Automatic speech recognition

Komal Kawaduji Balki, Dr. Dinesh Vittalrao Rojatkar
Student, Asst.Professor
Dept. Of Electronics & Telecommunication
Government College of engineering
Chandrapur (MS), India

***Abstract*- in today's world, Automatic Speech Recognition is very important & popular. Speech recognition is the process of converting spoken words into text. Designing a machine that mimics behavior of human, particularly the capability of speaking naturally and responding properly to spoken language, has intrigued engineers and scientists for centuries.**

Automatic speech recognition system today find widespread application in tasks that require a human machine interface, like automatic call processing in telephone network and query based information systems do things like provide new updated information, weather reports, stock price quotations, etc. Automatic Speech Recognition is an emergent field in research. This paper gives basic information about speech recognition, knowledge and idea on automated speech recognition and techniques used behind this method. Various researches are going on this automatic speech recognition field where speech is identified and then converted into text.

***Keywords:* Language Model, Hidden Markov Model, Feature extraction, Feature Matching, Vector Quantization (VQ).**

INTRODUCTION

Speech Recognition is the process of converting an acoustic signal, received by a microphone or a telephone, to a set of words. It is also known as Automatic Speech Recognition, speech to text, Computer Speech Recognition. The speech recognized words can be an end in themselves, as for applications such as commands & control, data entry. They can also serve as the input to further linguistic processing in order to achieve speech understanding.

The Speech is the most common and primary mode of communication among human beings. Speech is the most natural and efficient form of exchanging information between humans. Human speech convey much more information such as gender, emotion and identity of the speaker. Speech Recognition can be defined as the process of converting speech signal to a sequence of words. The objective of speech recognition is to determine which speaker is present based on the individual's characterization. The several techniques have been proposed for compensating the mismatch occurred among the testing and training sessions.

Speech is the most natural form of human communication and speech processing has been one of the most important areas of the signal processing. ASR system has made it possible for computer to follow human voice commands and understand human languages. The main purpose of speech recognition area is to develop systems and techniques for speech input to machine. Speech recognition is the ability of a machine to identify different words and phrases in spoken language and convert them to a machine readable or understandable format. Many speech recognition applications, such as simple data entry, voice dialing and speech-to-text are in existence today. Automatic speech recognition (ASR) systems were first made in the 1950s. These speech recognition systems tried to apply a set of grammatical and syntactical rules to identify speech. If the spoken words adhered to a certain rules, the system could recognize the words. However, human language has numerous exceptions to its own rules. The way of words are spoken can be vastly altered by accents, dialects and mannerisms. Therefore, today's speech recognition systems increasingly rely on statistical methodology, dynamic time warping, moving away from approaches such as template matching, and non probabilistically motivated distortion measures that were initially proposed . The statistical model that dominates the field today is the hidden Markov Model.

RELEVANT ISSUES OF ASR DESIGN

The main issues on which recognition accuracy depends have been presented in the bellow table 1.

Table 1: Relevant issues of ASR design.

Environment	Type of noise; signal/noise ratio; working conditions
Speakers	Speaker dependence/independence Sex, Age; physical and physical state
Transducer	Microphone; telephone
Vocabulary	Characteristics of available training data; Specific or generic vocabulary

LITERATURE SURVEY:

OVERVIWE:

The remainder of this survey is structured as follows: Section 1 discusses the different approaches to speech recognition. This is followed in Section 2 by a summary of the speech recognition process. Section 3 briefly discusses hidden Markov Models (HMM) and their application to speech processing. Section 4 looks at some research that has been done on speaker independent systems to handle various dialects. Finally, Section 5 summrises and concludes the paper.

Speech is the primary means of communication among humans. The natural ease with which we communicate through conversations masks the complexity of language. The diversity in language arises from given factors:

- Geographical - there are over thousands of languages comprising different various dialects.
- Cultural - the level of education and other things has a strong influence on speaking style.
- Physical - each individual‘s sound box have slightly different shapes hence differentiates between others in their speaking style and clarity.

1 VARIOUS APPROACHES TO SPEECH RECOGNITION:

The three broad approaches to automatic speech recognition are the acoustic-phonetic, pattern recognition and artificial intelligence (AI) approaches.The acoustic phonetic approach to speech recognition has not been very successful in practical speech recognition systems.

i).Pattern Recognition approach

This method has gained its popularity in the recent years in automatic speech recognition. Pattern recognition approach includes two basic steps they are

1. Pattern comparison

2. Pattern training

In this approach a direct comparison is made amongs the spoken words and the patterns learned in the training stage.

ii). The acoustic-phonetic approach

In acoustic phonetic approach, speech sounds were found and these sounds are labelled to produce text form. In this method, features are extracted from speech based on various classification like ratio of maximum and minimum frequencies, voiced and unvoiced classification. Acoustic phonetic approach followed the following sequence.

