic features and Gaussian Mixture Model (GMM), the following objectives have been set fourth:

- To review and critically examine the speaker recognition technology and to identify key issues that needs to be addressed in the real-world deployment of this technology.
- To study about the ways for improving speaker recognition performance and noise robustness for real-world operational conditions,
- To study about the alternatives to the present state-of-the-art approaches for speaker recognition, and to identify the ones that offer practical advantages,
- To demonstrate the operation of the speaker recognition technology in real-life or close to real-life scenarios.
- To find out the better modeling technique and pattern matching technique.
- To conduct a detailed study on Prosodic features extraction technique.

- To design an algorithm to extract the features of speaker's voice and to create a database for the same.
- To design and develop a speaker recognition system with better performance.
- To test the proposed system by analyzing the words spoken by a speaker to verify whether the speaker is a true speaker or not (i.e. speaker verification) and to identify a particular speaker among a group of persons by accepting or rejecting.
- To evaluate the performance of the proposed system.
- To compare the performance of the proposed speaker recognition system with the other recognition system in the area.
- To validate the proposed system.

6. Research Contribution

In the research, some basic but important factors which are a crucial part of the Automatic Speaker recognition (ASR) System such as Voice fundamental, sampling rate, pitch, equal error rate etc. have been discussed. These factors are used to design and construct an Automatic Speaker Recognition System. ASR is a technology by which a person can be recognized by his/ her voice using a recognition system where different comparisons are made with training and testing data. This technology is based on the voice features of a person which is most suitable to recognize a person or authenticating the person. ASR system can be designed for Speaker Verification & Speaker Identification. Speaker Recognition/ voiceprint recognition is an example of Biometrics i.e. it is a type of technology using which person's physiological/ behavioral characteristics can be measured.

The main focus of the research has explained about the feature extraction techniques. Speaker Recognition System makes use of a system based on comparison method is use to recognize the people from his/her voice. Prosodic features analysis of a speech signal has been elaborated in detail. In addition Gaussian mixture model generally used modeling technique explained in detail. The research has provided an elaborated Gaussian Mixture Model (GMM) and its component for speech signal. It has been revealed that Gaussian Mixture Model is very much appropriate for creating speech features model of a speaker. For speaker recognition, Gaussian mixture model is an essential appliance of statistical clustering. The task effortlessly performed by humans is not effortless for machine or computers such as voice recognition or face recognition. Speaker recognition technology produces a solution, using which the computers/machines out performs than humans. For achieving the research objective of speaker recognition, the research has made the following contributions.

- **Literature survey on speaker recognition:** An exhaustive review has been done on speaker recognition techniques during the initial phases of research work. As the outcome of literature survey it has been concluded that speaker recognition primary focused on the speaker identification, text-dependent, text independent etc. based speaker recognition system. The contribution has been published in [16, 17, 18]. In the last several decades research in speech and speaker recognition has been going on worldwide. Since speech is the basic and the most suitable form of communication to convey message among people. The development of speaker recognition system shows the interest of human in technology. It is as the first step towards natural human machine communication. Some considerable advances in speech along with speaker recognition are not likely to come solely from research in statistical pattern recognition and signal processing. Number of practical limitations has been encountered which hinder widespread deployment of application and services.

- **Analysis of various tools/models/techniques used at various level of Speaker Recognition System development:** Speaker recognition using prosodic and Gaussian mixture models as a technique to provide efficient and accurate solutions to problems of speaker recognition. A review is about the prosodic and about the GMM for speaker recognition technology has been performed. The studies show that mixture model analysis yields a large number of theorems, methods, applications and test procedures. There is much related theoretical work as well as research is available on Gaussian mixture applications. The research has explored the same only for automatic speaker recognition technology. The Detailed description is presented in chapter 2.

- **Voice Database creation:** Database of recorded male and female speakers from different channel has been developed. To build a voice database there are lots of medium available for recording the voice/speech. In the research, we have recorded the speech by using head phone, MATLAB program and some external/portable voice recording devices. To record voice a head phone or any other medium of recording is needed using which voice of speaker is recorded. In addition to Computer/Laptop, there is also some external audio/video devices required to store voice database. At the time of recording the speaker's speech of 2-3 minutes length has been recorded sometimes it is vary up to 5 minutes. Detailed description of the same is presented in chapter 4.

