EVALUATING THE GAPS IN SPEECH RECOGNITION TECHNIQUES

Varinderjit Singh[1], Dr Paramjeet Singh[2]
Dept. of Computer Science& Engineering,
GZS PTU CAMPUS, Punjab. India.
[1]vickyss550@gmail.com, [2]param2009@yahoo.com

Abstract
This paper presents an related work on different speech recognition techniques. Speech recognition is becoming an hot issue in real time security systems. The methods developed so far are working efficiently and giving good results but they are not much accurate. Neural networks take much time in training the neural and so the methods based on the neural network proved to be inefficient. This paper has evaluated the different gaps in existing research and techniques. It has been found that existing methods do not provide high accuracy and take lot of time in recognition. Therefore in-order to provide high accuracy and speedup to clients, a new technique is required.

Index terms: Speech, Security systems and Speech recognition

I. INTRODUCTION

The speech recognition [1] – [12] is the process of taking the spoken word as an input and matches it with the database of previously recorded speeches on basis of various parameters. This can be done by various methods. In this dissertation reverse wave transformation is used and speaker is recognized on basis of various parameters. The ultimate aim of Speech recognition research is to allow a computer to recognize matches of audio with 100% accuracy that are spoken by any person, independent of vocabulary size, noise, accent, or channel conditions.

Speech Recognition [1] – [12] is a process of automatically recognizing who is speaking on the basis of features of speaker of the speech signal. Basically, speaker recognition is classified in to speaker identification and speaker verification. Wide application of speech recognition system includes control access to services such as banking by telephone, database access services, voice dialling telephone shopping so on. Now, speech recognition technology is the most suitable technology to create new services that will make our everyday lives more secured. Speech utterances [1] from one speaker may vary due to age, sex, noise and background environment, and speaker tone and attitude. After fifty years of research of speech processing in the time domain, the accuracy of speech recognizers has not reached a desirable success rate. The complexity of speech recognition systems has increased nowadays.

The common problem with identification system [2] nowadays is that the system can easily be fooled. Although it uses biometric identification which is unique from everyone else, there are still ways to fool the system. As for fingerprint identification, it does not have a good psychological effect on the people because of its wide use in crime investigations. Also, when the surface of human fingerprint is hurt, the recognition system will have problems to recognize the user because the system recognizes the surface of the fingerprints while for face recognition, people are still working the pose and the illumination invariance

Speech is a natural mode [3] of communication for people. All the relevant skills are learned during early childhood, without instruction, and we continue to depend on speech communication throughout our lives. The human vocal tract and articulators are biological organs with nonlinear properties, whose operation are not just under conscious control but also affected by factors ranging from gender to upbringing to emotional state. As a result, vocalizations can vary widely in terms of their accent, pronunciation, articulation, roughness, nasality, pitch, volume, and speed; moreover, during transmission, our irregular speech patterns can be further distorted by background noise and echoes, as well as electrical characteristics (if telephones or other electronic equipment are used).
However [4], the task of speech recognition is difficult because:

- Lot of redundancy is present in the speech signal that makes discriminating between the classes difficult.
- Presence of temporal and frequency variability such as intra speaker variability in pronunciation of words and phonemes as well as inter speaker variability e.g. the effect of regional dialects.
- Context dependent pronunciation of the phonemes (co-articulation).

- Signal degradation due to additive and convolution noise present in the background or in the channel
- Signal distortion due to non-ideal channel characteristic.

2. CLASSIFICATION OF SPEECH RECOGNITION SYSTEMS

Most speech recognition systems can be classified according to the following categories [4]:

A. Speaker Dependent vs Speaker Independent

A speaker-dependent speech recognition [5] & [6] system is one that is trained to recognize the speech of only one speaker. Such systems are custom built for just a single person, and are hence not commercially viable. Conversely, a speaker-independent system is one that is independence is hard to achieve, as speech recognition systems tend to become attuned to the speakers they are trained on, resulting in error rates that are higher than speaker dependent system.

