Speech Recognition Technology: Applications & Future

Pankaj Pathak

Asst. Professor

Dept. of Computer Science

Shri Vaishnav Institute of Computer science

Indore, India

pankajpathak101@gmail.com

Abstract: Voice or speech recognition is "the technology by which sounds, words or phrases spoken by humans are converted into electrical signals, and these signals are transformed into coding patterns to which meaning has been assigned", .It is the technology needs a combination of improved artificial intelligence technology and a more sophisticated speech-recognition engine . Initially a primitive device is developed which could recognize speech, by AT & T Bell Laboratories in the 1940s. According to figures provided by industry satisfying the needs of consumers and businesses by simplifying customer interaction, increasing efficiency, and reducing operating costs, speech recognition is used in a wide range of applications. The paper focuses on understanding what speech recognition technology is, how it is performed, its application areas, and the future perspective of speech recognition technology.

Keywords: Speech Recognition, Technology, voice, Dragon system, Disable

I.   INTRODUCTION

Speech recognition technology provides ways in which computer technology can be utilized to tasks which uses natural (human) languages or speech. Now we have entered in the era of computer science concerned with the interactions between computers and human (natural) languages. Voice recognition, or speech recognition, technology utilizes audio input for entering data rather than a keyboard. Speaking into a microphone and it produces the same result as typing words manually with a keyboard. Simply called, voice recognition software is designed with an internal database of recognizable words or phrases. The program matches the audio frequencies of speech with corresponding entries in the database

The study of speech recognition and transcription began in the 1936 with AT&T's Bell Labs. Formerly, most research was funded and performed by Universities and the U.S. Government (primarily by the Military and DARPA - Defense Advanced Research Project Agency). In the early 1980's the Speech Recognition technology reached at the commercial market.

There was several competing research "camps", organized each working independently to develop speech recognition. The first company which launched a commercial product was Covox in 1982. Covox brought digital sound (via The Voice Master, Sound Master and The Speech Thing) to the Commodore 64, Atari 400/800, and finally to the IBM PC in the mid '80s.With this interaction of sound with computers Speech Recognition came into existence. Another company that was founded in 1982 and whose product has become the famous leader in the speech recognition market was **Dragon Systems**. Scan soft, Inc. now owns and manufactures this product, Dragon Naturally Speaking.

A.   *How is Speech Recognition Performed?*

How does a computer convert spoken words or speech into the data which it can understand and then executes accordingly? While using speech recognition technology a person has to speak in a microphone. The electrical signal from the microphone is digitized by an analog-to-digital (A/D) converter, and is stored in memory. To determine the meaning of this voice input, the computer attempts to match the input with a digitized voice sample, or template that has a known meaning to the computer. During the speech recognition a stream of audio samples are entered as input to the program. Before that, the speaker database must be ready and comparable feature must be computed from the input audio signal. The feature vectors are then matched to the contents of the database. This technique is similar to the traditional command inputs from a keyboard. The program contains the input templates, and attempts to match these templates with the actual input provided to the computer. In fact, in the early 1990s, the best speech recognizers were giving a 15% error rate on a relatively small 20,000 word dictation task. Now though, that error percentage has dropped to as low as 1-2%, although this can vary greatly between speaker to speaker.
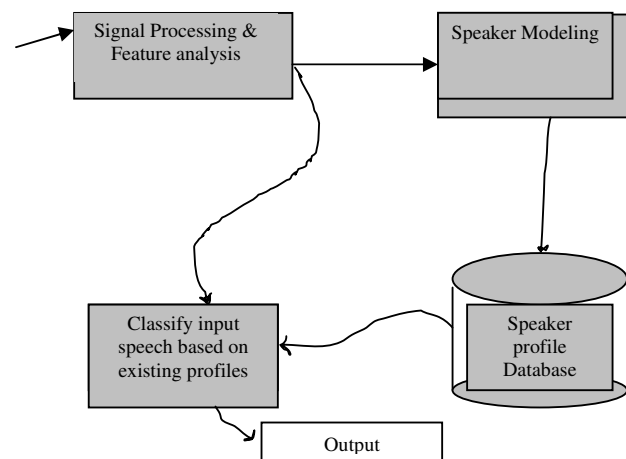

