

Voice Computing: Technology for Next Technical Era

Amit Ashok Mokashi

Professor at Sharadchandra Pawar College of Engg., Otur (Pune)

Abstract

Voice computing is the newly proposed system for latest technical generation. Today's computer system is useful if the operating person is even unable to speak but there is no provision for the people who are handicap by other organs. As the name of project indicates we are going to implement the concept of computing totally based on voice especially for blind or handicap people. Because if anyone is mentally feet then the Weakness of physical organ must not be the limitation for learning or operating computer. To overcome this problem we design a system consists of following facility in a single unit.

-Controlling operating system

-Talking editor

-People choice

-Mail reader

-Security module

The concept of voice computing is useful in many areas like: for blind people, officers, students, housewives etc.

Key words: *Speech Recognition, Voice computing, speech, text*

1. Introduction

This paper presents a brief survey on Automatic Speech Recognition and discusses the major themes and advances made in the past 60 years of research, so as to provide a technological perspective and an appreciation of the fundamental progress that has been accomplished in this important area of speech communication. After years of research and development the accuracy of automatic speech recognition remains one of the important research challenges (e.g., variations of the context, speakers, and environment). 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 classifiers, database and performance evaluation. The problems that are existing in ASR and the various techniques to solve these problems constructed by various research workers have been presented in a chronological order. Hence authors hope that this work shall be a contribution in the area of speech recognition. The objective of this review paper is to summarize and compare some of the well known methods used in various stages of speech recognition system and identify research topic and applications which are at the forefront of this exciting and challenging field.

1.1 Problem definition

The problem statement can be given as 'We want to design a system which replaces the current computer system by a strong, protected and efficient system.' The efficiency can be measured in terms of reusability, platform dependability, manpower, various resources etc.

1.2 Aim

Voice computing is the application specially designed for physically challenged people or blind person for whom physical disorder is the only obstacle in handling current or traditional system.

1.3 Purpose

Actually the basic purpose of this project is to overcome the limitations of use of key board and mouse. And other purpose is to give flexibility to the administrator for handling system and provide all in one system software. Speech is the primary means of communication between people. Speech recognition, generation of speech waveforms, has been under development for several decades. Automatic speech Recognition is a process by which a computer takes a speech signal and Converts it into words. It is the

process by which a computer recognizes what a person Said. Keyboard, although a popular medium is not very convenient, as it requires a certain amount of skill for effective usage .A mouse on the other hand requires a good hand eye co-ordination. Physically challenged people find computer difficult to use. Partially blind people find reading from a monitor difficult. All these constraints have to be eliminated. Speech interface help us to tackle these problems. The objective is to trap human voice in a digital computer and decode it into corresponding text. Speech recognition can be defined as the process of converting an acoustic signal, captured by a microphone or a telephone, to a set of words. When two people speak to one another, they both recognize the words and the meaning behind them. Computers, on the other hand, are only capable of the first thing: they can recognize individual words and phrases, but they don't really understand speech in the same way as humans do. Computer recognizes the command and software tells the computer what to do when that command is recognized. 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.

Speech recognition technology has made it possible for computer to follow human voice commands and understand human languages. The main goal of speech recognition area is to develop techniques and systems for speech input to machine. Speech is the primary means of communication between humans. These recognizers can be used on the construction of speech-based applications, but with some limitations due to the difficulty of integration with other software applications

Automatic Speech Recognition

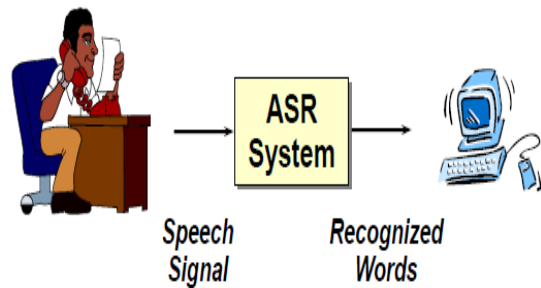
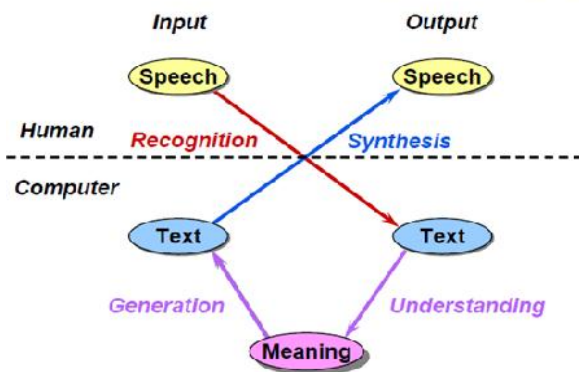


