Speech Recognition

Anjali I.P
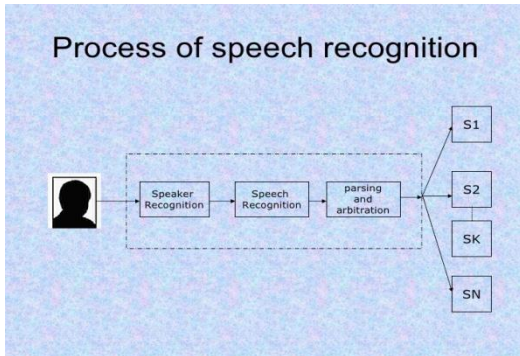Department Of Computer Science
Carmel College Mala
Thrissur, Mala

Sherseena P. M
Department Of Computer Science
Carmel College Mala
Thrissur,Mala

Abstract : **Today, voice and natural language processing are at the forefront of any human machine interaction environment. The chapter emphasizes the tremendous progress that has taken place in machine learning, statistical data-mining and pattern recognition approaches that can help in making speech interfaces more versatile and pervasive. The growing requirements of speech interfaces also warn against the impediments that may come in the way of successful implementation of acoustically robust natural interfaces. Finally, the chapter underlines the technical advances and research efforts to be undertaken for high performance real-time speech recognition that will completely change the way humans interact with their computing devices.**

INTRODUCTION

The field of computer science that deals with designing computer systems that can recognize spoken words .Note that voice recognition implies only that the computer can take dictation, not that it understands what is being said. Comprehending human languages falls under a different field of

computer science called natural language processing. A number of voice recognition systems are available on the market. The most powerful can recognize thousands of words.

However, they generally require an extended training session during which the computer system becomes accustomed to a particular voice and accent. Such systems are said to be speaker dependent.

Many systems also require that the speaker speak slowly and distinctly and separate each word with a short pause. These systems are called discrete speech systems. Recently, great strides have been made in continuous speech systems -- voice recognition systems that allow you to speak naturally. There are now several continuous-speech systems available for personal computers.

Because of their limitations and high cost, voice recognition systems have traditionally been used only in a few specialized situations. For example, such systems are useful in instances when the user is unable to use a keyboard to enter data because his or her hands are occupied or disabled. Instead of typing commands, the user can simply speak into a headset. Increasingly, however, as the cost decreases and performance improves, speech recognition systems are entering the mainstream and are being used as an alternative to keyboards.

This paper deals with the topic SPEECH RECOGNITION which can make a revolution in the years to come. Speech recognition acts as an interface between the user and the system. Its applications vary to the extent that it is a successful replacement for input devices like Keyboard ,mouse etc.

This paper contains information about Automatic Speech Recognition which decodes speech signals to phones, which is the basic building block of any word. Speech Recognition Systems are classified as Dependent and Independent Systems. Dependent systems recognize the sound generated by a single speaker whereas an Independent System recognizes sounds generated by multiple speakers. Speech recognition technologies allow computers equipped with a source of sound input, such as a microphone, to interpret human speech, e.g., for transcription or as an alternative method of interacting with a computer

Automatic Speech Recognition

Automatic speech recognition is the process by which a computer maps an acoustic speech signal to text. Automatic speech understanding is the process by which a computer maps an acoustic speech signal to some form of abstract meaning of the speech

Speaker dependent / adaptive / independent means

A speaker dependent system is developed to operate for a single speaker. These systems are usually easier to develop, cheaper to buy and more accurate, but not as flexible as speaker adaptive or speaker independent systems.

A speaker independent system is developed to operate for any speaker of a particular type (e.g. American English). These systems are the most difficult to develop, most

expensive and accuracy is lower than speaker dependent systems. However, they are more flexible.

A speaker adaptive system is developed to adapt its operation to the characteristics of new speakers. It's difficulty lies somewhere between speaker independent and speaker dependent systems.

VOCABULARY MEANS

The size of vocabulary of a speech recognition system affects the complexity, processing requirements and the accuracy of the system. Some applications only require a few words (e.g. numbers only), others require very large dictionaries (e.g. dictation machines). There are no established definitions, however, try

• Small Vocabulary - tens of words
• Medium Vocabulary - hundreds of words
• Large Vocabulary - thousands of words
• Very-Large Vocabulary - tens of thousands of words.

Continuous Speech and Isolated-word means

An isolated-word system operates on single words at a time - requiring a pause between saying each word. This is the simplest form of recognition to perform because the end points are easier to find and the pronunciation of a word tends not affect others. Thus, because the occurrences of words are more consistent they are easier to recognize.

A continuous speech system operates on speech in which words are connected together, i.e. not separated by pauses. Continuous speech is more difficult to handle because of a variety of effects. First, it is difficult to find the start and end points of words. Another problem is "co articulation". The production of each phoneme is affected by the production of surrounding phonemes, and similarly the start and end of words are affected by the preceding and following words. The recognition of continuous speech is also affected by the rate of speech (fast speech tends to be harder).

