International Journal of Engineering Science Invention (IJESI)
ISSN (Online): 2319 – 6734, ISSN (Print): 2319 – 6726
www.ijesi.org || PP. 83-87

A Review Regarding Speech Recognition by Machine

S.Kothbul Zeenath Banu[1], D.Viji[2].
[1,2] *(Assistant Professor, Krishnasamy College of Science, Arts and Management for Women, Cuddalore , TamilNadu, India)*

Abstract: *This paper presents a brief survey on Automatic Speech Recognition and discusses the major themes and advances made in the past years of research. After years of research and development the automatic speech recognition remains one of the most important research challenges. 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 Databases etc..The main aim of this review paper is to summarize and compare some of the well known methods used in various stages of speech recognition system*

Keywords: *Automatic Speech Recognition, Acoustic Phonetic, Pattern Recognition, Feature extraction, Speech Databases.*

I. Introduction

Speech Recognition (is also known as Automatic Speech Recognition (ASR), or computer speech recognition) is the process of converting a speech signal to a sequence of words, by means of an algorithm implemented as a computer program.

II. Basic Model Of Speech Recognition

Acoustic model is used to model the statistics of speech features for each speech unit of the language such as a phone or a word. Figure 1.shows the basic block diagram of a speech recognition system. As can be seen from Figure 1, acoustic models are required to analyze the speech feature vectors for their acoustic content. The acoustic models can be a template of the speech unit to be modeled. During recognition, an unknown word can be recognized by simply comparing it against all known templates and finding the closest match. But templates cannot model acoustic variability. Hence, acoustic models basically build using the probability distribution over the acoustic space. Probability distributions can be modeled by using both parametric and non-parametric techniques.

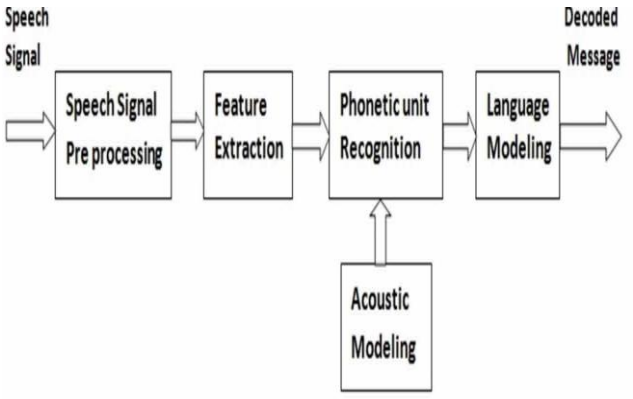

Fig 1.

The standard approach to large vocabulary continuous speech recognition is to assume a simple probabilistic model of speech production whereby a specified word sequence, W, produces an acoustic observation sequence Y, with probability P(W,Y). The goal is then to decode the word string, based on the acoustic observations equence, so that the decoded string has the maximum a posteriori (MAP) probability.

^
P(W/A)= arg maxw
W
P(W/A) . (1)

Using Baye s rule, equation (1) can be written as

P(W/A)=P(A/W)P(W) . (2)
P(A)
Since P(A) is independent of W, the MAP decoding rule
of equation(1) is
^ ^
W=argmaxw P(A/W)P(W) .. (3)
The first term in equation (3) P(A/W), is generally called the acoustic model, as it estimates the probability of a sequence of acoustic observations, conditioned on the word string. Hence P(A/W) is computed.

III. Various Methods Of Speech Recognition

Speech recognition systems can be separated in several different classes by describing what types of utterances they have the ability to recognize. These classes are classified as the following:

3.1 Isolated Word

Isolated word recognizers usually require each utterance to have quiet on both sides of the sample window. It accepts single words or single utterance at a time. These systems have "Listen/Not-Listen" states, where hey require the speaker to wait between utterances. Isolated Utterance might be a better name for this class.

3.2 Connected Words

Connected word systems are similar to isolated words, but allows separate utterances to be 'run-together' with a minimal pause between them.

3.3 Continuous Speech

Continuous speech recognizers allow users to speak almost naturally, while the computer determines the content.(Basically, it's computer dictation). Recognizers with continuous speech capabilities are some of the most difficult to create because they utilize special methods to determine utterance boundaries.

3.4 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 features such as words being run together, "ums" and"ahs", and even slight stutters

IV. Approaches To Speech Recognition

Basically there exist three approaches to speech recognition.
They are

- Acoustic Phonetic Approach.
- Pattern Recognition Approach.
- Artificial Intelligence Approach**.**

4.1 Acoustic phonetic approach

The earliest approaches to speech recognition were based on finding speech sounds and providing appropriate labels to these sounds. This is the basis of the acoustic phonetic approach. which postulates that there exist finite, distinctive phonetic units (phonemes) in spoken language and that these units are broadly characterized by a set of acoustics properties that are manifested in the speech signal over time. The first step in the acoustic phonetic approach is a spectral analysis of the speech combined with a feature detection that converts the spectral measurements to a set of features that describe the broad acoustic properties of the different phonetic units. The next step is a segmentation and labeling phase in which the speech signal is segmented into stable acoustic regions, followed by attaching one or more phonetic

