Proc. of Int. Conf. on Current Trends in Eng., Science and Technology, ICCTEST

Voiceprint Authentication System to Securely Verify and Protect Personal Identity

Shashi Ranjan[1] and Dr. Mahesh P K[2]
[1]Don Bosco Institute of Technology/ECE, Bangalore, India
Email: shashiranjanbe@gmail.com
[2]Don Bosco Institute of Technology/ECE, Bangalore, India
Email:maheshpk24@gmail.com

***Abstract*—Voice biometrics specifically was first developed in 1970, and although it has become a sophisticated security tool only in the past few years, it has been seen as a technology with great potential for much longer. Voice biometric has a history dating back some four decades and uses the acoustic features of speech that have been found to differ between individuals. The most significant difference between voice biometrics and other biometrics is that voice biometrics is the only commercial biometrics that process acoustic information. Most other biometrics is image-based. Another important difference is that most commercial voice biometrics systems are designed for use with virtually any standard telephone or on public telephone networks. The ability to work with standard telephone equipment makes it possible to support broad-based deployments of voice biometrics applications in a variety of settings. In contrast, most other biometrics requires proprietary hardware, such as the vendor's fingerprint sensor or iris-scanning equipment. By definition, voice biometrics is always linked to a particular speaker. The best-known commercialized forms of voice biometrics are Speaker Recognition. Speaker recognition is the computing task of validating a user's claimed identity using characteristics extracted from their voices.**

***Index Terms*—Biometrics, Voice Signal, Speech Recognisation.**

I. INTRODUCTION

A speaker's voice is extremely difficult to forge for biometrics comparison purposes, since a myriad of qualities are measured ranging from dialect and speaking style to pitch, spectral magnitudes, and format frequencies. The vibration of a user's vocal chords and the patterns created by the physical components resulting in human speech are as distinctive as fingerprints. Voice Recognition captures the unique characteristics, such as speed and tone and pitch , dialect etc associated with an individual's voice and creates a non-replicable voiceprint which is also known as a speaker model or template. This voiceprint which is derived through mathematical modeling of multiple voice features is nearly impossible to replicate. A voiceprint is a secure method for authenticating an individual's identity that unlike passwords or tokens cannot be stolen, duplicated or forgotten.

II. LITERATURE SURVEY

Researches on speaker recognition have been undertaken for more than 40 years and it continues to be an active area of spoken language processing. The development in speaker recognition technology is closely

Grenze ID: 02.ICCTEST.2017.1.182

concomitant with the advancement in speech and signal processing and computer technology. Speaker recognition by human was broadly studied in the 1960s. The motivation of these studies was to learn how human recognizes speakers and the reliability of human in recognizing a speaker [1] [2]. The most significant work which stimulated the further research on speaker recognition by machine was done by Kersta who introduced the spectrogram (where he noted as voiceprint) as a means of personal identification [3]. In the 1970s, attention was turned to speaker recognition by computer and came the so called automatic speaker recognition. Speaker recognition systems in this era only dealt with a small population (< 20 speakers) [3][4]. Fourier transforms, linear predictive and cepstral analysis techniques have been applied for generating feature parameters. Long time averages of these parameters were used as the speaker references. In the 1980s, more complicated statistical pattern recognition methods were investigated, for example, the Dynamic Time Warping (DTW) [5] [6] and Vector Quantization (VQ) [7] for large scale speaker recognition systems (> 100 speakers). The contribution of static and dynamic features for speaker recognition was also investigated. Since the 1990s, the availability of large scale speech database (for example, the YOHO corpus [9]) has boosted studies on more complicated models for speaker representation. These models include the stochastic models (for example, Hidden Markov Model (HMM) [10] [11], Gaussian Mixture Model (GMM) [12]), neural networks (for example, Multilayer Perceptron (MLP) [13], Radial Basis Function (RBF) [14]), and support vector machines (SVM) [15][16], so on. Among these modeling techniques, the GMM has been recognized to be the most effective in characterizing the density distribution of the speech data and has been the dominant modeling technique for speaker recognition. As for feature extraction, cepstral coefficients incorporating the auditory model, that is, the Mel-frequency cepstral coefficients (MFCC) and their dynamic coefficients have been the dominant feature parameters. Besides, various score normalization techniques have also been investigated for robust speaker recognition [17-19],[7]. A system with MFCC parameters, GMM modeling, and universal background model (UBM) for score normalization has been reported to achieve best results and have been widely accepted as the baseline for comparing new technologies [7]. To foster interaction among researchers in speaker recognition, a benchmark evaluation program has been carried out by NIST for different research communities to demonstrate their technology advancements [20]. In this practice, common test data and evaluation process are used so that different technologies and systems become comparable. For general and detailed overviews of speaker recognition, refer to [21-27][6].The captured voice may contain unwanted background noise, unvoiced sound, and there can be a device mismatch, environmental mismatch between training and testing voice data which subsequently leads to degradation in the performance of Speaker Recognition System. The process of removal of this unwanted noise, dividing sounds into voiced and unvoiced sounds and channel compensation etc for the enhancement of speech/voice is called pre-processing.