Fig 2: Voice Computing using Speech recognition system

Speech is the primary means of communication between people. Speech recognition, generation of speech waveforms, has been under development for several decades. Automatic speech Recognition is a process by which a computer takes a speech signal and Converts it into words. It is the process by which a computer recognizes what a person Said. Keyboard, although a popular medium is not very convenient, as it requires a certain amount of skill for effective usage .A mouse on the other hand requires a good hand eye co-ordination. Physically challenged people find computer difficult to use. Partially blind people find reading from a monitor difficult. All these constraints have to be eliminated. Speech interface help us to tackle these problems. The objective is to trap human voice in a digital computer and decode it into corresponding text. Speech recognition can be defined as the process of converting an acoustic signal, captured by a microphone or a telephone, to a set of words.

When two people speak to one another, they both recognize the words and the meaning behind them. Computers, on the other hand, are only capable of the first thing: they can recognize individual words and phrases, but they don't really understand speech in the same way as humans do. Computer recognizes the command and software tells the computer what to do when that command is recognized.

Communication via Spoken Language



and possible license restrictions.

Fig 1: Speech recognition system

Fig 3: Voice Computing software: Module I

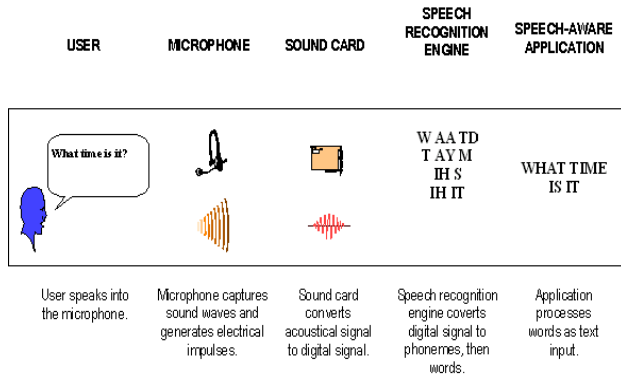
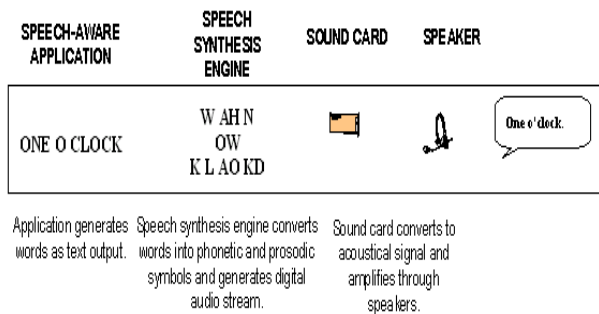


Fig 4: Voice Computing software: Module II



Today, speech recognition software can even be downloaded for free or come standard with cell phones. These types of software basically work by taking natural language, spoken words or commands and translating them into a language easily understood by the computer. This occurs when the computer picks up your voice through a microphone and then converts your voice into an analog signal; it then is processed by your computers sound card and from there is translated into a binary code so that your computer can understand it. Through that process the software either turns the voice to text or uses it to carry out the consumers command.

Speech recognition software can help a wide range of people from the busy teenager, to the disabled. Disabled individuals, who are unable to operate computers through mouse or keyboard use, can now control their computers with ease and confidence. Software is now available that supports completely hands-free controlling from everything to computer games to sending important business emails. The option to ask your computer how to perform tasks can help those who

have trouble using computer. Speech recognition software can be incorporated into all of our lives. We have all seen the commercials for speech recognition software, a depiction of a college student writing an entire paper just by speaking into their PC, or a busy mom asking Siri to set a reminder for an important event. Speech recognition is a part of many people's everyday life and maybe it is time for you to discover how it can make your life easier.

