A Survey on Automatic Recognition of Speech via Voice Commands

Ayisha Zubain Bhandari
Department of Computer Science and Engineering
B.L.D.E.A's Dr.P.G.Halakatti College of Engineering and Technology, Vijayapur, Karnataka, India

Prof.B C.Melinamath
Department of Computer Science and Engineering
B.L.D.E.A's Dr.P.G.Halakatti College of Engineering and Technology, Vijayapur, Karnataka, India

Abstract- Speech has the ability to express the thoughts and feelings of an individual through a medium known as sound. In this paper, our focus is on various approaches of speech recognition. There are lots of languages in the world. Speech recognition is a process in which a system can recognize natural language. In this process a system is able to perform tasks based on voice commands. The first step is taking speech signal as an input. The next step is to eliminate the noise if present. The third step is matching the patterns to recognize the input speech. The mechanism used here is error tolerance so as to decrease human errors and, environmental noise. A database is used in order to keep records of various voice patterns, patterns are fixed in a way such that system can easily match the patterns without help of an end user.

Keywords – Speech Recognition, HMM, MFCC, Wavelet, Pattern Recognition, Acoustic-Phonetic

I. INTRODUCTION

Speech is the most basic, common and efficient form of communication method for people to interact with each other. People are very comfortable with speech therefore people would also like to interact with computers via speech, rather than using keyboards and pointing devices. In this paper we focus on storing patterns instead of voice. The voice which acts as input is decoded into patterns. These patterns decide the probability that a particular word is spoken or not. These patterns are used for indexation of weighted automata & generate a value of recognition. To achieve the robustness of speech lots of work is done so as to reduce the problems such as mismatch in speech signal, lots of environmental noise which may corrupt original data and this may lead to misinterpretation, different styles of speech. Fig. 1 represents the generalized Speech recognition technique that is used now days.

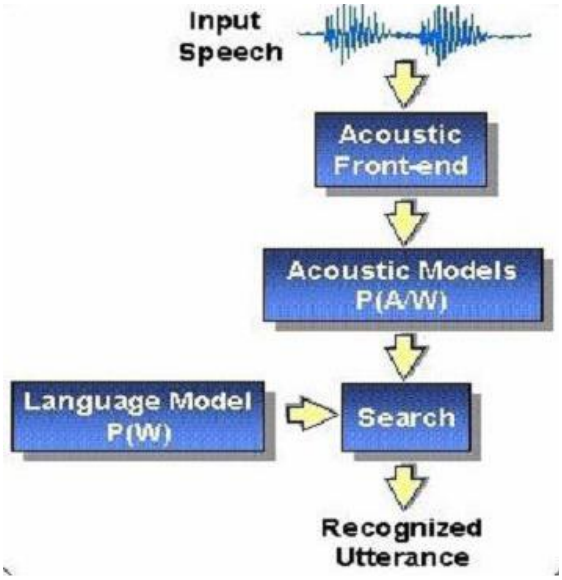

Figure 1 Schematic representation of generalized Speech recognition technique.

The rest of the paper is organized as follows. Literature Survey is explained in section II. Concluding remarks are given in section III.

II. LITERATURE SURVEY

To study and analyze more about speech recognition techniques, the following literature survey has been done. In this review paper some of the important techniques of speech recognition presented in the past that has been discussed. In [1] authors have worked on speaker independent speech recognition for developing a speaker free speech recognition system. There are many techniques used for feature extraction such as Linear Discriminant Analysis (LDA), Linear Predictive Coding Analysis (LPD), Perceptually Based Linear Predictive Analysis (PLP) and Mel-Frequency Cepstrum Coefficient (MFCC). In this work the authors have made use of Mel-Frequency Cepstrum Coefficient (MFCC). The authors have also made use of vector quantization in which the features that are extracted are used for testing the system. This work has made use of Mel-Frequency Cepstrum Coefficient (MFCC) inorder to focus on the attributes of speech generated from the speaker independent speech recognition system. This technique makes use of different speech samples of different persons and these samples are stored in the database of the system. The MFCC can be calculated using equation (1).

$$m = 2595\, log10\left(1 + \frac{f}{700}\right)$$

Where, m is the Mel-scale frequency and f is the perceived frequency. The Cepstral signal can be calculated using equation (2).

$$c(n) = ifft\ (\log\ |fft(s(n))$$

Where, c(n) is a cepstral signal, s(n) is the sample of speech signal, ifft refers to inverse fast fourier transform and fft refers to fast fourier transform. The combination of MFCC along with cubic log compression is used for processing of featured vectors of speech signal and is given by equation (3).

$$e(n) = \sum log\ \ 3(s[k]) * \ cos[n * (k + 0.5) * \frac{\pi}{m}]$$

Where s[k] represents energy of each Mel window ranging from $1 \leq k \leq M$, M signifies Mel window number ranging from 20 t0 24 i.e. from $1 \leq n \leq L$ and L is order of M. This method makes use of MATLAB. The authors have tested this technique for 20 persons where the 20 persons have recorded their speech for the counting numbers that begins from zero to nine in a less noisy environment with a frequency of 44.1KHZ based on their frequency of speech & based on the output, it was found that the system gave correct result as that was predicted. This system showed 90% accuracy as it makes use of vector quantization and execution rate is much faster due to vector coefficients.

