VOICE RECOGNITION USING FAST WAVE TRANSFORMATION

Varinderjit Singh[#1], Dr Paramjitsingh[*2]
[1]*Dept. of Computer Science& Engineering*
GZS PTU CAMPUS, Punjab, India
vickyss550@gmail.com
[2]*Dept. of Computer Science& Engineering*
GZS PTU CAMPUS, Punjab, India
Param2009@yahoo.com

Abstract-**This paper presents a review of the papers on voice recognition techniques. The different techniques are considered which can achieve objective of voice recognitionby covering different aspects and circumstances. Thisresearch workconsider theaudio voices in .wav format using wave transformation. The problem of noise in the input signal is also considered in this paper. It is found that the most of researchers has neglected the noise in the signals.**

Keywords-**Reverse Wave Transformation, Speaker, VoiceRecognition, Wave format**

I. INTRODUCTION

Voice recognition [1] is a technology that allows spoken input into systems. You talk to your computer, phone or device and it uses what you said as input to trigger some action. The technology is being used to replace other methods of input like typing, clicking or selecting in other ways. It is a means to make devices and software more user-friendly and to increase productivity. There are plenty of applications [2] –[3] and areas where voice recognition is used, including the military, as aid for impaired persons (imagine a person with crippled or no hands or fingers), in the medical field, in robotics etc. In the near future, nearly everyone will be exposed to voice recognition due to its propagation among common devices like computers and mobile phones. Voice recognition, in its version known as Voice to Text (STT), has also been used for a long time to translate spoken words into text. “You talk, it types”, as Via Voice would say on its box. But there is one problem with STT as we know it. More than 10 years back, I tried Via Voice and it did not last a week on my computer. Why? It was grossly inaccurate and I ended up spending more time and energy speaking and correcting than typing everything. Via Voice is one of the best in the industry, so imagine the rest. The technology has matured and improved, but voice to text still makes people ask questions. One of its main difficulties is the immense variations among people in pronouncing words. Voice recognition [3] is proving to be better off as an input method for new phones and communication technologies like VoIP, than as a productivity tool for mass text input.

This research work focus on providing better performance in audio recognition algorithm by integrating digital signal transposition with audio recognition techniques. Main emphasis is to recognize audio with reverse wave transformation to achieve better results.

Voice recognition is becoming popular in real time security systems. The methods developed so far are working efficiently and giving good results. Neural networks take much time in training the neural and so the technique is going to be formed that take much less time by simply processing the signal by using reverse wave transposition.

This paper deals with efficiently recognizing voice by using signal transposition and also gives better results than technique used using neural networks. To achieve this, a new hybrid methodology will be proposed which will recognize the audio.

II. LITERAUTURE SURVEY

WahyuKusuma R. and Prince Brave Guhyapati has been studied in the simulation voice recognition system for controlling robotic applications [1]. 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.

This focused on analysis of matching process to give a command for multipurpose machine such as a robot with Linear Predictive Coding (LPC) and Hidden Markov Model (HMM), where LPC is a method to analyze voice signals by giving characteristics into LPC coefficients. In the other hand, HMM is a form of signal modeling where voice signals are analyzed to find maximum probability and recognize words given by a new input based from the defined codebook.

T. Khalid Al-Sarayreh et al has been researched on the using the sound recognition techniques to reduce the electricity consumption in highways [2]. 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. The author aimed at using the sound recognition techniques in order to turn the lights on only when there are cars on the highway and only for some period of time. In more details, Linear Predictive Coding (LPC) method and feature extraction will be used to apply the sound recognition. Furthermore, the Vector Quantization (VQ) will be used to map the sounds into groups in order to compare the tested sounds.

LindasalwaMuda and MumtajBegam et al has studied the voice recognition algorithms using Mel Frequency Cepstral Coefficient (MFCC) and Dynamic Time Warping (DTW) techniques [3]. 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. Therefore the digital signal processes such as Feature Extraction and Feature Matching are introduced to represent the voice signal. Several methods such as Liner Predictive Predictive Coding (LPC), Hidden Markov Model (HMM), Artificial Neural Network (ANN) and etc are evaluated with a view to identify a straight forward and effective method for voice signal. The extraction and matching process is implemented right after the Pre Processing or filtering signal is performed. The non-parametric method for modelling the human auditory perception system, Mel Frequency Cepstral Coefficients (MFCCs) are utilize as extraction techniques. The non-linear sequence alignment known as Dynamic Time Warping (DTW) introduced by Sakoe Chiba has been used as features matching techniques. Since it's obvious that the voice signal tends to have different temporal rate, the alignment is important to produce the better performance. This has presented the viability of MFCC to extract features and DTW to compare the test patterns.