Speech Enhancement (Denoising): Numerous schemes have been proposed and implemented that perform speech enhancement under various constraints/assumptions and deal with different issues and applications [28-35]. Channel Compensation: Channel effects, are major causes of errors in speaker recognition and verification systems. The main measures to improving channel robustness of speaker recognition system are channel compensation and channel robust features [33-40]. Feature Extraction: The purpose of this module is to convert the speech waveform to some type of parametric representation (at a considerably lower information rate).The heart of any speaker recognition system is to extract speaker dependent features from the speech. They are basically categorized into two types: low level and high level features. Low level features are short range features [41-45]. Higher level features are long range features of voice that have attracted attention in automatic speaker recognition in recent years [46-49]. Speaker Model Generation: The feature vectors of speech are used to create a speaker's model/template. The recognition decision depends upon the computed distance between the reference template and the template devised from the input utterance.

III. OBJECTIVE

In today's digital world ensuring the security of our personal identity strikes everyone. Protecting the individual personally, the employee in their work surroundings and also the organization themselves to ensure financial, personal and corporate information is safeguarded against criminals is of paramount importance while financial transaction. A voice authentication service that utilizes an individual's biometric voice print to verify that person is whom they say they are. Authenticate their voice remotely during financial transaction, where strong authentication is essential.

The goal of this research is to investigate the performance of the popular recognition technique using voiceprint. As the subjects enter the scene the system should:

- Captures the voice sample
- Preprocess and analysis of voice sample
- Extraction of information of given sample
- Normalization and decision logic
- Authenticate the person.

The principal objective of the research is to study and compare existing feature extraction method and apply a hybrid solution to the present problem.

IV. METHODOLOGY

The underlying premise for speaker recognition is that each person's voice differs in pitch, tone, and volume enough to make it uniquely distinguishable. Several factors contribute to this uniqueness: size and shape of the mouth, throat, nose, and teeth, which are called the articulators and the size, shape, and tension of the vocal cords. The chance that all of these are exactly the same in any two people is low. The manner of vocalizing further distinguishes a person's speech: how the muscles are used in the lips, tongue and jaw. Speech is produced by air passing from the lungs through the throat and vocal cords, then through the articulators. Different positions of the articulators create different sounds. This produces a vocal pattern that is used in the analysis.

A visual representation of the voice can be made to help the analysis. This is called a spectrogram also known as voiceprint, voice gram, spectral waterfall, and sonogram. A spectrogram displays the time, frequency of vibration of the vocal cords (pitch), and amplitude (volume). Pitch is higher for females than for males. Each speaker recognition system has two phases: Enrollment and verification. During enrollment, the speaker's voice is recorded and typically a number of features are extracted to form a voice print, template, or model. In the verification phase, a speech sample or "utterance" is compared against a previously created voice print Speaker recognition systems fall into two categories: Text-Dependent and Text- Independent. In a text-dependent system, text is same during enrollment and verification phase. In Text-independent systems the text during enrollment and test is different. In fact, the enrollment may happen without the user's knowledge, as in the case for many forensic applications.

V. CONCLUSION

The outcome of the research work would be a secured and efficient Voiceprint Authentication System, which typically improves system performance in the current technology.

REFERENCE

[1] R. H. Bolt, F. S. Cooper, E. E. David, P. B. Denes, J. M. Pickett, and K. N. Stevens. "Identification of a speaker by speech spectrograms: How do scientists view its reliability for use as legal evidence?", Science, 166:338-343, 1969.

[2] K. N. Stevens, C. E. Williams, J. R. Carbonell, and B. Woods. "Speaker authentication and identification: A comparison of spectrographic and auditory presentations of speech material". J. Acoust. Soc. Am., 44:1596-1607, 1968.

[3] L. G. Kersta. "Voiceprint identification". Nature, 196:1253{1257, 1962.

[4] B. S. Atal. "Automatic recognition of speakers from their voices". Proc.IEEE, 64:460-475, 1976.

[5] G.R.Doddington. "Speaker recognition-identifying people by their voices". Proc. IEEE, 73:1651-1664, 1985. Page 6

[6] AL. Higgins and R. E. Wohlford. "A new method for text-independent speaker recognition". In Proc. IEEE Int. Conf. on Acoust., Speech, Signal Processing (ICASSP), pages 869-872, 1986.

[7] F. K. Soong and A. E. Rosenberg. "On the use of instantaneous and transitional spectral information in speaker recognition". IEEE Trans. Acoust., Speech, Signal Processing, 36(6):871-879, 1988.

[8] S. Furui, "Cepstral analysis technique for automatic speaker verification", IEEE Transactions on Acoustics, Speech and Signal Processing, vol. 29, pp. 254-272, 1981.