B. Isolated vs. Continuous

In isolated speech [7] & [8], the speaker pauses shortly between every word, while in continuous speech the speaker speaks in a continuous and possibly long stream, with little or no breaks in between. Isolated speech recognition systems are easy to build, as it is unimportant to determine where one word ends and another starts, and each word tends to be more cleanly and clearly spoken. Words spoken in continuous speech on the other hand are subjected to the co-articulation effect, in which the pronunciation of a word is modified by the words surrounding it. This makes training a speech system difficult, as there may be many inconsistent pronunciations for the same word.

C. Keyword based vs. Subword unit based

A speech recognition system [5] - [9] can be trained to recognize whole words, like dog or cat. This is useful in applications like voice-command-systems, in which the system need only recognize a small set of words. This approach, while simple, is unfortunately not scalable. As the dictionary of recognized words grow, so too the complexity and execution time of the recognizer.

3. Fundamentals of the speech recognition system

Speaker Recognition [1] - [12] has always focuses on security system of controlling the access to secure data or information from being accessed by anyone. Speaker recognition is the process of automatically recognizing the speaker voice according to the basis of individual information in the voice waves. It is a branch of biometric authentication where it is one of the fast gaining popularity as means of security measures due to its unique physical characteristics and identification of individuals. According to the earliest method of biometric identification includes fingerprints and handwriting while the recent ones are using eye scan, face scan or voice print [7].

Speaker Identification [8] is the process of using the voice of speaker to verify their identity and control access to services such as voice dialing, mobile banking, database access services, voice mail or security control to a secured system. The areas that are mainly use the speech or voice processing are:

- Access control to confidential areas in facility or terminal.
- Mobile purchases through online bank transaction
- Remote monitoring
- Credit card activation through mobile.
- Forensic voice sampling.

A. BIOMETRICS

Biometric can be defined as study of life which includes humans, animals and plants. The

Word is taken from the Greek word where ‘Bio’ means life and ‘Metric’ means measure. It is a system of identifying or recognizing the identity of a living person based on physiological or behavioural characteristics. If the Biometric needs to be success in the security system, Biometric should have the uniqueness where different people have its own traits like the DNA of humans. The comparison of some Biometric Identification and the patterns are given into the following in Table 1 [7]:

- Distinctiveness: characteristic in pattern among population
- Robustness: repeatable, not subject to large changes
- Accessibility: easily presented to sensor
- Acceptability: perceived as non-intrusive by users

Table 1: Comparison between Biometric Identification

Biometric Patterns	Iris	Face	Finger prints	Voice
Distinctiveness	High	High	High	Avg
Robustness	High	High	Avg	Avg
Accessibility	Low	Avg	Avg	High
Acceptability	Avg	High	Avg	High

From the Table 1, it shows that the biometric patterns of voice are both high accordance with

accessibility and acceptability. But for distinctiveness and robustness, it is moderate. Actually, the security measures three different goals that are:

- Protect the confidentiality
- Protect the integrity
- The availability of data for authorized use

For the third goal, the accessibility and acceptability have fully met the goal of the security measures [7].

Speaker recognition is the computing task of validating a user's claimed identity using characteristics extracted from their voices. In contrast to other biometric technologies which are mostly image based and require expensive proprietary hardware such as vendor's fingerprint sensor or iris-scanning equipment, the speaker recognition 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 automated speaker recognition the speech signal is processed to extract speaker-specific information. The speaker specific information's are used to generate voiceprint which cannot be replicated by any source except the original speaker. This makes speaker recognition a secure method for authenticating an individual since unlike passwords or tokens; it cannot be stolen, duplicated or forgotten.

Table 2. Typical applications of speaker recognition systems [8]

Areas	**Specific Applications**
Authentication	Remote Identification & Verification, Mobile Banking Using ATM Transaction , Access Control
Information Security	Personal Device Logon, Desktop Logon, Application Security, Database Security, Medical Records, Security Control For Confidential Information
Law Enforcement	Forensic Investigation, Surveillance Applications
Interactive Voice	Banking Over A Telephone Network, Information And Reservation
Response	Services, Telephone Shopping, Voice Dailing, Voice Mail

A speaker's voice [9] 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.