1. Spectral analysis

2. Feature detection

3. Segmentation and labeling

4. Recognizing valid word

iii).Learning based approaches

To overcome the disadvantage of the Hidden Markov Models machine learning methods which was introduced in neural networks and genetic algorithm learning based approaches has been taken. In learning based approaches, the machine can be learned automatically through emulations or evolutionary process.

iv). Knowledge based approaches

The guidance or information should be taken from an expert knowledge about variations in speech is hand coded into a system. This approach gives the advantage of apparent modeling but this situation is difficult or hard to obtain and cannot used successfully. Knowledge based approach uses the information regarding to linguistic, spectrogram and phonetic. Vector Quantization (VQ) is applied to automatic speech recognition. It is useful for speech coders, i.e. efficient data reduction.

v).Artificial intelligence approach

The artificial intelligence (AI) approach coordinates the recognition procedure according to the person who applies it. The intelligence of any person such as visualizing, analyzing etc. is used for making a decision on the measured acoustic features.

2 SPEECH RECOGNITION PROCESS:

When a person speaks in microphone, compressed air from the lungs is forced through the vocal tract as a sound wave that varies as per the variations in the lung pressure and the vocal tract. This sound wave is interpreted as speech when it falls up on a person's ear. Speech waveforms have the characteristic of being continuous in both time and amplitude.

2.1 Fluent speech Recognition

Fluent speech recognition is a more complicated task than isolated word recognition. In this case the task is to recognize a continuous string of words from the vocabulary.

2.2 Feature Extraction and Pattern Recognition

The input into an automatic speech recognition system is the speech signal. The two major tasks involved in speech recognition are feature extraction and pattern recognition.

2.2.1 Feature Extraction

In all speech recognition systems the first step in the process is signal processing. Initially a spectral and temporary analysis of the speech signal is performed to give observation vectors which can be used to train the HMMs. The extraction of the features of the parameters which represent an acoustic signal is an important task to produce a best recognition performance. The capability of this method is important for the next method since it affects its behavior. Various feature extraction methods available with their features.

i) In Principal Component analysis (PCA), It uses non linear feature extraction method and eigenvector based.

ii) In linear Discriminate Analysis (LDA), it depends on non linear feature extraction method, gives supervised linear map and eigenvector based. This method is better than principal component analysis.

iii) The linear predictive coding uses static feature extraction method which has lower order coefficient. It is used for extracting features at the lower order.

iv) In Mel-frequency cestrum. It has the property that the Power spectrum is scaled by performing Fourier analysis.

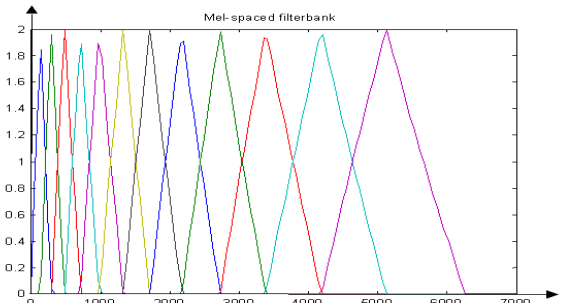

Fig.1 Mel Frequency Cepstral Coefficients.

v) The wavelength analysis gives better time resolution than Fourier transform because It replaces the fixed BW of Fourier transform with one proportional to frequency which allow better time resolution at large frequencies than Fourier Transform.

FEATURE MATCHING

Many techniques used in feature extraction such as dynamic time wrapping (DTW), vector quantization (VQ), LBG, etc. Each technique has own feature matching function and specification.

DTW:

Dynamic time wrapping (DTW) is an algorithm for measuring similarity between two temporary sequences which may vary in time or speed. DTW is a method that calculates an optimal match among two given sequences. DTW has been applied to temporary sequences of audio, video and graphics data indeed, any data which can be turned into a linear sequence can be analyzed with dynamic time wrapping.

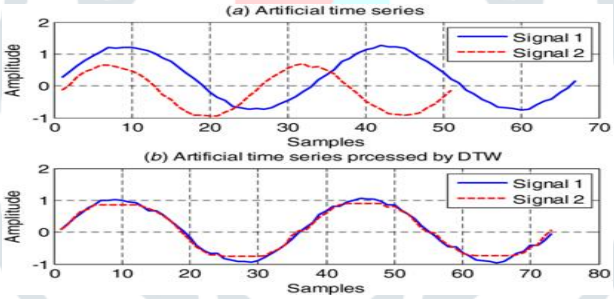

Fig.2.Dynamic Time Wrapping of two speech signal.

2.2.2 Pattern Recognition

The pattern recognition process consists of training and testing. During training, a model of vocabulary word must be created. Each model consists of a features extracted from the speech signal. During testing, a similar model is created for the unknown word. The patterned recognition algorithm compares the model of the known world with the models of known words and selects the word whose model score is highest.