Figure 1.   Working Model of Speech Recognition Software

B. Applications

Speech recognition applications are different from any other kind of computer application. It opens up a world of possibilities for the developers. Speech applications can also present users with more options at any given time, as they are not limited by the number of keys on a phone keypad nor do users have to remember complex numerical figures. Users can simply say what they want and get through their interactions must faster. There are some areas where speech recognition can works efficiently are as follows-

C. People with Disabilities

The most appreciating areas for the application of speech recognition is in helping handicapped and disable people. Within the last decade, many improvements have been done in the performance of speech recognizers and current technology in relation to the needs of the disabled population. Blind people were among the first people with disabilities to benefit significantly from speech recognition computer technology. Using simple machines that could read the words on early text-based computer screens and convert them to synthesized speech, as well as accepts commands spoken by the blind person; make them able to operate computers efficiently. It is also useful for people who have difficulty with or unable to use their hands, causes repetitive stress injuries to involved disabilities that require alternative input for support with accessing the computer system.

D. Health care

Speech recognition is a noble technology for modern health care. The advancement of speech recognition from creating text to capturing and reading information is very helpful for doctors. Speech recognition facilitates information capturing in electronic health record systems as it allows physicians to interact with the system by merely speaking the words. Dictation for a transcriptionist would be also very easy at this point.

Speech recognition is still under development and its use in most likely in certain segments of health care, such as radiology, pathology and emergency medicine. An example of a speech recognition system available to the area of health care is Dragon Naturally Speaking Medical Solutions. Future speech recognition applications will incorporate content solutions, such as medical path technology and coding databases.

E. Helicopters

Efficient test and evaluation programs have been carried out later in speech recognition systems applications in helicopters, notably by the U.S. Army Avionics Research and Development Activity (AVRADA) and by the Royal Aerospace Establishment (RAE) in the UK. Work in France has included speech recognition in the Puma helicopter. There has also been much useful work in Canada. Results have been amazing and encouraged, voice recognition applications to include in control of communication radios; setting of navigation systems; and control of an automated target handover system.

By now there are very interesting solutions using voice commands. When Speech Recognition is used with Microsoft Flight Simulator it gives the feeling that one is in the cockpit of an aircraft, giving instructions to a co- pilot.

The new voice-driven system will be used to train air traffic control personnel, including those in the U.S. military. Since November 2001

The voice recognition software enables students to verbally guide aircrafts into and out of assigned radar sectors.The multi-modal solution with never known accuracy to recognize and instantly respond via visual changes in the radar images to the standard voice command and control phrases used by air traffic control operators, such as "Mr. X, you are number three to land; follow three eight seven on downwind, over."

F. Telephony

Recent advances in wireless communication and speech recognition have made it possible to access the web from any place, at any time, by using only a phone. Some possible applications are browsing the web, getting share prices, verifying flight schedules, getting maps and directions, checking email, etc.

Phones can also offer voice access to the web. Companies like IBM, Motorola, Lucent and AT&T founded in 1999 the VoiceXML Forum to work with speech recognition technologies that make Internet accessible. The VoiceXML Forum aims to drive the market for voice-enabled Internet services through the creation of a common specification based on existing Internet standards. VoiceXML, a markup language for voice applications based on extensible Markup Language (XML), is expected to revolutionize the Internet industry by providing voice access to Web content and services.

Advantages of phone (voice) access to the web:

a. Easy to use: Unlike a computer interface, a voice interface needs no keyboard, no mouse, no screen, freeing users from these barriers to access and action. It requires no training. It is accessible to anyone with a telephone.

b. Access from anywhere: Voice is mobile—information can be sent and retrieved from anywhere. Since customers can have access at anytime from anywhere, voice makes it possible to use time more effectively. Fast and efficient, voice frees users from not only the desktop, but even the laptop.

G. Windows Speech Recognition

Speech recognition software enables the operating system to convert spoken words to written text. Windows Speech Recognition in Windows Vista facilitates users to interact with their computers by voice. User can dictate documents and emails in mainstream applications use voice commands to start and switch between applications, control the operating system, and even fill out forms on the Web. An internal driver, called a speech recognition engine, recognizes words and converts them to text. In order to be successful, the targeted system would need to have the speech recognition feature previously activated and configured. Additionally the system would need to have speakers and a microphone installed and turned on. The speech recognition feature picked up commands through the microphone such as "copy", "delete", "shutdown", etc. and acting on them. These commands would be coming from an audio file that is being played through the speakers. This would be heard and the actions taken would be visible to the user if they were in front of the PC during the execution of attempted command. There are also additional

barriers that, microphone feedback, and the clarity of the dictation by which results can be affected.

