# A Survey on Single and Multi-Speaker Voice Recognition In Android

Mr. Tanmay Naskar[1], Mr. Anket Sah[2], Mr. Akshay Mense[3], Ms. Neha Patekar[4], Ms. Nikhatjahan Kankurti[5], Dr. Mrs. Neeta Deshpande[6], Mrs. Suvarna Kadam[7]

[1, 2,3,4,5] *BE Scholar, Department of Computer Engineering, D Y Patil College of Engineering, Akurdi, Pune, MH, India*

[6,7] *Asst. Professor, Department of Computer Engineering, D Y Patil College of Engineering, Akurdi, Pune, MH, India*

**Abstract** – *For communication voice is the most preferred natural form and the foremost goal of developing this system is to make interaction with machines as easy as possible. Voice recognition and speech recognition are considered same however, they have distinct meanings. Speech recognition detects only the words, while the voice recognition disregards the language or anything linguistic for that matter and detects the identity of the person speaking. In this given paper, we discuss two parts, In the first part of the paper, we discuss the means of processing the speech signal using the Mel Frequency Cepstral Coefficient and the basics of voice characteristics. At hardware level the problem of voice recognition had been already implemented, but here we propose a solution with higher level of accuracy on the software end. Moreover, both single speaker and multi-speaker single recognition application being developed on android is discussed in the second part. The multispeaker feature enables the system to identify more than one person speaking by splitting the input into smaller blocks and treating each block as a single speaker identification problem. For the case of single speaker recognition, the algorithm is able to correctly identify the individual 90.3% of the time. In the multispeaker recognition case, up to 3 individuals are correctly identified roughly 40% of the time. For a database consisting of 415 samples, the average time for identification is 220ms.*

**Keywords -** *Voice biometric, Voice Recognition, Vice identification, Multi Speakers, MFCC, LPCC, VAD, Smartphone Application.*

## I. INTRODUCTION

Voice recognition is a multi-disciplinary technology which uses vocal attributes of a speaker's voice and help us to of identifying or confirming the identity of an individual based on his voice. Speaker Recognition is a branch of biometrics that is used for identifying, verifying and classifying individual or multiple speakers. Voice recognition and speech recognition are often used interchangeably and considered as same; however, they have distinct meanings. Speech recognition detects only the words, whereas the voice recognition disregards the language or anything linguistic for that matter and detects the identity of the person speaking. Hence voice biometrics would use the voice recognition. Enormous capacity of information can be carried by a voice signal. Voice recognition system which involves voice identification as well as speaker verification, is based on the fact that a person's speech reflects its several unique attributes. Voice signal can be transmitted over long distances via telephone media and even seen as a non-evasive biometric that can be collected with or without the knowledge of person. A speaker's voice cannot be lost, stolen or at most forgotten unlike different forms of identification, such as a password or a key. A secure method of authenticating and identifying speakers is allowed by speaker recognition application.

While there are several different biometric recognition systems available like fingerprint recognition system, retinal scans, face recognition etc. these are more dependable methods used across security and access control system for identification nowadays. In near future it's expected that voice recognition will be used widely to make it feasible, the use of speaker's voice for verifying and validating their identity and for controlling access to multiple services such as voice mail, information services, voice dial, telephone shopping, banking by telephone, security control for confidential information areas or providing remote access of computers. The work of RnD (research and development) on voice recognition method and related techniques has been done for more than six decades until now and still it continues to be an active area. For representing the voice signal different feature matching and extraction methods are developed during this span of six decades. The two most popular algorithms of voice feature extraction presented are: Mel Frequency Cepstral Coefficients (MFCC) and Linear Predictive Cepstral Coefficients (LPCC). After a test was done on 415 different voices with users from both genders, it was proven that the MFCC is more efficient than the LPCC and can reach up to an accuracy rate of 90% whereas the LPCC would hit 75%. Therefore, using MFCC for feature extraction proved to be the best solution. Voice Activity Detection or VAD is also important for the system. It needs to know whether a person is

currently speaking, or if it's just noise. On the basis of speech attributes, differentiation of voiced signal into silence and speech is done. Slicing of signal is done into contiguous frames. With each frame a non-negative and real-valued parameter is associated. This parameter is average energy content along with number of Zero Crossings of the frame for the time-domain algorithms. This parameter is spectrum as well as variance of the spectrum of a frame for the frequency-domain algorithms. The signal frame is classified as ACTIVE if this parameter exceeds a certain threshold, otherwise, it is classified as INACTIVE.