4.2 Pattern Recognition approach

The pattern-matching approach involves two essential steps namely, pattern training and pattern comparison. The essential feature of this approach is that it uses a well formulated mathematical framework and establishes consistent speech pattern representations, for reliable pattern comparison, from a set of labeled training samples via a formal training algorithm. A speech pattern representation can be in the form of a speech template or a statistical model and can be applied to a sound (smaller than a word), a word, or a phrase. In the pattern-comparison stage of the approach, a direct comparison is made between the unknown speeches (the speech to be recognized) with each possible pattern learned in the training stage in order to determine the

identity of the unknown according to the goodness of match of the patterns. The pattern-matching approach has become the predominant method for speech recognition in the last six decades

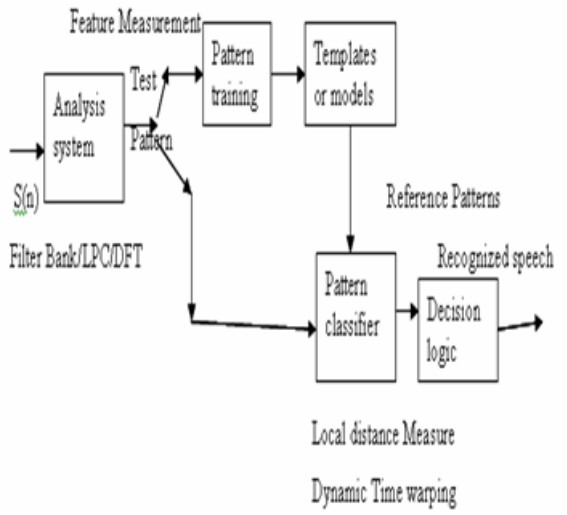

Fig 2.

4.3 Artificial Intelligence approach (Knowledge Based approach)

The Artificial Intelligence approach is a hybrid of the acoustic phonetic approach and pattern recognition approach. In this, it exploits the ideas and concepts of Acoustic phonetic and pattern recognition methods. Knowledge based approach uses the information regarding linguistic, phonetic and spectrogram. Some speech researchers developed recognitions ystem that used acoustic phonetic knowledge to develop classification rules for speech sounds. While template based approaches have been very effective in the design of a variety of speech recognition systems; they provided little insight bout human speech processing, thereby making error analysis and knowledge-based system enhancement difficult. On the other hand, a large body of linguistic and phonetic literature provided insights and understanding to human speech processing. In its pure form, knowledge engineering design involves the direct and explicit incorporation of expert speech knowledge into a recognition system. This knowledge is usually derived from careful study of spectrograms and is incorporated using rules or procedures. Pure knowledge engineering was also motivated by the interest and research inexpert systems. However, this approach had only limited success, largely due to the difficulty in quantifying expert knowledge. Another difficult problem is the integration of many levels of human knowledge phonetics, phonotactics, lexical access, syntax, semantics and pragmatics. Alternatively, combining independent and asynchronous knowledge sources optimally remains an unsolved problem. In more indirect forms, knowledge has also been used to guide the design of the models and algorithms of other techniques such as template matching and stochastic modeling. This form of knowledge application makes an important distinction between knowledge and algorithms Algorithms enable us to solve problems. Knowledge enable the algorithms to work better. This form of knowledge based system enhancement has contributed considerably to the design of all successful strategies reported.

V. Feature Extraction

In speech recognition, the main goal of the feature extraction step is to compute a parsimonious sequence of feature vectors providing a compact representation of the given input signal. The feature extraction is usually performed in three stages.

The first stage is called the speech analysis or the acoustic front end. It performs some kind of spectro temporal analysis of the signal and generates raw features describing the envelope of the power spectrum of short speech intervals. The second stage compiles an extended feature vector composed of static and dynamic features. Finally, the last stage(which is not always present) transforms these extended feature vectors into more compact and robust vectors that are then supplied to the recognizer. Although there is no real consensus as to what the optimal feature sets should look like, one usually would like them to have the following properties: they should allow an automatic system to discriminate between different through similar sounding speech sounds, they should allow for the automatic creation of acoustic models for these sounds without the need for an excessive amount of training data, and they should exhibit statistics which are largely invariant across speakers and speaking environment.

VI. Speech Databases

Speech databases have a wider use in Automatic Speech Recognition. They are also used in other important applications like, Automatic speech synthesis, coding and analysis including speaker and language identification and verification. All these applications require large amounts of recoded database. Different types of databases that are used for speech recognition applications are discussed along with its taxonomy.