- **Development of framework for speaker recognition system:** A new approach is developed for speaker identification system. The developed framework aims to extract speech features using Prosodic feature extraction technique. Prosodic features achieve better recognition rate by considering supplementary information sources. The aim is to design a speaker identification system, and apply it to a speech of an unknown speaker (text-independent). By investigating the extracted features of the unknown speech and comparing them to the stored speaker models for each different speaker, an effort is made to identify the unknown speaker. Prosodic showed a significant improvement in automatic speaker recognition system performance. The Detailed description is presented in chapter 3.

- **Implementation of the Proposed Framework for Improving Speaker Recognition System Performance:** The proposed framework is implemented in chapter 4 and all the phases of the framework are discussed in detail. Each and every phase of the proposed framework has been implemented and tested using MATLAB as well as Praat software. Speaker models are created by using Gaussian mixture modeling technique, and stored for training and testing purpose. After creation of speaker models matching is performed and on the basis of match score decision is made that either speaker is accepted or rejected. In addition system performance is evaluated on the basis of equal error rate metric. The detailed description is shown in chapter 4.

- **Validation of Prosodic feature extraction technique:** It was validated that features based on Prosodic provided better performance compare to the conventional MFCC. For validation of the proposed approach an experimental tryout has been carried out. For voice sample collection first we created a voice database for speaker's (male & female) after that feature extracted from speech signal using different feature extraction methods such as MFCC, Prosodic .Using feature extraction method performance of a speaker recognition system was calculated. The Detailed description is presented in chapter 6.

7. Methodology for Validation

Validation is a procedure to estimate the comparison between computational results from the simulation and the actual (hypothesis) data from the system. Primary goal of validation is identification and quantification of the error, uncertainty in the conceptual models, and calculation of the numerical error in the computational solution, evaluation of the simulation uncertainty, and at last, comparison between the computational results and the actual data. Therefore, accuracy is calculated in terms to real/hypothetical data. But this approach does not believe that the real/hypothetical data are more accurate than the computational results [19-20].

For validation of the proposed approach an experimental tryout has been carried out. For data collection first we created a voice database for speaker's (male & female) after that feature extracted from speech signal using different feature extraction methods such as MFCC and Prosodic. Using feature extraction method accuracy of a speaker recognition system was calculated. Calculated values of recognition rate using MFCC, Prosodic have been given in table- 1. Figure- 2 is the graphical representation of recognition rate for MFCC and prosodic.

Training Language	**Testing Language**	**Recognition Rate (%) MFCC**	**Recognition Rate (%) Prosodic**
English	English	89.70	95.74
	Hindi	86.50	94.61
Hindi	English	86.50	94.61
	Hindi	89.70	95.74

Table- 1: Calculated values for MFCC and Prosodic

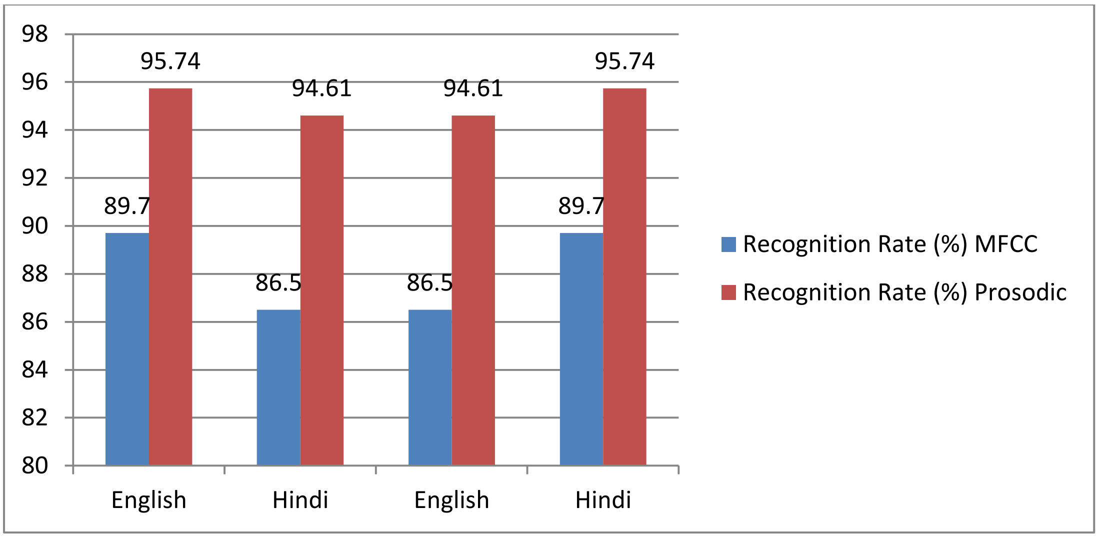


Figure- 2: Comparison Performance of Automatic Speaker Recognition System using Prosodic and MFCC

8. Hypothesis Testing

A null hypothesis reflects that there is no significant relationship between two or more parameters whereas alternate hypothesis affirms the relationship. Rejection of a null hypothesis provides a stronger base to accept the relationship or to accept the alternate hypothesis [21]. This study relates improvement of speaker recognition system performance by using prosodic features. During the research work a framework has been developed by using Prosodic features of speech and modeling by Gaussian mixture model.