4. Related work

Wahyu et al. (2012) [1] has been studied in the simulation voice recognition system for controlling robotic applications. Voice recognition is a system to convert spoken words in well-known languages into written languages or translated as commands for machines, depending on the purpose. The input for that system is "voice", where the system identifies spoken word(s) and the result of the process is written text on the screen or a movement from machine's mechanical parts.

Khalid et al. (2009) [2] has been researched on the using the sound recognition techniques to reduce the electricity consumption in highways. The lighting is available for the highways to avoid accidents and to make the driving safe and easy, but turning the lights on all the nights will consume a lot of energy which it might be used in another important issues.

Lindasalwa et al. (2010) [3] has studied the voice recognition algorithms using Mel Frequency Cepstral Coefficient (MFCC) and Dynamic Time Warping (DTW) techniques. Digital processing of speech signal and voice recognition algorithm is very important for fast and accurate automatic voice recognition technology. The voice is a signal of infinite information. A direct analysis and synthesizing the complex voice signal is due to too much information contained in the signal.

Sonam et al. (2012) [4] has been studied the controlling of device through voice recognition using MATLAB. Speech Recognition is the process of automatically recognizing a certain word spoken by a particular speaker based on individual information included in speech waves. Hardware components used for developing a technique, serial port, MAX232 voltage level converter controller to take input and generate output.

Reena et al. (2012) [5] has been researched in the speech recognition and synthesis tool: assistive technology for physically disabled persons. Attempt has been made to

develop a Speech Recognition and Synthesis Tool (SRST) as assistive technology to provide a solution for communication between two physically disabled persons; blind and deaf. An off-line chat room has made where two physically challenged persons can communicate to each other in US English accent via USB Serial Adaptor.

Santosh et al. (2010) [6] have studied the review on speech recognition technique. The Speech is most prominent & primary mode of Communication among of human being. The communication among human computer interaction is called human computer interface. Speech has potential of being important mode of interaction with computer.

Cristian et al. (2008) [7] has been researched the building domain specific languages for voice recognition applications. The proposed method has implemented the voice recognition for the control of software applications. The solutions proposed are based on transforming a subset of the natural language in commands recognized by the application using a formal language defined by the means of a context free grammar. At the end of the proposed technique has presented the modality of integration of voice recognition and of voice synthesis for the Romanian language in Windows applications.

Katti et al. (2009) [8] has researched the Speech Recognition by Machine. A brief survey on Automatic Speech Recognition and discusses the major themes and advances made in the past 60 years of research, so as to provide a technological perspective and an appreciation of the fundamental progress that has been accomplished in this important area of speech communication. After years of research and development the accuracy of automatic speech recognition remains one of the important research challenges (eg. variations of the context, speakers, and environment).

C. Vimala et al. (2012) [9] has studied in the review on speech recognition challenges and approaches. Speech technology and systems in human computer interaction have witnessed a stable and remarkable advancement over the last two decades. Today, speech technologies are commercially available for an unlimited but interesting range of tasks. These technologies enable machines to respond correctly and reliably to human voices, and provide useful and valuable services.

Ibrahim Patel et al. (2010) [10] has studied the speech recognition using HMM with MFCC as an analysis using frequency spectral decomposing technique. The author described an approach to the recognition of speech signal using frequency spectral information with Mel frequency for the improvement of speech feature representation in a HMM based recognition approach. Frequency spectral information is incorporated to the conventional Mel spectrum base speech recognition approach.

C.Y. Fook et al. (2012) [11] has researched Malay speech recognition and audio visual speech recognition. Automatic speech recognition (ASR) is an area of research which deals with the recognition of speech by machine in several conditions. ASR performs well under restricted conditions (quiet environment), but performance degrades in noisy environments. This paper presents a brief survey on Automatic Speech Recognition on Malays Corpus and multi-modal speech recognition on others Corpus.