In [2], authors firstly describe the factors that lead to inaccurate result in speech recognition system, the first factor is environment and environment plays a vital role in speech recognition system where if the system has to work under noisy environments, the background noise may corrupt the original data and leads to misinterpretation. The next factor is transducer, it is nothing but the device that records the speech, the transducers should be in a good condition a transducer may be a microphone or a telephone. The next factor is speaker, the devices that record speech are speaker dependent devices or speaker independent devices, whether the speaker is a male or a female, whether the speaker is a child or a young person or an old person. The next factor is the speech style, this describes the speech is high pitched or low pithched, whether the speech is said in an isolated word format or continuous format or it is a spontaneous utterance. Next the authors describe three different techniques used for speech recognition namely Acoustic phonetic approach, Pattern recognition approach and Artificial intelligence approach. In the acoustic phonetic approach the speech is decoded based on the relationship between the acoustic features and phonetic symbols. The next approach is the Pattern recognition approach this approach directly makes use of speech without explicit feature determination and segmentation. In this approach a test pattern is created and the test pattern is matched with the existing reference pattern and based on the best match the output is computed. The reference pattern is based on template based approach and stochastic approach. The template based approach makes use of dynamic programming to measure the differences in speaking rates across talkers as well as across repetitions of the word by the same talker. The stochastic approach makes use of Hidden Markov Models (HMM). The third approach is artificial intelligence approach it is also known as Knowledge based approach. It a combination of both acoustic approach and pattern recognition approach. This approach makes use of linguistic, phonetic and spectrogram for extracting the information. The authors summarize the above three techniques and come to a conclusion that the

Hidden Markov Model (HMM) has given an accurate result in speech recognition systems and now a day's many of the systems make use of this models.

In [3], author describes different speech feature extraction techniques and their decision based recognition through artificial intelligence and statistical techniques; they also include the various techniques such as analysis technique, feature extraction technique, modeling technique, matching technique. The author discusses the applications of speech recognition systems such as automatic transcription, multimedia content analysis and natural human-computer interfaces. The issues faced by automatic speech recognition design are environment, transducer, channel, speakers, speech styles, vocabulary. Accuracy and speed are the two criteria that measures performance of speech recognition system. The authors have presented a survey on speech recognition system i.e.; it deals with identification of person who is speaking, characterizing their voices. The authors come to a conclusion that MFCC along with HMM provide more accurate and performance is high.

In [4], authors gave an overview of different techniques that are used in speech recognition system. The authors first described the types of speech utterances i.e. isolated words; this makes use of single utterance at a point in time. Connected words, this should be separated by a pause. Continuous speech, these are the natural way of speech and hence it is difficult to develop a system that makes use of continuous speech. Spontaneous speech, it makes use of all the above types of utterances. The authors also describe the two types of speaker models those are: Speaker dependent models, designed for a particular speaker and it gives more accurate results, less costly and easy to develop but it is not flexible where as the other type of model is Speaker independent model, designed for variety of speakers and it gives less accurate results, more costly and difficult to develop but it is more flexible compared to that of speaker dependent models. The authors also described the phases of speech recognition technique, the first phase is analysis and it makes use of three techniques Segmentation analysis, Sub-segmental analysis and Supra-segmental analysis, it is found that supra-segmental analysis gives accurate analysis of frames ranging from 100-300ms. The second phase is speech feature extraction phase it consists of the following techniques namely, Linear Predictive Loading, Mel-Frequency Cepstral Coefficients (MFCC). The third phase is modeling phase and the techniques used in this phase are Acoustic-phonetic approach, Pattern recognition approach, Knowledge based approach, Statistical based approach and Dynamic Time Wrapping (DTW). The last phase is matching phase it uses two techniques namely, whole-word matching and sub-word matching, the whole-word matching technique is more efficient compared to that of sub-word matching. The authors come to a conclusion that HMM approach along with MFCC feature is more advantageous and offer accurate results. In [5], authors describe the above mentioned techniques that are used for speech recognition systems. The authors conclude that HMM and MFCC provide an accurate result.