- ✓ **Null Hypothesis (H_{01}):** Performance of speaker recognition system cannot be improved by using Prosodic.
- ✓ **Alternative Hypothesis (H_{11}):** Performance of speaker recognition system can be improved by using Prosodic.
- ➢ **Interpretation**

By observing calculated values in table- 2, it can be observed very easily that the prosodic features equal error rate shows that system performance is improved. The evaluated values for the recognition rate of Prosodic and MFCC are shown in table- 1. The result shows that table-1 is shown that prosodic have better performance than MFCC. Experimental values show that methodology followed for improving system performance is appropriate. Hence it has been conclude that performance of recognition system could be improved. Hence the initial claim that prosodic feature is able to improve speaker recognition system performance proved true. Table- 2 shows the calculated values of EER for speaker recognition system. Figure- 3 shows the EER of MFCC and Prosodic

Training Language	Testing Language	EER (MFCC)	EER (Prosodic)
English	English	10.3	4.26
	Hindi	13.5	5.39
Hindi	English	13.5	5.39
	Hindi	10.3	4.26

Table-2: Calculated values of EER for Speaker Recognition

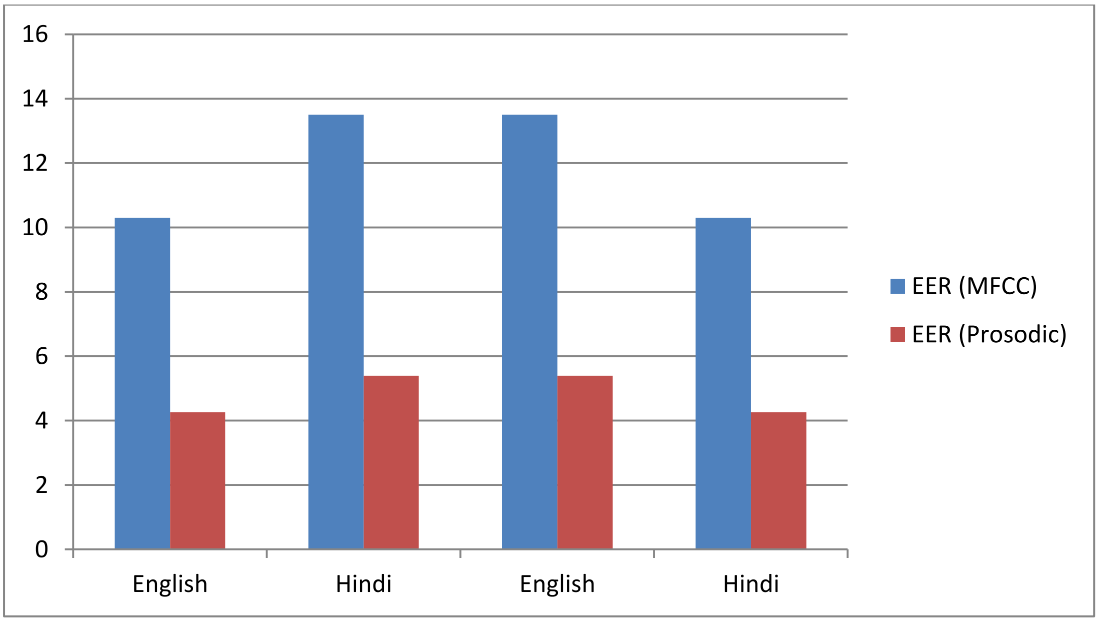


Figure- 3: EER of MFCC and Prosodic

- **Level of Significance of the proposed Framework**

To find out the significance difference between MFCC and Prosodic; the means of MFCC and Prosodic are calculated as shown in table-3. Pearson coefficient of correlation is 1. The degree of freedom is 4. For application of the t-test in the scenario, homogeneity of variances i.e. F value must be tested. The homogeneity can be obtained by dividing the larger variance by the lower. The large variance is 3.4134 for MFCC and the smaller one is 0.4256 for Prosodic.