Any speech recognition system involves five major steps:

1 Converting sound into electrical signal.

2 Background noise removal.

3 Breaking up words into phonemes.

4 Matching & choosing character combination.

5 Language analysis.

6 After that, there will be grammar check. It tries to find out whether or not the combination of words any sense. That is there will be a grammar check package.

7 Finally the numerous words constitution the speech recognition programs come with their own word processor, some can work with other different word processing package like MS word and word perfect.

3 HIDDEN MARKOV MODELS IN SPEECH RECOGNITION:

HMM can be used to model an unknown process that produces a sequence of observable outputs at discrete intervals, where the outputs are members of some finite alphabet. These models are called hidden Markov models (HMMs) precisely because the state sequence that produced the observable output is not known it's hidden.

HMM is represented by a set of states, vectors defining transitions among certain pairs of those states, probabilities that apply to state to state transitions, sets of probabilities characterizing observed output symbols, and also initial conditions. In the three state diagram of HMM model where states are denoted by nodes and transitions by directed arrows (vectors) between nodes. In hidden markov model the circles represent states of the speaker's vocal system specific configuration of tongue, lips, etc that produce a given sound. The arrows represent possible transitions from one state to another. Transition may occur only from the tail to the head of a vector. A state can have more than one transition leaving it and more than one leading to it.

4 Speaker-Independent SR Systems:

Speaker independent speech recognition software is designed to recognize anyone's voice, so no training is involved in speaker independent speech recognition system. This means it is the only unaffected option for applications such as interactive voice response systems where businesses cannot ask callers to read pages of text before using this system. The downside is that speaker independent SR software is generally less exact than speaker dependent software. Speaker dependent software is used more widely in dictation software, where only single person will use the system and there is a need for a large grammar.

The LumenVox Speech Engine, which powers all of our SR software, is speaker independent. It is not orthography software, it is not capable of recognizing an immensurable number of words at once, and it is not the same as voice recognition. It is designed for recognizing specific information, primarily by callers into a telephone IVR. It works well as a call router, auto consequential or any other application where designers have an idea what sort of words a speaker is likely to say. Because the audio we use to build the models contains hundreds of speakers, the Engine has a wide variety of voices it can recognize. This is what makes it speaker–independent.

AREAS OF APPLICATION:

Growing interest researches in the development of new techniques for different phases of automatic speech recognition systems have lead to the applications of Automatic Speech Recognition in different fields such as:

i. Command and Control Systems:

- Automated Call-Type Recognition
- Call Distribution by Voice commands

ii. Demotic appliance control and content based spoken audio search.

iii. Automated data entry: e.g. entering a credit card numbers.

iv. Preparation of structured documents: e.g. radiology report.

v. Speech to text processing: e.g. word processors or emails.

vi. Direct voice input: e.g. In aircraft cockpits

vii. Home automation

viii. Transcription of speech to mobile text message.

CONCLUSION:

There has been much progress in the field of automatic speech recognition since it's in the 1950s. Various approaches to ASR have been mentioned. Current speech recognition systems are generally based on hidden Markov models (HMMs) as these models have lead to the best results in speech recognition systems thus far. Although HMMs have been very successful, there are a few limitations of the models that were mentioned. The need for, and usefulness of speaker adaptation in speaker independent systems was highlighted. We are a long way from achieving perfect speech recognition and there is much research still to be done in the field of automatic speech recognition.

REFERENCES:

[1] J. Vepa and H. Bourlard, "Improving Speech Recognition Using a Data-Driven Approach," Proc. Eurospeech, Lisbon, Spet 2005, pp. 3333-3336.

[2] L.R Rabiner. A Tutorial on Hidden Markov Models(HHMs) and Selected Applications in Speech Recognition. Proceedings of the IEEE. 77(2):257-286. 1989

[3] Abdul Kadir K(2010), "Recognition of Human Speech using q-Bernstein Polynominals," International Journal of Computer Application, Vol.2 – No.5, pp.22-28.

[4] IEEE Information Technology (2010), page 557-562. Rajesh Kumar Aggarwal and M. Dave, "Acoustic modeling problem for automatic speech recognition system: advances and refinements Part (Part II)", Int J Speech Technol, pp. 309–320, 2011

[5] Grosjean, F. (1980) "Temporal variables within and between languages'" in towords a Cross-Linguistic Assessment of Speech Production, edited by H. W. Dechert and M. Raupach(Lang, Frankfurt), pp.39-53.

[7] Friesen, L. M., and Wang, X. (2001), "Speech recognition in noise as a function of the number of spectral channels: Comparison of acoustic hearing and cochlear implants,"J. Acoust. Soc. Am. 110(2), 1150–1163.

[8] Rohini B. Shinde and V.P. Pawar(2012), "A Review on acoustic phonetic approach for Marathi Speech Recognition". International Journal of Computer Applications 59(2): 40-44.

[9] Y. Linde, A. Buzo and R. M. Gray, "An algorithm for vector quantizer design," *IEEE Trans. on Communication*, Vol. COM-28, pp. 84-95, Jan. 1980.