## II.    LITERATURE REVIEW

As nowadays, both physical and data security has become a major concern, new methods and approaches are being used for recognizing and uniquely identifying the person. Some major techniques has being developed over the years such as Hidden Markov Model (HMM) for pattern matching, Dynamic Time Warping (DTW) for matching given two temporal sequences, Support Vector Machine(SVM) for analysing data used for classification purpose.

Even though speaker recognition may seem to be a well-known topic, it's still in its early stage of development .The following table below shows the different approaches used for speaker identification over the past few years:

| Sr. No | Paper | Key Features | This paper refer to: |
|---|---|---|---|
| 1 | Intelligent Voice Assistant Using Android Platform[1] | • Focus on the Android development over voice control and analysing corresponding commands automatically.<br>• Google products and relevant APIs (Google weather and search)<br>• Wikipedia API and mobile device references ranging from Speech-To-Text, Text-To-Speech technology<br>• Camera, messaging and other technology needed in daily life. | • HMM (Hidden Markov Model)<br>• TTS (Text to Speech)<br>• Intent Android OS |
| 2 | Speech Recognition in Marathi Language on Android O.S[2] | • To study different methods for feature extraction and comparison techniques for ASR systems<br>• Selecting the most appropriate technique for improving the accuracy of the system.<br>• An experimental database of phonemes is collected for speaker dependent system. | • Acoustic Model, Dynamic Time Warping (DTW)<br>• Language Model, Mel Frequency Cepstral Coefficient<br>• (MFCC), Phoneme recognition, Speech recognition. |
| 3 | Speech to Text Conversion using Android Platform[3] | • Developed an on-line speech-to-text engine.<br>• The system acquires speech at run time through a microphone and processes the sampled speech to recognize the uttered text.<br>• The recognized text can be stored in a file. | • HMM, Android OS, DVM, Speech Recognition, Intents |
| 4 | Smart Voice Search Engine[4] | • Share a domain independent search engine with smart search which searches user speech automatically without the user's request and provide him with evidence on his speech, this engine was called SVSE. | • Speech Recognition<br>• Speaker Recognition<br>• Search Engine |

| 5 | Voice Recognition Based Secure Android Model For Inputting Smear Test Results[5] | • The voice recognition technology is applied into a laboratory information system for identifying each technician's voice. i.e. By using the user's voice sample<br>• a secure authentication system is developed where the unique features of the user's voice are extracted and stored at the time of registration. | • Voice Recognition<br>• Speaker Identification<br>• Speech Recognition, Smear Test<br>• Fast Fourier Transform. |
|---|---|---|---|
| 6 | An Outdoor Navigation with Voice Recognition Security Application for Visually Impaired People[6] | • A research of a navigation system for blind people is written in order to provide more precise location information.<br>• To identify the position and orientation and location of the blind person any of those solutions rely on Global Positioning System (GPS) technology | • TTS<br>• Visually Impaired Persons<br>• Navigation System. |
| 7 | Voice To Text Transcription Of Word Recording [7] | • The benefits of same-language transcription of media content and goes on to outline the details of a technical feasibility study. | • Transcription, Universal Design for Learning (UDL)<br>• Same Language Subtitles (SLS) |
| 8 | Gender Recognition System Using Speech Signal [8] | • The digital processing of a speech signals (pronounced "A" and "B") which are taken from 10 persons, 5 of them are Male and the rest of them are Female.<br>• Power Spectrum Estimation of the signal is examined | • Gender Recognition, Feature Extraction<br>• First Fourier Transform (FFT), Font-end, Back-end. |
| 9 | Robotic Control using Speech Recognition and Android[9] | • Use practical approach of robotics through a popular platform android and the speech recognition method Mel Frequency Cepstral Coefficients (MFCC).<br>• It gives the industry an optimized method for basking in information regarding temperature, humidity, gas leakage in challenging surroundings and provides security with voice authentication. | • MFCC, Android, Bluetooth, Cepstrum<br>• Smartphone, RF module, Sensors<br>• Speech Recognition, Linde-Buzo-Gray,<br>• Fourier Transform |
| 10 | Software Independent Speech Recognition System[10] | • Introduces the basics of speech recognition and speech to text technology and also highlights the difference between different speech recognition systems.<br>• The most common algorithms and basic block diagram which are used to do speech recognition are also discussed along with the current and its future use. | • Speech Recognition<br>• Microphone, Voice, |