The t value comes out to be -7.2218. As the value exceeds the t critical value of 0.0019 for a two tail test at the 0.01 level for 4 degree of freedom, the null hypothesis H_{01} is strongly rejected and the alternate hypothesis H_{11} is accepted. Hence it is validated that performance of speaker recognition system can be improved by using Prosodic features.

- **T-Test: Two-Sample Assuming Unequal Variances**

Statistical Observation	Prosodic	MFCC
Mean	4.825	11.9
Standard deviation	0.565	1.6
Variance	0.425633333	3.413333333
Observations	4	4
Pearson Coef. of Correlation(r)		1
Hypothesized Mean Difference	0	
Degree of Freedom (df)		4
t Stat		-7.221863395
P(T<=t) one-tail		0.000974887
t Critical one-tail		2.131846786
P(T<=t) two-tail		0.001949775
t Critical two-tail		2.776445105
Test for homogeneity of variances(F)		0.124697265624998

Table-3: T-Test: Two-Sample Assuming Unequal Variances

Acceptance of any new approach by society or industry depends upon validation of that approach. It is the validation which proves the usefulness of the approach in society or in industry. For testing the usefulness of the integrated approach for recognition improvement rate of an Automatic Speaker Recognition System, a systematic validation is carried out.

9. Significance of the Work

Automatic speaker recognition is used for the purpose of authentication to improve security of an automatic system. In today's scenario it is more useful because of its voice based biometric technique. The world-wide collaboration and free exchange of ideas led to technology boost in the speaker recognition field as well. Applications of speaker recognition system provide prominent alternatives to biometrics such as finger prints, retina scans and face recognition. The key advantage of speaker recognition over these techniques is being its low costs, not harmful to human body and non-invasive etc. The proposed technique can be successfully implemented in the following areas:

- Forensics Department
- Remote access control security
- Web services, online Calling
- Personalization of services and customer relationship management
- Voice based biometric system
- Transactions authentication/Voice based banking
- Surveillance/criminal investigation etc.

10. Future Direction

In future work we wish to work with many more prosodic features with the fusion of some other types of speech features. Also we will try to work with prosodic features with another modelling technique. To till date many recent advances have been achieved in the field of speaker recognition but there are still many problems remains unsolved for which good solutions need to be found. These problems mainly arise from speaker variability, channel variability, recording condition, background noise etc. Also it is key point to find feature parameters that are more useful to improve performance of speaker recognition system. The aim of finding such speech features that are stable over time, unaffected to the variation of speaking manner (including speaking rate and level) and robust against variation against voice quality such as voice disguise or cold, is a very important task. There is also need to develop a method to deal with the problem of distortion such as telephone sets and channel and background noises. In addition, we will try to find out the reason behind rejected voice (unmatched voice), i.e. it is occurs in due to testing database or training database.

11. Limitations

In order to keep the research precise and within the time boundary, the thesis has few limitations. These are as follows:

- The proposed work focuses only on speaker Identification and not on speaker verification.
- Voice database is limited and it also contains some background noise.
- The performance of proposed system is tested only within laboratory setup. Implementation of the same has not been done in real scenario.
- Training and testing is performed only on English and Hindi voice database.
- During testing, channel mismatch has not been considered.

12. Thesis Outline

It is expected that the proposed research will make the speaker identification task more accurate. The thesis presents detailed study about the same. Apart from annexure, references and other components, this study includes seven chapters. A summary of each chapter is presented below.

Chapter- 2: Literature Review

This chapter provides concise definition and discussion about speaker recognition technology. It presents the literature review, basic terminology of speaker recognition and speaker recognition methodology in details. It also presents the general overview of human speech production, and consequently introduces the review literature of speaker modeling technology and the estimated model.

Chapter- 3: Framework for Speaker Recognition

In this chapter, a framework for development of speaker recognition system is proposed along with the premises of framework. A guideline is given for the proposed framework. The proposed framework has the six phases including acquiring speech signal, feature extraction, modeling, pattern matching and decision phase and performance evaluation phase. In addition, limitation of the proposed framework has also been discussed.

Chapter- 4: Implementation of the Proposed Framework Using Prosodic Features

In this chapter, a detailed description is presented on how the proposed framework for speaker identification system is implemented. A detailed study of prosodic features with their extraction procedure is also presented. Most of the components of this system are implemented individually by MATLAB as well as Praat software as per the suitability of framework component. Database creation process is also given in detail. Voice database is created for training and testing of speaker recognition. At the time of voice recording some background noise exists. In addition, discussions about prosodic features which are used in this research along with an algorithm for speaker identification are presented.