6.1 Taxonomy of Existing Speech Databases

The intra-speaker and inter-speaker variability are important parameters for a speech database. Intra-speaker variability is very important for speaker recognition performance. The intra speaker variation can originate from a variable speaking rate, changing emotions or other mental variables, and inenvironment noise. The variance brought by different speakers is denoted inter-speaker variance and is caused by the individual variability in vocal systems involving source excitation, vocal tract articulation, lips and/or nostril radiation.If the inter-speaker variability dominates the intra-speaker variability, speaker recognition is feasible. Speech databases are most commonly classified into single-session and multisession. Multi-session databases allow estimation of temporal intra-speaker variability. According to the acoustic environment, databases are recorded either in noise free environment, such as in the sound booth, or with office/home noise. Moreover, according to the purpose of the databases are designed for developing and evaluating speech recognition, for instance TIMIT, and some are specially designed for speaker recognition, such as SIVA,Polycost and YOHO. Many databases were recorded in one native language of recording subjects; however there are also multi-language databases with non-native language of speakers, in which case, the language and speech recognition become the additional use of those databases.

VII. Conclusions

Speech is the primary, and the most convenient means of communication between people. Whether due to technological curiosity to build machines that mimic humans or desire to automate work with machines, research in speech and speaker recognition, as a first step toward natural human-machine communication, has attracted much enthusiasm over the past five decades. we have also encountered a number of practical limitations which hinder a widespread deployment of application and services. In most speech recognition tasks, human subjects produce one to two orders of magnitude less errors than machines. There is now increasing interest in finding ways to bridge such a performance gap. What we know about human speech processing is very limited. Although these areas of investigations are important the significant advances will come from studies in acoustic phonetics, speech perception, linguistics, and psycho acoustics.Future systems need to have an efficient way of representing, storing, and retrieving knowledge required for natural conversation. This paper attempts to provide a comprehensive survey of research on speech recognition and to provide some year wise progress to this date. Although significant progress as been made in the last two decades, there is still work to be done, and we believe that a robust speech recognition system should be effective under full variation in: environmental conditions, speaker variability etc. Speech Recognition is a challenging and interesting problem in and of itself. Speech recognition is one of the most integrating areas of machine intelligence, since, humans do a daily activity of speech recognition. Speech recognition has attracted scientists as an important discipline and has created a technological impact on society and is expected to flourish further in this area of human machine interaction. We hope this paper brings about understanding and inspiration amongst the research communities of ASR.

References

[1]. Sadaoki Furui, 50 years of Progress in speech and Speaker Recognition Research , *ECTI Transactions on Computer and Information Technology,Vol.1. No.2November 2005.*

[2]. K.H.Davis, R.Biddulph, and S.Balashek, *Automatic Recognition of spoken Digits, J.Acoust.Soc.Am.,24(6):637-642,1952.*

[3]. H.F.Olson and H.Belar, *Phonetic Typewriter ,J.Acoust.Soc.Am.,28(6):1072-1081,1956.*

[4]. D.B.Fry, *Theoritical Aspects of Mechanical speech Recognition , and P.Denes, The design and Operation of the Mechanical Speech Recognizer at Universtiy CollegeLondon, J.British Inst. Radio Engr., 19:4,211-299,1959.*

[5]. J.W.Forgie and C.D.Forgie, *Results obtained from avowel recognition computer program , J.A.S.A.,31(11),pp.1480-1489.1959.*

[6]. J.Suzuki and K.Nakata, *Recognition of Japanese Vowels Preliminary to the Recognition of Speech, J.RadioRes. Lab37(8):193-212,1961.*

[7]. T.Sakai and S.Doshita, *The phonetic typewriter, information processing 1962 , Proc.IFIP Congress, 1962.*

[8]. K.Nagata, Y.Kato, and S.Chiba, *Spoken Digit Recognizer for Japanese Language , NEC Res.Develop.,No.6,1963.*

[9]. T.B.Martin, A.L.Nelson, and H.J.Zadell, *Speech Recognition b Feature Abstraction Techniques ,Tech.Report AL-TDR-64-176,Air Force AvionicsLab,1964.*

[10]. Frederick Jelinek , *Statistical methods for Speech Regognition System ,*

[11]. Herre A. Bourland ,Nelson Morgan ,*Connectionist of Speech Recognition A Hybrid Approach,*
[12]. Chin-HuiKee,Frank k.Soong, Kuldip K. Paliwal. *Automatic Speech and Speaker Recognition ,*
[13]. F.Itakura, *Minimum Prediction Residula Applied to Speech Recognition ,IEEE Trans.Acoustics,Speech,Signal Proc., ASSP-23(1):67-72,February 1975.*
[14]. H.Sakoe and S.Chiba, *Dynamic programming algorithm 200 http://sites.google.com/site/ijcsis/optimization for spoken word recognition ,IEEE Trans Acoustics, Speech, Signal Proc., ASSP-26(1).pp. 43-49,1978.*
[15]. H.Sakoe and S.Chiba, *Dynamic Programming Algorithm Optimization for Spoken Word Recognition, IEEE Trans.Acoustics, Speech, Signal Proc.,ASSP-26(1):43-49,February 1978.*
[16]. H.Sakoe and S.Chiba, *Dynamic Programming Algorithm Optimization for Spoken Word Recognition, IEEE Trans.Acoustics, Speech, Signal Proc.,ASSP-26(1):43-49,February 1978.*