In [6], authors describe the development of speech recognition system using techniques like MFCC, vector quantization and HMM. The speaker identification phase makes use of MFCC along with vector quantization. The speaker recognition phase makes use of HMM. Overall 98% efficiency is obtained. The authors made use of MATLAB and found that the MFCC along with distance minimum algorithm gave 95% of accurate result and the HMM algorithm gave the accurate result for isolated words. Overall these three techniques gave a 98% of accurate result.

In [7], authors have used MFCC technique for feature extraction and HMM for pattern recognition along with various techniques and found out that MFCC is best suitable for feature extraction and HMM are best suited for pattern recognition.

In [8], authors have used the technique of ANN i.e. Artificial Neural Networks to improve performance of speech recognition through a model known as ANN-HMM for large vocabulary speech recognition systems. These techniques gave high accuracy and less error rate and can develop robust systems.
In [9], authors gave a conclusion that HMM can be used for pattern recognition and can develop a user machine interface system. This system can be advantageous for the disable persons who cannot make use of keyboard and mouse. This system can also be used for the persons who do not feel comfortable with English language and can use their native language.

In [10], authors concentrate on feature extraction technique that makes use of MFCC for developing a human computer interface system for Marathi language. It is found that MFCC is used widely for feature extraction and GHM and HMM are best modeling techniques. It provides more accurate results.

In [11], authors made use of phonetic lattices generation technique for efficient retrieval of audio based on phone sequence queries. It makes use of inverted index structure that acts as features. This approach is helpful in extracting small number of audios hence its efficiency is very less.

In [12], authors use dynamic spectral sub band centroid technique for feature extraction. This technique works well in noisy environments and is more efficient as compared to that of MFCC technique.

III CONCLUSION

This paper presents some of the existing techniques of speech recognition. Authors have implemented the different techniques on speech recognition. Authors have also introduced new techniques which are capable of reducing errors in order to achieve robustness. These techniques are used for the systems that work on voice commands. These techniques give more accurate results.

REFERENCES

[1] Neeraj Kaberpanthi and Ashutosh Datar, "Speaker Independent Speech Recognition using MFCC with Cubic-Log Compression and VQ Analysis," International Journal of Computer Applications (0975 – 8887), Volume 95, No.26, JUNE 2014.

[2] Miss Himanshu, Sarbjit Kaur and Vikas Chaudhary, "Literature Survey on Automatic Speech Recognition System," International Journal of Advanced Research in Computer Science and Software Engineering, Volume 4, Issue 7, JULY 2014, pp 398-402.

[3] Bhavneet Kaur, "Major Challenges of Voice Command Recognition Technique," International Journal of Scientific & Engineering Research, Volume 5, Issue 8, AUGUST-2014.

[4] Om Prakash Prabhakar and Navneet Kumar Sahu, "A Survey On: Voice Command Recognition Technique," International Journal of Advanced Research in Computer Science and Software Engineering, Volume 3, Issue 5, MAY 2013, pp. 576-585.

[5] Harpreet Singh and Ashok Kumar Bathla, "A Survey on Speech Recognition," International Journal of Advanced Research in Computer Engineering & Technology (IJARCET), Volume No. 2, Issue No. 6, JUNE 2013.

[6] Suma Swamy and K.V Ramakrishnan, "AN EFFICIENT SPEECH RECOGNITION SYSTEM," Computer Science & Engineering: An International Journal (CSEIJ), Vol. 3, No. 4, AUGUST 2013.

[7] Preeti Saini and Parneet Kaur, "Automatic Speech Recognition: A Review," International Journal of Engineering Trends and Technology-Volume4, Issue2- 2013.

[8] Wiqas Ghai and Navdeep Singh, "Literature Review on Automatic Speech Recognition," International Journal of Computer Applications (0975 – 8887), Volume 41– No.8, MARCH 2012.

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

[10] Santosh K.Gaikwad, Bharti W.Gawali and Pravin Yannawar, "A Review on Speech Recognition Technique," International Journal of Computer Applications (0975 – 8887), Volume 10– No.3, NOVEMBER 2010.

[11] Olivier Siohan and Michiel Bacchiani, "Fast Vocabulary—Independent Audio Search Using Path—Based Graph Indexing," INTERSPEECH SEPTEMBER 4—8 2005.

[12] Jingdong Chen, Yiteng (Arden) Huang, Qi Li and Kuldip K. Paliwal, "Recognition of Noisy Speech Using Dynamic Spectral Subband Centroids," IEEE SIGNAL PROCESSING LETTERS, VOL. 11, NO. 2, FEBRUARY 2004..