SonamKumari et al has been studied the controlling of device through voice recognition using MATLAB [4]. 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. To drive the relays we have used ULN 2803 IC which has arrays of 8 Darlington pair of transistors. Darlington pair of transistors are capable of provide larger amount of current to drive the relays. 8 LEDs are used as indicator each corresponding to 8 data lines of the data port. This technique makes it possible to use the speaker's voice to verify his/her identity and provide controlled access to services like voice based biometrics, database access services, voice based dialing, voice mail and remote access to computers. The algorithms for the speech recognition has been developed and implemented on MATLAB. These algorithms can be used for any security system in which the person authentication is required.

F. Reena Sharma and S. Geetanjali Wasson has been researched in the speech recognition and synthesis tool: assistive technology for physically disabled persons [5]. 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. SRST is designed in Microsoft.NET 3.5 framework using C# Programming in Microsoft Visual Studio 2008 Environment. Microsoft Windows Speech Application Programming Interface (SAPI) 5.3 and, system speech recognition and system speech synthesis namespaces are used for speech to text conversion and vice-versa. The blind and deaf students trained on the tool and window speech recognition system inbuilt in windows vista. Then this communication tool is implemented in real setting for the physically challenged persons. It has been observed that they communicated to each other effectively only in noise free environment and blind students taken more time than deaf students to familiar with tool commands and speech recognition commands.

Santosh K.Gaikwad et al have studied the review on speech recognition technique [6]. 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. It given an overview of major technological perspective and appreciation of the fundamental progress of speech recognition and also gives overview technique developed in each stage of speech recognition. It helped in choosing the technique along with their relative merits & demerits. A comparative study of different technique is done as per stages. This technique concludes with the decision on feature direction for developing technique in human computer interface system using Marathi Language.

Cristian Ionita has been researched the building domain specific languages for voice recognition applications [7]. 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.

M.A.Anusuya and S.K.Katti has researched the Speech Recognition by Machine [8]. 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).The design of Speech Recognition system requires careful attentions to the following issues: Definition of various types of speech classes, speech representation, feature extraction techniques, speech classifiers, and database and performance evaluation. The problems that are existing in ASR and the various techniques to solve these problems constructed by various research workers have been presented in a chronological order. Hence authors worked in the area of speech recognition. The objective of this review research has to summarize and compared some of the well-known methods used in various stages of speech recognition system and identify research topic and applications which are at the forefront of this exciting and challenging field.

C. Vimala has been studied in the review on speech recognition challenges and approaches [9]. 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. Recent research concentrates on developing systems that would be much more robust against variability in environment, speaker and language. Hence today's researches mainly focus on ASR systems with a large vocabulary that support speaker independent operation with continuous speech in different languages. This research has given an overview of the speech recognition system and its recent progress. The primary objective of this research has to compare and summarized some of the well-known methods used in various stages of speech recognition system.

Ibrahim Patel has studied the speech recognition using HMM with MFCC- an analysis using frequency specraldecomposion technique [10]. 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. The Mel frequency approach exploits the frequency observation for speech signal in a given resolution which results in resolution feature overlapping resulting in recognition limit. Resolution decomposition with separating frequency is mapping approach for a HMM based speech recognition system. The Simulation results show an improvement in the quality metrics of speech recognition with respect to computational time, learning accuracy for a speech recognition system.

C.Y. Fook has researched Malay speech recognition and audio visual speech recognition [11]. Automatic speechrecognition (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 SpeechRecognition on Malays Corpus and multi-modal speechrecognition on others Corpus. The audio only speechrecognition has been done by many researchers few decades. After years of research and development the performance of automatic speechrecognition remains one of the important research challenges (eg., variations of the context, database, and environment). The criteria for designing SpeechRecognition system are pre-processing filter, end-point detection, feature extraction techniques, speech classifiers, database, and performance evaluation. The existing problems that are in Automatic SpeechRecognition (ASR)-noise environments and the various techniques to solve these problems had constructed. The objective of this review has to summarize and compared some of the well know methods used by previous researcher.

V. Mitrahas studied in the normalized amplitude modulation features for large vocabulary noise-robust speech recognition [12]. Background noise and channel degradations seriously constrain the performance of state-of-the-art speechrecognition systems. Studies comparing human speechrecognition performance with automatic speechrecognition systems indicate that the human auditory system is highly robust against background noise and channel variabilities compared to automated systems. A traditional way to add robustness to a speechrecognition system is to construct a robust feature set for the speechrecognition model.

YuShao has been researched on bayesian separation with sparsity promotion in perceptual wavelet domain for speech enhancement and hybrid speech recognition [13]. Speechrecognition 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 speechrecognition performance in the presence of background noise. The proposed robust speechrecognition system is realized by implementing speech enhancement preprocessing, feature extraction, and a hybrid speech recognizer in the time-frequency space.