**Table1: Literature Survey**

## III.     APPROACHES FOR SPEAKER IDENTIFICATION

Different ways in which the speaker identification techniques can be classified are as follows:
1.  **Text Dependent Vs. Text Independent**:
    Text Dependent system involves same text to be spoken both during the training and testing phase while, in Text Independent system there is no any restriction on the text to be spoken.

2. **Identification Vs. Verification**:
Identification technique determines from the set of known registered speakers, whereas in Verification performs accepting or rejecting the identity claims of the input.

Mel Frequency Cepstral Coefficients algorithm is extremely effective for voice feature extraction. After a test was done using The Mel Frequency Cepstral Coefficients (MFCC) and Linear Predictive Cepstral Coefficients (LPCC) on 415 different voices with users from both genders, it was proven that the MFCC is more efficient than the LPCC and can reach up to an accuracy rate of 90% whereas the LPCC would hit 75%. Therefore, using MFCC for feature extraction proved to be the best solution. The distance between the recorded voice print and the saved voice print is computed and compared with the Universal Voice print in order to obtain a likelihood ratio. The voice print with the highest likelihood ratio is the identified individual.

Following are the advantages of the speaker recognition system:
1. Provide higher level of accuracy.
2. The multi-speaker feature implementation for the system.
3. Techniques that help to improve its efficiency.

## IV.     CONCLUSION

This paper shows that the speaker recognition application can be accurately adapted to work on a software, specifically on an Android mobile phone. The results obtained already look promising and with proper care, a voice print can be used as an additional security measure on our mobile phones, in addition to pin locking and finger-print locking. VRS which capable of detect voice and recognize it to know if its sound of clap or ringing sound or Whistle sound. VRS enables user some option to get alarm when voice detect it can send email with alarm type or by SMS to mobile number VRS enables user to choose song and run it when voice is detected and system can choose song automatically. VRS enables user to choose how much second mobile will vibrate when sound detected. Finally, VRS is flexible, effective, and attractive system. The proposed system expansion by introducing new modules and implementing new interfaces is undemanding and hassle-free.

## V.     REFERENCES

[1]     Sutar Shekhar "Intelligent Voice Assistant Using Android Platform", International Journal of Advance Research in Computer Science and Management Studies, Volume 3, Issue 3, March 2015.

[2]     Saroj B. Jadhav, Jayshree Ghorphade and Rishikesh Yeolekar "Speech Recognition in Marathi Language on Android O.S", International Journal of Research in Computer and Communication Technology, Vol 3, Issue 8, August – 2014.

[3]     B. Raghavendhar Reddy, "Speech to Text Conversion using Android Platform", International Journal of Engineering Research and Applications (IJERA), Vol. 3, Issue 1, January -February 2013, pp.253-258.

[4]     Shahenda Sarhan, "Smart Voice Search Engine", International Journal of Computer Applications (0975 – 8887) Volume 90 – No.3, March 2014.

[5]     Teenu Therese Paul, "Voice Recognition Based Secure Android Model For Inputting Smear Test Results", International Journal of Engineering Sciences & Emerging Technologies, Dec. 2013.

[6]     Nandish M S, "An Outdoor Navigation with Voice Recognition Security Application for Visually Impaired People", International Journal of Engineering Trends and Technology (IJETT) – Volume 10 Number 10 - Apr 2014.

[7]     Dr. Stuart Dinmore, "Voice to Text Transcription of Word Recording", show me the learning, Nov-27-30.

[8]     Md. Sadek Ali, "Gender Recognition System Using Speech Signal", International Journal of Computer Science, Engineering and Information Technology (IJCSEIT), Vol.2, No.1, February 2012.

[9]  Gaurav Chauhan, Prasad Chaudhari, "Robotic Control using Speech Recognition and Android", International Journal of Engineering Research and General Science Volume 3, Issue 1, January-February, 2015.

[10]  Pranali Yawle, "Software Independent Speech Recognition System", International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering, Vol. 5, Issue 3, March 2016.