ISSN: 2278 – 1323
International Journal of Advanced Research in Computer Engineering & Technology (IJARCET)
Volume No. 2, Issue No. 6, June 2013

A Survey on Speech Recognition

[1]Harpreet Singh,[2]Ashok Kumar Bathla

[1,2]Department of Computer Engineering, Punjabi University
Yadavindra College of Engineering, Talwandi Sabo
Punjab, India

ABSTRACT

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. Speech recognition is the process of the computer identifying human speech to generate a string of words or commands. The output of speech recognition systems can be applied in various fields. Besides, there are many artificial intelligent techniques available for Automatic Speech Recognition (ASR) development. This paper gives an overview of the speech recognition system and its recent progress. The primary objective of this paper is to compare and summarize some of the well known methods used in various stages of speech recognition system.

Keywords— ***Speech Recognition; Feature Extraction***

I. INTRODUCTION

Speech is the most basic, common and efficient form of communication method for people to interact with each other. People are comfortable with speech therefore persons would also like to interact with computers via speech, rather than using primitive interfaces such as keyboards and pointing devices.

This can be accomplished by developing an Automatic Speech Recognition (ASR) system which allows a computer to identify the words that a person speaks into a microphone or telephone and convert it into written text. As a result it has the potential of being an important mode of interaction between human and computers. Since the 1960s computer scientists have been researching ways and means to make computers able to record interpret and understand human speech. Throughout the decades this has been a daunting task. Even the most rudimentary problem such as digitalizing (sampling) voice was a huge challenge in the early years. It took until the 1980s before the first systems arrived which could actually decipher speech. Communication among the human being is dominated by spoken language, therefore it is natural for people to expect speech interfaces with computer .computer which can speak and recognize speech in native language. Machine reorganisation of speech involves generating a sequence of words best matches the given speech signal.

II TYPES OF SPEECH

Speech recognition system can be separated in different classes by describing what type of ullerances they can recognize.

A. Isolated word

Isolated word recognizes attain usually require each utterance to have quiet on both side of sample windows. It accepts single words or single utterances at a time .This is having “Listen and Non Listen state”. Isolated utterance might be better name of this class.

B. Connected word

Connected word system are similar to isolated words but allow separate utterance to be “run together minimum pause between them.

C. Continuous speech

Continuous speech recognizers allows user to speak almost naturally, while the computer determine the content. Recognizer with continues speech capabilities are some of the most difficult to create because they utilize special method to determine utterance boundaries.

D. Spontaneous speech

At a basic level, it can be thought of as speech that is natural sounding and not rehearsed .an ASR System with spontaneous speech ability should be able to handle a variety of natural speech feature such as words being run together.

III. RELATED WORK

In 2004 Jingdong Chen and et al has discussed that despite their widespread popularity as front-end parameters for speech recognition, the cepstral coefficients derived from either linear prediction analysis or a filter-bank are found to be sensitive to additive noise. In this letter, we discuss the use of spectral subband centroids for robust speech recognition. We show that centroids, if properly selected, can achieve recognition performance comparable to that of the mel-frequency cepstral coefficients (MFCCs) in clean speech, while delivering better performance than MFCCin noisy environments. A procedure is proposed to construct the dynamic centroid feature vector that essentially embodies the transitional spectral information.

In 2005 Esfandiar Zavarehei and et al has studied that a time-frequency estimator for enhancement of noisy speech signals in the DFT domain is introduced. This estimator is based on modeling and filtering the temporal trajectories of the DFT components of noisy speech signal using Kalman filters. The time-varying trajectory of the DFT components of speech is modelled by a low order autoregressive (AR) process incorporated in the state equation of Kalman filter. A method is incorporated for restarting of Kalman filters, after long periods of noise-dominated activity in a DFT channel, to mitigate distortions of the onsets of speech activity. The performance of the proposed method for the enhancement of noisy speech is evaluated and compared with MMSE estimator and parametric spectral subtraction. Evaluation results show that the incorporation of temporal information through Kalman filters results in reduced residual noise and improved perceived quality of speech.

