IPASJ International Journal of Electronics & Communication (IIJEC)
Web Site: http://www.ipasj.org/IIJEC/IIJEC.htm
A Publisher for Research Motivation ..... Email:editoriijec@ipasj.org
Volume 5, Issue 11, November 2017 ISSN 2321-5984

# A Critical Review on Automatic Speaker Recognition

**Ecclestone, Bernie**

University of Minnesota

## ABSTRACT

***Automatic Speaker Recognition (ASR) is use to recognizing persons from their voice. Since the voice of every human is not same because their vocal tract shapes, larynx sizes and other parts of a human voice production system. Automatic Speaker recognition is a procedure to automatically recognizing a speaker or who is speaking by the individual information counted in speech signal/waves. Automatic speaker recognition technique makes it possible to use the speaker's speech to verify their identity. It have many applications for example control access to services such as voice mail, voice dialing, banking by telephone, remote access to computers, telephone shopping, information services, database access services and security control for confidential information areas.***

## 1. INTRODUCTION

Today's biometric information use to distinguish between different person's identities. Normally it may be consider that face is one most important feature to recognize someone but through biometric there are some other unique features, such as retina, voice, finger-prints, iris, are often used [1]. A different way to recognize a person is from the voice/acoustic since each person's voice are different, and this is the one area of speech processing, automatic speaker recognition [2].The fundamental research to discover innovative speaker individual features is to encode them into more affluent and more informative speaker models. Also to evaluate the effectiveness of these speech features and speaker models for speaker recognition [3]. There are many way to extract the features from speech signal such as pursuing mutually a knowledge-based, building on existing linguistic constructs and directed by perceptions from psycholinguistics and human performance studies also a more speculative datadriven approach, seeking idiosyncratic vocal performances; spectro temporal patterns with high speaker characterizing power, independent of linguistic constraints etc.[4][5][6] In recent Speech and speaker recognition have an important research and development area in biometric. The development here is the desire to create a natural form of communication between human and machine. As it is well accepted that speech is our most natural appearance of communication, ASR has the capability to impact on numerous fields of research and development [6]. Prosodic features play an important role to retrieve features from human speech. The information contained by prosodic features is different from information contained in cepstral features. Prosodic features are extracted by larger frame size while acoustical features extracted by small window. Since prosodic features exist over a long speech segment such as syllables. Prosody can be defined as it is the structure that organizes sound for example Pitch (tone), Energy (loudness) and normalized duration (rhythm) are the main components of prosody for a speaker. Prosodic features can vary from speaker to speaker and depend on long-term information of speech signal [7] [8] [9]. In this paper describes how to build a simple & complete automatic speaker recognition system. Such type of speaker recognition systems have been used in many security applications. [10]. ASR Systems can be classified as Automatic Speaker Verification (ASV) and Automatic Speaker Identification (ASI) systems. Speaker verification systems are use to verify whether an input speech signal matches to the claimed identity where as speaker identification objective to identify an input speech by selecting one speaker model from a set of enrolled speakers models. Sometimes speaker verification will follow speaker identification in order to validate the identification results [11].

## 2. DEVELOPMENT OF SPEAKER RECOGNITION

The first type of speaker recognition system comes in to the existence at 1960's, which uses spectrogram of voices for identification and this system known as voiceprint analysis. Acoustic spectrum of the speech is like to the fingerprint however such type of analysis could not fulfill the objective of automatic recognition system. In the 1980's numerous methods was proposed to extract features from speech signal for speaker recognition [6] [8]. These features are represented in time domain and frequency domain or in both domains. Result Shows those acoustic features of speech signal be different between individuals and these acoustic features consist of both learned behavioral features such as

pitch, accent and anatomy such as shape of the vocal tract and mouth [12]. The most commonly extracted features are the prosodic, Mel-Frequency Cepstral Coefficients (MFCC), Linear Predictive Coding (LPC) and Linear Predictive Cepstral Coefficient (LPCC) these features extracted to the short time analysis of speech signal which provides information on the vocal tract [13][14][15].

After feature extraction the main steps to create models for speaker's speech, different modeling techniques are available to model voiceprint extracted from speech signal. There are many models were introduced for example pattern matching i.e. Dynamic Time Warping (DTW) which does direct template matching between training and testing voice database. But study shows that direct template matching is time consuming, as the number of feature vectors increases [8]. To overcome from this clustering is a process to decrease the number of features vector through using a codebook to represent centers of the feature vectors e.g. Vector Quantization (VQ). For Vector Quantization (VQ) the LBG (Linde, Buzo and Gray) algorithm and the k-means algorithm are the most familiar algorithms. Also some other methods was proposed for speaker modeling such as neural networks (NN) and stochastic models that uses probability distribution for example Hidden Markov Model (HMM) and Gaussian Mixture Model (GMM) etc [11] [13] [16].