While the cool features of each software vary from provider to provider, below you'll see some common features we've found are available as a core function of almost every software provider out there. Consider the points below the 'baseline for determining the best':

Supports Multiple Languages

Simple Dictation

Grammar Checks

Easy Installation and Set up

Ability to Understand a Wide Range of Accents and Dialects
Commands Compatible

2. Proposed system

Language is man's most important means of communication and speech its primary medium. Speech provides an international forum for communication among researchers in the disciplines that contribute to our understanding of the production, perception, processing, learning and use. Spoken interaction both between human interlocutors and between humans and machines is inescapably embedded in the laws and conditions of Communication, which comprise the encoding and decoding of meaning as well as the mere transmission of messages over an acoustical channel. Here we deal with this interaction between the man and machine through synthesis and recognition applications. The paper dwells on the speech technology and conversion of speech into analog and digital waveforms which is understood by the machines. Speech recognition, or speech-to-text, involves capturing and digitizing the sound waves, converting them to basic language units or phonemes, constructing words from phonemes, and contextually analyzing the words to ensure correct spelling for words that sound alike.

We sucked all the filtered features from each software and consider the most important attributes all in together. The modules provided in our proposed software are described in detail as follow:

- **Controlling operating system:**
In this module we use the human voice for controlling the operations of operating system. For this facility we use human voice as an input and the output of this facility is the operations performed as per human voice. Suppose the user give the command of 'START' through his voice that time system show the menus present in 'START' button. For performing this operation we use the facility provided by speech engine. By considering feature of speech engine we develop the GUI for controlling this facility which will easily managed by targeted people.
- **Talking editor:**
In this module we provide facility for blind person to listen the words or document present in front of them. Blind people are unable to operate computer only due to their physical disorder but mentally they are able to give the same efforts and output as normal person. This tool is useful in such a situation. In this module the input is user selected document and the output is the voice of computer system.
- **People choice:**
By using this facility any person is able to write the data on editor by using their voice. This tool is useful when someone want to write large amount of data .This module reduce the efforts of normal person but basically it design concentrating on the persons who are physically challenged by their hand.
- **Mail reader:**
In this tool we can hear our mail in mailbox automatically without reading them personally. This tool is specially designed for people who are in software profession. The facility is useful for the professional who are busy almost all the time, they get there mail updates after some time of interval atomatically.They can read their mailbox any time they want.

As the concept relates to voice, so security is the very important issue.

➤ Security module:

So to implement the security we provide the option for encryption and Decryption of data. It will perform using AES, RSA algorithm. In this module we provide the advance security option for strong security. Because password or shoulder sniffing are not comparatively strong. Cryptography is the latest technique and easy to implement. In encryption we convert the normal text to cipher text and key is provided to receiver. At receiver end decryption is done in which cipher text is converted to normal text.

2.1 Application

The concept of voice computing is useful in many areas like:

- for blind people
- for officers
- for students
- for housewives etc....

2.2 Speech technology

Three primary speech technologies are used in talking WORD editor processing applications: stored speech, text-to speech and speech recognition. Stored speech involves the production of computer speech from an actual human talking word that is stored in a computer's memory and used in any of several ways. Speech can also be synthesized from plain text in a process known as text-to speech which also enables talking word processing applications to read from textual database. The first step in voice recognition is for an individual to produce an actual voice sample. Voice production is a fact of life in which we take for granted every day, and the actual process is complicated. The production of sound originates at the vocal cords. In between the vocal cords is a gap. When we attempt to communicate, the muscles which control the vocal cords contract. As a result, the gap narrows, and as we exhale, this breathe passes through the gap, which creates sound. The unique pattern of an individual's voice is then produced by the vocal tract. The vocal tract consists of the laryngeal pharynx, oral pharynx, oral cavity, nasal pharynx, and

the nasal cavity. It is these unique patterns created by the vocal tract which is used by voice recognition systems. Even though people may sound alike to the human ear, everybody, to some degree, has a different or unique enunciation in their speech.