As previous versions of the operating system do not appear affected in speech recognition technology. Windows Vista's sophisticated speech recognition allows for easier operation and extended support for commands. This has been largely used to facilitate computing use especially for users that are affected by serious difficulties or impairments.

Different voice recognition software available and which are popular in the market are: Dragon Systems' NaturallySpeaking 3.0 Preferred Edition, IBM ViaVoice 98 Executive, L&H Voice Xpress Plus, and Philips FreeSpeech98. All four programs run on Pentium-powered PC's utilizing Windows 95, 98 or NT 4.0 and require 16-bit SoundBlaster-compatible sound cards. Random access memory (RAM) requirements for software run under Windows NT are higher for all of these programs [5]

Each of the four software uses "wizards" to install and configure the hardware, and all software programs support macros for frequently used phrases.

a. Dragon Systems' NaturallySpeaking uses wizard to train the system to recognize the particular user's voice within 4 minutes. Text Material is provided to the person so that about 30 minutes of reading will improve the accuracy of the outcome. Electronic medical documents are analyzed automatically to "learn" new specialized terms and proper names. Its Command Wizard feature facilitates any user to create medical-specialty macros. Frequently used and required medical forms, electronically stored, can be easily fetched up and the user is prompted to fill out each section of the form.

b. IBM's ViaVoice software also trains the system by reading from selected texts for about 30 minutes, and its wizard makes adjustable the microphone and speaker volume levels.

c. L&H Voice Xpress Plus tells the user to read chapters of a book, and it tests, about 75 minutes for the process.

d. Philips FreeSpeech98 tells the user to read selected text for about 15 minutes; ten training topics are available for the user's review.

Installations of all of the program software are easy and simple, and the initial basic "training" is not very much time-consuming for any of the Voice recognition products. All software products provide macros and the medical customization features of Dragon Systems' product are very good. Though this software initially requires more time and document input, accuracy is increased.

H. Future perspective of VRS

Speech recognition technology has making great progress day by day. Every year, it's 10 percent to 15 percent better than the year before. So, over the next five years, it will reach or exceed human-level performance. Speech Recognition Technology will be providing the most effective interface with computers for everyone. Computers have gradually moved towards human ways of communicating. So what are the future applications that we will soon be seeing in the market place? There are several fields where the speech recognition technology would impacts like: interactive online books, robotic control, interactive internet searching, self driving car interface and many more. Universal translator is yet to achieve, however -- it's very difficult to build a system that combines automatic translation with voice activation technology. One problem is making a system that can flawlessly handle roadblocks like slang, dialects, accents and background noise. The different grammatical structures used by languages can also pose a problem. For example, Arabic sometimes uses single words to convey ideas that are entire sentences in English.

II. CONCLUSION

Speech recognition technology has exploited throughout many industries. Some companies have implemented a robust system that performs as expected and sends the call to their intended destination. This technology would have to be compatible with all software and hardware. This technology would require the CPU to simultaneously process voice input and data access. People with disabilities are great users of this technology, and they tell what is wrong with the software. The people who use speech recognition because of their disabilities are motivated to help in improving the technology.

The speech recognition capabilities in Windows XP, especially when combined with Office XP productivity software, can enhance computing in such areas as gaming, data entry, or editing. Further, third-party software vendors are taking the speech recognition capabilities of your PC into new areas, including things like home automation and telephony.

Voice recognition technology is still in its infancy, and has provided with some laughing and frustrating moments. Basic voice recognition systems exist today. This technology is very new and has limited capability. Some challenges are also to this technology. Possible limitation might be cost. This technology would require absolute refinement. To make this technology accessible for the general public, it would have to be affordable to the average user. With the advent of intelligent, filtering microphones and near-perfect speech-recognition, we will hopefully see a new era of human-computer interaction evolve.

III. REFERENCES

[1] A. Wager, Frances Wickham Lee and John P. Glaser. Managing health care information systems: A practical approach for health care executives (I), Jossey-Bass : A Wiley imprint San Francisco, 2005.

[2] Rajeev Agarwal - Voice Browsing the Web for Information access: http://www.w3.org/Voice/1998/Workshop/RajeevAgarwal.html

[3] http://www.businessweek.com/1998/08/b3566022.htm

[4] http://www.microsoft.com/enable/demos/windowsvista/speech.aspx

[5] http://www.hitl.washington.edu