Chapter- 5: Experiments and Result

In this chapter, the experimental results are presented. During experiment training condition, enrollment condition and test conditions are discussed. Results obtained from the individual feature extraction techniques have also been discussed. This chapter deals with the results obtained in the study and its data analysis. Calculated results are presented graphically. Comparison of accuracy results of speaker recognition systems using different feature extraction methods are calculated and compared. It has also shown that accuracy of automatic speaker recognition system depends on various factors including which feature extraction technique is used, what speaker modeling technique is used, recording conditions of speech etc.

Chapter- 6: Validation of the Framework

In this chapter, concept of validation has been discussed and methodology for validation has been designed. During validation, hypothesis have been formulated and tested on the basis of statistical analysis. Student t-test has been used for testing the hypothesis. Additionally, some issues have been discussed which occur at the time of development of speaker recognition system. The main issues including noise, headset mismatch, sampling rate of speech signal etc. are discussed here. In addition, the ways to improve speaker recognition system performance have been described.

Chapter- 7: Conclusion and Future Work

This chapter describes the findings of this research. It presents an overview of the research and outlines in terms of its major findings. In addition, it demonstrates the significant contribution of this research in reference to speaker recognition technology. It also discusses probable limitations of the research and proposes directions for future research.

References

1. U. Sandouk, “Speaker Recognition Speaker Diarization and Identification”, PhD Thesis, University of Manchester, School of Computer Science, pp. 14-101, 2012.
2. M. Savvides, “Introduction to Biometric Technologies and Applications”, ECE & CyLab, Carnegie Mellon University. http://www.biometricscatalog.org/biometrics/biometrics_101.pdf or biometrics_101.pdf
3. B. S. Atal, “Automatic Recognition of Speakers from their Voices”, Proc. IEEE, vol. 64(4), pp. 460-475, 1976.
4. Q. Jin, “Robust Speaker Recognition”, PhD Thesis, Language Technologies Institute School of Computer Science Carnegie Mellon University, Pittsburgh, pp. 23-177, 2007.
5. A. Rajsekhar G., “Real Time Speaker Recognition using MFCC and VQ”, PhD Thesis, Department of Electronics & Communication Engineering, National Institute of Technology Rourkela, pp. 9-71, 2008.
6. J. Luettin, “Visual Speech and Speaker Recognition”, PhD Thesis, Department of Computer Science University of Sheffield, pp. 16-156, May 1997.
7. T. Kinnunen, “Spectral Features for Automatic Text-Independent Speaker Recognition”, Licentiate’s Thesis, University of Joensuu, Department of Computer Science Joensuu, Finland, pp. pp. 1-151, December 21, 2003.
8. M. R. Srikanth, “Speaker Verification and Keyword Spotting Systems for Forensic Applications”, PhD Thesis, Department of Computer Science and Engineering Indian Institute of Technology Madras, pp. 1-135, Dec. 2013.
9. M. El. Ayadi, Abdel-Karim S.O. Hassan, Ahmed Abdel-Naby, Omar A. Elgendy, “Text-independent Speaker Identification using Robust Statistics Estimation”, Speech Communication vol-92, pp. 52–63, 2017.
10. S. Sarkar, Sreenivasa, Rao, K., “Stochastic Feature Compensation Methods for Speaker Verification in Noisy Environments” Appl. Soft Comput. 19, pp. 198–214, 2014.
11. G. Doddington, Liggett, W., Martin, A., Przybocki, M., and Reynolds, D, “Sheeps, Goats, Lambs and Wolves: A Statistical Analysis of Speaker Performance in the NIST 1998 Speaker Recognition Evaluation”, In Proc. Int. Conf. on Spoken Language Processing (ICSLP 1998) (Sydney, Australia), 1998.
12. S. Furui, “Speaker-Dependent-Feature Extraction, Recognition and Processing Techniques”, ESCA Workshop on Speaker Characterization in Speech Technology Edinburgh, Scotland, UK, pp. 10-27, June 26-28, 1990.