The current applications for voice recognition systems are for physical access entry and where "remote identity verification" is required. Examples of this include call center automation, and transaction processing applications via the telephone or computer. Popular applications in this area are financial transactions (account access; funds transfer; bill payment; trading of financial instruments) and credit card processing (address changes; balance transfers; loss prevention). Voice recognition has also made an impact in the penal system. This technology has been used for inmates on parole, juvenile inmates, and those under house arrest. However, voice recognition technology has not been as widely adopted and utilized as the other biometric technologies examined in previous articles (iris recognition, fingerprint recognition, hand geometry recognition, and facial recognition).

Some future applications for voice recognition systems include Customer Relationship Management (CRM) applications, wireless products, and Voice over IP (VOIP).

What is Voice Computing?

Voice recognition is an alternative to typing on a keyboard. Put simply, you talk to the computer and your words appear on the screen. The software has been developed to provide a fast method of writing on a computer and can help people with a variety of disabilities. It is useful for people with physical disabilities who often find typing difficult, painful or impossible. Voice-recognition software can also help those with spelling difficulties, including users with dyslexia, because recognized words are almost always correctly spelled.

Voice-Computing Software

Voice-recognition software programs work by analyzing sounds and converting them to text. They also use knowledge of how English is usually spoken to decide what the speaker most probably said. Once correctly set up, the systems should recognize around 95% of what is said if you speak clearly. Several programs are available that provide voice recognition. These systems have mostly been designed for Windows operating systems, however programs are also available for Mac OS X. In addition to third-party software, there are also voice-recognition programs built in to the operating systems of

Windows Vista and Windows 7. Most specialist voice applications include the software, a microphone headset, a manual and a quick reference card. You connect the microphone to the computer, either into the soundcard (sockets on the back of a computer) or via a USB or similar connection.

3. Future scope

The system can be extended to continuous word recognition with large vocabulary based on a phone acoustic model. This work can be taken into more detail and more work can be done on the project in order to bring modifications and additional features. The current software doesn't support a large vocabulary, the work will be done in order to accumulate more number of samples and increase the efficiency of the software. The current version of the software supports only few areas of the notepad but more areas can be covered and effort will be made in this regard. The scope of project can be extended to implement the system on small microchip so as to make it more popular, cost effective and user friendly.

4. Conclusion

Through this paper, we present a scheme to convert speech to text as well as text to speech. The key factor in designing such system is the target audience. For example, physically handicapped people should be able to wear a headset and have their hands and eyes free in order to operate the system. A word based acoustic model is used. This model can be used only for limited vocabulary. As the size of the vocabulary increases performance of the system decreases. The system cannot properly distinguish between similar words. Like to and two because they have similar sound phonemes. At last we conclude that this project can be used at very large scale with very little modifications. During the experiment work medium size vocabulary system was implemented. The system can be extended to continuous word recognition with large vocabulary based on a phone acoustic model, using the HMM Technique or using other growing techniques like Artificial Neural Network.

5. References

- [1]. Dat Tat Tran, Fuzzy Approaches to Speech and Speaker Recognition, A thesis submitted for the degree of Doctor of Philosophy of the university of Canberra.
- [2]S. Young, G. Evermann, T. Hain, D. Kershaw,

G. Moore J. Odell, D. Ollason, D. Povey,
V. Valtchev, and P. Woodland, *The HTK Book
for HTK V3.2*, Cambridge University Press, Cambridge,
UK, 2004.

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

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

[5] “Speaker Recognition”, Joseph P. Campbell, Jr.
Article is from the book: “Biometrics: Personal
Identification in Networked Society”, By Anil Jain,
Ruud Bolle, and Sharath Pankati.

[6] S. Young. Large vocabulary continuous speech
recognition: A review IEEE Signal Processing
Magazine, 13(5):45–57, 1996.

[7] S. Young. Statistical modelling in continuous
speech recognition. Proc. Int. Conference on
Uncertainty in Artificial Intelligence, Seattle, WA,
August 2001.

IJERT