In 2008 Chunyi Guo and et al has presented that speech is one of the most direct and effectivemeans of human communication, it's natural to apply biomimetic processing mechanism to automatic speech recognition to solve the existing speech recognition problems.Three typical techniques were selected respectively: Simulated evolutionary computation(SEC), artificial neural network(ANN) and fuzzy logic and reasoning technique, from intelligence building processing simulation, intelligence structure simulation and intelligence behavior simulation, to identify their applications in different stages of speech recognition. In 2009 Negar Ghourchian has presented that the use of a new Filtered Minima-Controlled Recursive Averaging (FMCRA) noise estimation technique as a robust front-end processing to improve the performance of a Distributed Speech Recognition (DSR) system in noisy environments. The noisy speech is enhanced by using a two-stage framework in order to simultaneously address the inefficiency of the Voice Activity Detector (VAD) and to remedy the inadequacies of MCRA. The performance evaluation carried out on the Aurora 2 task showed that the inclusion of FMCRA in the front-end side leads to a significant improvement in DSR accuracy.

In 2010 Richard M Stern and et al has described a way of designing modulation filter by data driven analysis which improves the performance of automatic speech recognition systems that operate in real environments. The filter for each nonlinear channel output is obtained by a constrained optimization process which jointly minimizes the environmental distortion as well as the distortion caused by the filter itself. Recognition accuracy is measured using the CMU SPHINX-III speech recognition system and the DARPA Resource Management and Wall Street Journal speech corpus for training and testing. It is shown that feature extraction followed by modulation filtering provides better performance than traditional MFCC processing under different types of background noise and reverberation.

In 2012 Kavita Sharma and et al has presented Speech Recognition is a broader solution which refers to a technology that can recognize a speech without being targeted at single speaker such call system can recognize arbitrary voice. The fundamental purpose of speech is communication, i.e., the transmission of messages. The problem in speech recognition is the speech pattern variability. The most challenging sources of variations in speech are speaker characteristics including accent, co-articulation and background noise. The filter bank in the front-end of a speech recognition system mimics the function of the basilar membrane. It is believed that closer the band subdivision to human perception better is the recognition results. Filter constructed from estimation of clean speech and noise for speech enhancement in speech recognition systems.

In 2012 Patiyuth Pramkeaw and et al has studied that the way to implement the Low-PassFilter with the Finite Impulse Response via using Signal Processing Toolbox under Matlab environment, successfully compassing analytical design of FIR filter and computational implementation, and evaluating its performance at Signal-to-Noise (S/N) ratio levels in which the desirable speech signal is intentionally corrupted by Gaussian White Noise. Results on word recognition are significantly improved, when the speech signals of the spoken word are first filtered by the implemented LPF, as compared with those of speech signals without filtering

In 2012 Bhupinder Singh has presented that phase of Speech Recognition Process using Hidden Markov Model. Preprocessing, Feature Extraction and Recognition three steps and Hidden Markov Model (used in recognition phase) are used to complete Automatic Speech Recognition System. Today's life human is able to interact with computer hardware and related machines in their own language. Research followers are trying to develop a perfect ASR system because we have all these advancements in ASR and research in digital signal processing but computer machines are unable to match the performance of their human utterances in terms of accuracy of matching and speed of response. In case of speech recognition the research followers are mainly using three different approaches namely Acoustic phonetic approach, Knowledge based approach and Pattern recognition approach.