V. Mitra (2012) [12] has studied in the normalized amplitude modulation features for large vocabulary noise-robust speech recognition. Background noise and channel degradations seriously constrain the performance of state-of-the-art speech recognition systems. Studies comparing human speech recognition performance with automatic speech recognition systems indicate that the human auditory system is highly robust against background noise and channel variabilities compared to automated systems.

Yu Shao (2011) [13] has been researched on bayesian separation with sparsity promotion in perceptual wavelet domain for speech enhancement and hybrid speech recognition. Speech recognition accuracy can be improved by the removal of noise. However, errors in the estimated signal components can also obscure the recognition. This research has presented a framework of wavelet-based techniques to harness the automatic speech recognition performance in the presence of background noise. The proposed robust speech recognition system is realized by implementing speech enhancement pre-processing, feature extraction, and a hybrid speech recognizer in the time-frequency space.

5. Conclusion

This paper presents a literature survey on different speech recognition techniques and finds the gaps in existing literatures. It is concluded that the methods developed so far are working efficiently and giving good results but they are not much accurate. Neural networks take much time in training the neural and so the methods based on the neural network proved to be inefficient. So in near future work will be extended to proposed a new technique which is time effective in nature and also provide better results.

References

[1] Wahyu Kusuma R. and Prince Brave Guhyapati, **"**simulation voice recognition system for controlling robotic applications", Journal of Theoretical and Applied Information Technology, Vol. 39, pp. 188- 196, 15 May 2012.

[2] Khalid T. Al-Sarayreh, Rafa E. Al-Qutaish and Basil M. Al-Kasasbeh, "Using The Sound Recognition Techniques To Reduce The Electricity Consumption In Highways", Journal of American Science, pp. 1-12, 2009.

[3] Lindasalwa Muda, Mumtaj Begam and I. Elamvazuthi, "Voice Recognition Algorithms Using Mel Frequency Cepstral Coefficient (MFCC) and Dynamic Time Warping (DTW) Techniques", Journal Of Computing, Vol. 2, pp. 138-143 , March 2010.

[4] Kumari Sonam, Arya Kavita, Saxena Komal, "controlling of device through voice recognition using MATLAB", International Journal of Advanced Technology & Engineering Research (IJATER), Vol. 2, pp. 177-179, March 2012.

[5] Sharma F. Reena and Wasson S. Geetanjali, "speech recognition and synthesis tool: assistive technology for physically disabled persons", International Journal of Computer Science and Telecommunications, Vol. 3, pp. 86-91, April 2012.

[6] Santosh K.Gaikwad, Bharti W.Gawali and Pravin Yannawar "review on speech recognition technique", International Journal of Computer Applications, Vol. 10, pp-16-24, Nov. 2010.

[7] Ionita Cristian, "Building Domain Specific Languages For Voice Recognition Applications", Revista Informatica Economica, pp. 105-109, 2008.

[8] M.A.Anusuya and S.K.Katti, "Speech Recognition by Machine", International Journal of Computer Science and Information Security, Vol. 6, pp. 181-205, 2009.

[9] Vimala C ., "review on speech recognition challenges and approaches", World of Computer Science and Information Technology Journal, Vol. 2,pp. 1-7, 2012.

[10] Ibrahim Patel, "Speech Recognition Using HMM with MFCC- An Analysis Using Frequency Specral Decomposion Technique", Signal & Image Processing An International Journal (SIPIJ) Vol.1, pp. 101-110 , December 2010.

[11] Fook C.Y., "Malay Speech Recognition and Audio Visual Speech Recognition", International Conference on Biomedical Engineering (ICoBE), pp. 479-484, Feb. 2012.

[12] Mitra, V. , "normalized amplitude modulation features for large vocabulary noise-robust speech recognition", IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Vol. 1, pp. 4117-4120, March 2012.

[13] Yu Shao, "Bayesian Separation with Sparsity Promotion In Perceptual Wavelet Domain For Speech Enhancement And Hybrid Speech Recognition", IEEE Transactions on Systems, Vol 41, pp. 284-294, March 2011.