A.A.M. Abushariahhas researched on the English digits speech recognition system based on hidden Markov models [14]. This research has aimed to design and implement English digits speechrecognition system using MATLAB (GUI). This work was based on the Hidden Markov Model (HMM), which provides a highly reliable way for recognizing speech. The system is able to recognize the speech waveform by translating the speech waveform into a set of feature vectors using Mel Frequency Cepstral Coefficients (MFCC) technique This study focuses on all English digits from (Zero through Nine), which is based on isolated words structure. Two modules were developed, namely the isolated words speechrecognition and the continuous speechrecognition. Both modules were tested in both clean and noisy environments and showed a successful recognition rates

V. OBJECTIVES OF THE SURVEY

1. Time

The time is the time taken by the algorithm to run. Time is directly proportional to number of inputs elements. Speed is inversely proportional to time. Speed increases when time taken decreases and decrease with increase in time taken to recognize voices. The time must be reduced.

2. Hit Ratio

It is defined as the number of voices recognized to the total number of voices that are given input. Hit ratio should be maximum in order to achieve full accuracy. Hits are the number of times the input sample is recognized.

3. Error rate

The error rate must be reduced so that accurate recognition of input voices must be there. Basically,

ERROR RATE= 1- HIT RATIO

Error rate must be reduced.

4. Accuracy rate

Accuracy rate must be improved to achieve good results. Accuracy increases with decrease in error rate and increase in hit ratio

Accuracy=Hit Ratio/ Total * 100

Table 1: Metrics

Serial No.	Metrics	Need To Be Maximized Or Minimized
1.	Time	Minimized
2.	Hit Ratio	Maximized
3.	Error Rate	Minimized
4.	Accuracy Rate	Maximized

VI. HOW TO COMPARE RECORDINGS

A. Frequency Domain

It is quite evident that any voice analysis [5] in time domain would be extremely impractical. Instead, an analysis of the frequency spectra in a voice (which remains predominately unchanged as speech is slightly varied) turned out to be a more viable option. Converting all recordings into frequency domain (by applying the Discrete Fourier Transform) greatly simplified the process of comparing two recordings. That being said, working in frequency domain also provided a new set of issues that required attention.

B. Finding a Norm

Due to nature of human speech [3], all data pertaining to frequencies above 600Hz can safely be discarded. Therefore, once a recording is converted into frequencey domain, it could then be simply regarded as a vector in 600-dimensional Euclidean space. At this point, a comparison between two vectors could easily be carried out by normalizing the vectors (giving them length 1) then computing the norm of the difference betweeen the two (of course, the difference between two vectors in R600 is performed by subtracting componentwise). Unfortunately, exactly which norm to use is not immediately clear. After carefully comparing and contrasting the use of the Taxicab, Euclidean, and Maximum norms, it became clear that the Euclidean norm most accurately measured the closeness between different frequency spectra. Once the norm function was chosen, all that remained was to decide exactly how small the norm of the difference of two vectors had to be in order to determine that both recordings originated from the same person.

C. Chebyshev's Inequality

Recall that Chebyshev's Inequality states [4] that in particular, at least 3/4 of all measurements from the same population fall within 2 standard deviations of the mean. Hence, in response to the problem posed at the end of the previous paragraph, the following solution can be formulated:

By requiring that the norm of the difference fall within 2 standard deviations of the normal average voice, we are then ensured that at least 3/4 of the time, the algorithm would recognize a voice correctly.

IX. CONCLUSION AND FUTURE DIRECTIONS

This paper has proposed new voiceidentificationalgorithm which has great significant in improving the voice recognition performance. The technique is able to authenticate the particular speaker based on the individual information that is included in the audio signal and the recognition is done using reverse wave transformation. The results show that proposed technique provides high accuracy rate.

In the future, the focus can be on reducing the noise or background disturbance that is introduced in the audio samples automatically while recording. The various filtering techniques can be applied in order to reduce disturbance.

REFERENCES

[1] *WahyuKusuma 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] *LindasalwaMuda, MumtajBegam 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] *KumariSonam, Arya Kavita, SaxenaKomal, "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 PravinYannawar"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", RevistaInformaticaEconomica, 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] *VimalaC ., "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 SpecralDecomposion 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 WithSparsity Promotion In Perceptual Wavelet Domain For Speech Enhancement And Hybrid Speech Recognition", IEEE Transactions on Systems, Vol 41, pp. 284-294, March 2011.*

[14] *Abushariah, A.A.M., " English Digits Speech Recognition System Based On Hidden Markov Models", International Conference on Computer and Communication Engineering (ICCCE), pp. 1-5, May 2010.*