13. B. Wubishet , “ Noise Robust Speaker Verification using SVM based GMM Supervector”, PhD Thesis, Addis Ababa University, School of Graduate Studies Addis Ababa Institute of Technology, Electrical and Computer Engineering Department, pp. 1-102, Dec. 2012.
14. S. Patra, “Robust Speaker Identification System”, PhD Thesis, Super Computer Education and Research Centre, Indian Institute of Science, Bangalore, pp. 1-131, Dec. 2007
15. S. Furui, “50 years of Progress in Speech and Speaker Recognition”, Department of Computer Science, Tokyo Institute of Technology, pp. 1-9, 2004.
16. N. Singh and Khan R. A., “Underlying of Text Independent Speaker Recognition”, in IEEE Conference (ID: 37465) (10th INDIACom 2016 International Conference on Computing for Sustainable Global Development), held on 16th -18th March, 2016 at BVICAM, New Delhi, pp. 11-15, 2016.
17. B. Martínez-Gonzalez, M. Jose Pardo, D. Julian Echeverry-Correa, Ruben San-Segundo, “Spatial Features Selection for Unsupervised Speaker Segmentation and Clustering”, Expert Systems with Applications, vol. 73, pp. 27–42, 2017.
18. T. Kinnunen and Haizhou Li, “An Overview of Text-Independent Speaker Recognition: From Features to Supervectors”, Speech Communication vol.52, pp.12–40, 2010.
19. J.P.C. Kleijnen, “Validation of Models: Statistical Techniques and Data Availability”, Proceedings of the 1999 Winter Simulation Conference, P. A. Farrington, H. B. Nembhard, D. T. Sturrock and G. W Evans, (eds.), pp.647-654, 1999.
20. J.D. Sterman, Business Dynamics, “Systems Thinking and Modeling for a Complex World”, IRWIN McGraw-Hill, pp.845, 2000.
21. F.K. Soong, et. al., “A Vector Quantization Approach to Speaker Recognition”, AT & T Technical Journal, vol.66 (2), pp.14-26, 1987.

List of Publication (Journal and Conferences)

1. Nilu Singh, Alka Agrawal and R. A. Khan, "Voice Biometric: A Technology For Voice Based Authentication", Advanced Science, Engineering and Medicine ISSN: 2164-6627 (print); EISSN: 2164-6635 (online) (Accepted).
2. Nilu Singh, Alka Agrawal and R. A. Khan, "Automatic Speaker Recognition: Current Approaches and Progress in Last Six Decades", Global Journal of Enterprise Information System. Volume-9, Issue-3, July-September, pp. 38-45, ISSN: 0975-1432, 2017.
3. Nilu Singh, Khan R A, "Underlying of Text Independent Speaker Recognition", in IEEE Conference (ID: 37465) (10th INDIACom 2016 International Conference on Computing for Sustainable Global Development), held on 16th -18th March, 2016 at BVICAM, New Delhi, pp. 11-15.
4. Nilu singh & R. A. Khan "Extraction and Representation of Prosodic Features for Automatic Speaker Recognition Technology", Fifth International Conference on AITMC (AIM-2015), Proceedings of Advanced in Engineering and Technology, Published by: Mc Graw Hill Education, pp.1-7, 2016. ISBN-10:93-85965-79-4.
5. Nilu Singh, Alka Agrawal, R. A. Khan, "Gaussian Mixture Model: A Better Modeling Technique for Speaker Recognition", Proceedings of (DIAL-2017) National Conference on Digital India-Altering Landscape, Tracking the Journey to Digitally Empowered New India, 9th -10th Dec 2017, Organized by CSI Lucknow Chapter, pp. 14-18, 2017. ISBN: 978-93-5291-226-1.
6. Nilu Singh, Alka Agrawal, R. A. Khan, "Speaker Recognition: A digital Evidence for Authentication", International Seminar on "Modern Trends in Engineering and Sciences", Organized by Dept. of ECE, Physics & EE under the Aegis of HRDC, Integral University, Lucknow, 03 Aug 2017.
7. Nilu Singh, Alka Agrawal, R. A. Khan, "Speaker Recognition: An Evolution", National Conference on Emerging Trends in High Performance Computing, Organized by Feroze Gandhi Institute of Engineering & Technology, Raebarely (UP), 1st -2nd May 2017.
8. Nilu Singh, Alka Agrawal, R. A. Khan, "Improving Speaker Identification Using Gaussian Mixture Model", 4th Lucknow Science Congress (LUSCON-2017), 3-4 March 2017, organized by Babasaheb Bhimrao Ambedkar University, Lucknow
9. Nilu Singh, Alka Agrawal, Raees Ahmad Khan, Presented a paper on "Text Independent Speaker Identification System and Challenges" in National Seminar on Emerging Trends & Advancements in Cyber Security –April-2016, at Integral University, Lucknow.