**Pre-Processing**

Pre-processing is a first step in automatic speaker recognition system. It is critical process performed on speech signal input in order to develop a robust and efficient speaker recognition system [14] [19]. In this process the task such as A/D, end point detection and pre-emphasis are completed. C. Feature Extraction Feature extraction from speech signal is also called front end processing and is performed in both recognition (testing) and training phase. Feature extraction use to converts digital speech signal into sets of numerical descriptors or feature vectors that hold significant characteristics of the speaker's voice [20].

**Mel Frequency**

Cepstrum coefficient (MFCC) Mel-frequency Cepstral coefficient (MFCC) is one of the most common method used for the voice feature extraction from speech signal. The basic difference among the MFC and cepstral analysis is that the MFC have frequency components with a Mel scale modeled which is based on the human ear perception of sound instead of a linear scale [21]. The Melfrequency cepstrum characterizes the short-term power spectrum of a speech signal using a linear cosine transform of the log power spectrum of a Mel scale [19].

**Speaker Modeling**

The purpose of modeling techniques is to create speaker models for feature matching of the speaker's voice. Speaker models can be defined as, it is the models that contain enriched speaker specific information with a compressed amount [14] [24]. At the time of training or enrolling phase speaker models are created using the specific features extracted from the current speaker and in the recognition phase, the speaker model is used to compare with the current speaker model for identification or verification. Gaussian Mixture Model (GMM) is generally used for speaker modeling [25].

## 3. PRACTICAL APPROACH TO ASR

The development in the field of speaker recognition technology is going on rapidly. But still there are a few inherent difficulties occurs which is needed to be solved. As the reliability of the speaker recognition system drops drastically when a huge voice database is used or when data acquire under a noisy environment. Still research is in progress and moving fast and it may be possible to improve the robustness of the existing recognition systems to solve some of the issues [14]. There are many factors which affect the speaker recognition i.e. speaker identification and speaker verification. The required Practical implications of using a speaker identification system have to be assessed are designed and implemented [6] [27].

## 4. CONCLUSION

Automatic speaker recognition technology is in full growing now days. A number of algorithms have been developed to improve the performance and robustness of Automatic Speaker Recognition systems. Automatic Speaker recognition systems are increasingly common and used in very different acoustic conditions. The use of Mel frequency cepstral coefficients (MFCCs) for feature extraction from speech signal and it is one of the standard methods used in ASR systems for information retrieval. Acknowledgement This work is sponsored by the CST-UP, Lucknow, India, under CST/D-413.

## REFERENCES

[1] Nilu Singh, "A study on speech and speaker recognition technology and its challenges." procedings of national conference on Information Security Challenges. Lucknow: DIT, BBAU, 2014. 34-37.
[2] Marcel Kockmann, Lukas Burget "Contour Modeling of Prosodic and Acoustic Features for Speaker Recognition" Speech@ FIT, Brno University, Czech Republic, pp.1-4.
[3] DOI: www.icsi.berkeley.edu/icsi/researchareas
[4] DOI: https://prezi.com/support
[5] DOI: minhdo.ece.illinois.edu/teaching/speaker_recognition/speaker _recognition.html
[6] David Michael Graeme Watts, "Speaker Identification - Prototype Development and Performance" Year 2006, pp.1-116
[7] S Furui, "50 years of progress in speech and speaker recognition research", ECTI Transactions on Computer and Information Technology, Vol. 1, No.2, November 2005.
[8] Thang Wee Keong "Voice Print Analysis For Speaker Recognition" Sim Universityschool Of Science And Technology 2009, Pp. 1-75
[9] Singh Nilu and Khan R. A. "Extraction of Prosodic Features for Speaker Recognition Technology and Voice Spectrum Analysis" International Journal of Scientific & Engineering Research (IJSER). volume 5.Issue 5 (May 2014): 600-605.
[10] http://www.ifp.uiuc.edu/~minhdo/teaching/speaker_recognitio n
[11] Utpal Bhattacharjee and Kshirod Sarmah, "SPEAKER VERIFICATION USING ACOUSTIC AND PROSODIC FEATURES" Advanced Computing: An International Journal ( ACIJ ), Vol.4, No.1, January 2013
[12] Singh Nilu, Khan R. A., and Raj Shree. "Equal Error Rate and Audio Digitization and Sampling Rate for Speaker Recognition System." American Scientific Publishers. Volume 20, .Numbers 5-6, (May 2014): pp. 1085-1088(4).