IV. TABLE OF COMPARISON

Author(s)	Year	Paper Name	Technique	Results
Jingdong Chen et al.	2004	Recognition of Noisy Speech Using Dynamic Spectral Subband Centroids	Use of spectral subband centroids	It showed that the new dynamic SSC coefficients are more resilient to noise than the MFCC features.
Esfandiar Zavarehei et al.	2005	Speech Enhancement using Kalman filters for Restoration of short-time DFT trajectories	Concept sequence modeling, two-level semantic-lexical modeling, and joint semantic-lexical modeling	Increase the semantic information utilized and tightness of integration between lexical and semantic items
Chunyi Guo et al.	2008	Research on the Application of Biomimetic Computing in Speech Recognition	Simulated evolutionary computation(SEC), Artificial neural network(ANN) and Fuzzy logic	All three techniques shows the accuracy of speech recognition more than 95% and also lower the error rate.
Negar Ghourchian, et al	2009	Robust Distributed Speech Recognition using Two-Stage Filtered Minima Controlled Recursive Averaging	Filtered Minima-Controlled Recursive Averaging (FMCRA)	Improve the accuracy of the estimated noise spectrum and to reduce the speech leakage
Richard M Stern et al.	2010	MINIMUM VARIANCE MODULATION FILTER FOR ROBUST SPEECH RECOGNITION	CMU SPHINX-III speech recognition system, DARPA Resource Management and Wall Street Journal speech corpus	Improved speech recognition accuracy compared to traditional MFCC processing under different background noises
Kavita Sharma et al.	2012	Speech Denoising Using Different Types of Filters	FIR, IIR, WAVELETS, FILTER	Use of filters shows that estimation of clean speech and noise for speech enhancement in speech recognition
Bhupinder Singh et al.	2012	Speech Recognition with Hidden Markov Model	Hidden Markov Model	Develop a voice based user machine interface system.
Patiyuth Pramkeaw and et al	2012	Improving MFCC-based Speech Classification with FIR Filter	FIR filter	Shows the improvement in recognition rates of spoken words

I. Conclusions

Speech recognition has been in development for more than 50 years, and has been entertained as an alternative access method for individuals with disabilities for almost as long. In this paper, the fundamentals of speech recognition are discussed and its recent progress is investigated. The various approaches available for developing an ASR system are clearly explained with its merits and demerits. The performance of the ASR system based on the adopted feature extraction technique and the speech recognition approach for the particular language is compared in this paper. In recent years, the need for speech recognition research based on large vocabulary speaker independent continuous speech has highly increased. Based on the review, the potent advantage of HMM approach along with MFCC features is more suitable for these requirements and offers good recognition result. These techniques will enable us to create increasingly powerful systems, deployable on a worldwide basis in future

References

[1] Jingdong Chen, *Member*, Yiteng (Arden) Huang, Qi Li, Kuldip K. Paliwal "Recognition of Noisy Speech Using Dynamic Spectral Subband Centroid*s" in IEEE SIGNAL PROCESSING LETTERS, VOL. 11, NO. 2, FEBRUARY 2004.*

[2] Hakan Erdogan, Ruhi Sarikaya, Yuqing Gao "Using semantic analysis to improve speech recognition" performance in *Elsevier 2005*
.

[3] Chunyi Guo, Runzhi Li, Lei Shi "Research on the Application of Biomimetic Computing in Speech Recognition" *in IEEE 2008*
.

[4] Negar Ghourchian, Sid-Ahmed Selouani, Douglas O'Shaughnessy "Robust Distributed Speech Recognition using Two- Stage Filtered Minima Controlled Recursive Averaging" *in IEEE 2009*
.

[5] Yu-Hsiang Bosco Chiu, Richard M Stern "MINIMUM VARIANCE MODULATION FILTER FOR ROBUST SPEECH RECOGNITION" *in 2009.*

[6] Chadawan Ittichaichareon, Patiyuth Pramkeaw "Improving MFCC-based Speech Classification with FIR Filter" in International Conference on Computer Graphics, Simulation and Modeling (ICGSM'2012) July 28-29, 2012 Pattaya (Thailand).

[7] Kavita Sharma,Prateek Haksar " Speech Denoising Using Different Types of Filters" in International Journal of Engineering Research and Applications Vol. 2, Issue 1,Jan-Feb 2012, pp.809-811.

[8] Bhupinder Singh, Neha Kapur, Puneet Kaur "Speech Recognition with Hidden Markov Model: A Review" in International Journal of Advanced Research in Computer Science and Software Engineering Volume 2, Issue 3, March 2012.

Harpreet Singh received his B.Tech degree in Computer Engineering from Yadavindra college of Engineering Guru Kashi Campus Talwandi Sabo(Bathinda) under Punjabi University Patiala in 2010 and pursuing M.Tech (Regular) degree in computer engineering from Yadavindra College of Engineering Punjabi University Guru Kashi Campus Talwandi Sabo (Bathinda), batch 2011-2013. His research interests include Improving Speech Recognition using FIR-1 Filter. At present he has guided many students regarding the research study in Speech Recognition.