The Process of Speech Recognition

There are several approaches to automatic speech recognition:

• **Acoustic-Phonetic** -- This approach is based on the idea that all spoken words can be split up into a finite group of phonetic units. If all of these phonetic units can be characterized computationally, one should be able to figure out what phonetic units have been spoken, and then decode them into words.

• **Pattern Recognition** -- This approach uses a training algorithm to teach a recognizer about the patterns present in specific words. It is similar to the acoustic-phonetic approach, but rather than defining the patterns explicitly (as phonetic units), Hidden Markov Model(HMM) based pattern recognizer finds it's own set of patterns.

• **Artificial Intelligence** -- This approach mixes the previous two approaches by combining phonetic, syntactic, lexical, and/or semantic based analysis with pattern recognition.
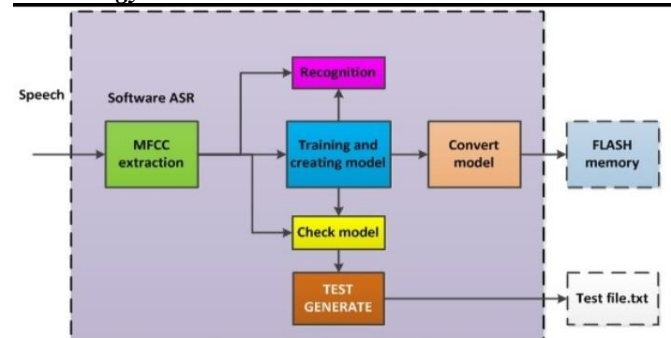
Speech Detection

The first task is to identify the presence of a speech signal. This task is easy if the signal is clear, however frequently the signal contains background noise. The signals obtained were in fact found to contain some noise. Two criterions are used to identify the presence of a spoken word. First, the total energy is measured, and second the number of zero crossings are counted. Both of these were found to be necessary, as voiced sounds tend to have a high total energy, but a low frequency, while unvoiced sounds were found to have a high frequency. Only background noise was found to have both low energy and low frequency. The method was found to successfully detect the beginning and end of the several words tested. Note that this is not sufficient for the general case, as fluent speech tends to have pauses, even in the middle of words (such as in the word 'acquire', between the 'c' and 'q'). In fact reliable speech detection is a difficult problem, and is an important part of speech recognition.

Blocking

The second task is blocking. Older speech recognition systems first attempted to detect where the phones would start and finish, and then block the signal by placing one phone in each block. However, phones can blend together in many circumstances, and this method generally could not reliably detect the correct boundaries. Most modern systems simply separate the signal into blocks of a fixed length. These blocks tend to overlap, so that phones which cross block boundaries will not be missed. Here is what a typical block might

Significance of Automatic Speech Recognition Technology

One of the modern applications that aim at achieving efficient transcription is automatic speech recognition software. This software is designed in such a way that it understands or recognizes human voice and converts the spoken words into text format in a matter of seconds.Accurate and timely documentation of relevant data proves to be a crucial element in different organizations. Especially in industries such as health care, quality of data and data integrity play a much more critical role. This necessitates verbatim and accurate transcripts accomplished through reliable and efficient processes. Here's where innovative technologies such as voice recognition play a vital role in assuring the reliability and timeliness of the documentation process.

Benefits of the Speech Recognition Software

Voice recognition technology is faster: speaking is normally faster than writing or typing among most of the individuals – speech recognition software offers to get words into documents without delay

Accuracy is fairly good: Although transcripts from automatic speech recognition software need to be proofed and checked for quality, its quality is fairly good

Hands-free, focused work: Voice recognition makes it possible for the dictating professional to focus on his or er core function, without the need for paying attention to the routine task of typing – you just need to dictate while even attending to your core activities

Spelling: Speech to text process relieves individuals from having to pay attention to spelling – being able to dictate directly into the digital device ensures that spelling errors are reduced considerably

Challenges Related to Speech Recognition Software

Automatic speech recognition software doesn't understand the complexities of the jargon – different industries use their own vocabulary and idioms that the software may not be able to understand fully

Accuracy is not reliable – transcripts, especially those related to medical industry, need to be of the maximum possible accuracy- which cannot be guaranteed in the case of automatic speech recognition software

Training is needed: the software needs to be trained to understand and recognize the voice of the dictator – transcribing voice data from more than one person is further difficult

CONCLUSION

Factors such as environmental changes and mild changes in appearence inpact the technology to a greater degree than many expect.For implimentations where the biometric system must verify and identify users reliably over time, facial scan be a very difficult,but not possible,technology to implement successfully.It hopes to create this software that can instandly translate two languages with atleast 90% accuracy.DARPA is also funding an R&D effort called TRANSTAC to enable the soldiers to communicate more effectively with civilian population in non-English-speaking countries.

REFERENCE

[1] https://en.wikipedia.org/wiki/Speech_recognition
[2] https://en.wikipedia.org/wiki/Speech_recognition
[3] https://www.answers.com/Q/Advantages_and_Disadvantages_of_Speech_Recognition