Artificial Intelligence for Speech Recognition

Kapil Kumar[1], Neeraj Dhoundiyal[2] and Ashish Kasyap[3]

[1,2,3] *Department of Computer Science Engineering, IIMT College of Engineering, GreaterNoida, UttarPradesh, (India)*

Abstract- It is the branch of science which is used develop an ability to develop a machine or program to identify words and phrases in spoken language and convert them into a format which is understandable by machine. Artificial intelligence is based on two basic ideas. Firstly, it involves analyzing and studying the overall activity of human beings. Second one is to represent those processes via machines (like computers, robots, etc).

I. INTRODUCTION

Artificial intelligence is based on two basic ideas. Firstly, it involves analyzing and studying the overall activity of human beings. Second one is to represent those processes via machines (like computers, robots, etc.).

One of the main benefits of speech recognition system is that it provide user do other works simultaneously with an ease. The user can concentrate on observation and manual operations, and still control the machinery by voice input commands.

A speech recognition system can be used in many different modes (speaker-dependent or independent, continuous speech , for small medium or large vocabulary).

Natural language processing (NLP) refers to artificial intelligence methods of communicating with a computer in a natural language like English whose main objective isto understand an input and perform action accordingly.

1.1 PURPOSE

Speech is a very natural way to interact, and it is not necessary to sit at a keyboard or work with a remote control as well as there is no need of training required for users.By the mean of speech recognition technology we can achieve lots of benefits in various fields. It will helps physically challenged skilled persons to do their work with an ease. They can do their works with an ease only just by giving command through voice and rest of thing is done by machines. Similarly ASR technology is also used in military weapons and in Research centers. Now a days this technology was also used by CID officers who used to track the any unusual or criminal activities.

1.2 DEFINITION

Speech Recognition is the branch of science which is used develop an ability to develop a machine or program to identify words and phrases in spoken language and convert them into a format which is understandable by machine.

II. HISTORY

- Work started soon after World-War II.
- Name is coined in 1957.
- Several names that are proposed are…
- Heuristic programming
- Machine Intelligence
- Computational Rationally

III. FOUNDATION

- **1950s and 1960s: Baby Talk**
  The first speech recognition systems could understand only digits. Bell Laboratories designed in 1952 the "Audrey" system, which recognized digits spoken by a single voice. Ten years later, IBM demonstrated at the 1962 World's Fair its "Shoebox" machine, which could understand 16 words spoken in English.

- **1970s: Speech Recognition Takes Off**
  Speech recognition technology made major strides in the 1970s, DoD's DARPA Speech Understanding Research (SUR) program, from 1971 to 1976, was one of the largest of its kind in the history of speech recognition, and among other things it was responsible for Carnegie Mellon's "Harpy" speech-understanding system. Harpy could understand 1011 words.

- **1980s: Speech Recognition Turns Toward Prediction**
  Over the next decade, speech recognition vocabulary jumped from about a few hundred words to several thousand words, and had the potential to recognize an unlimited number of words. One major reason was a new statistical method known as the *hidden Markov model*.

Philosophy	(428 B.C.-present)
Mathematics	(c.800-present)
Economics	(1776-present)
Neuroscience	(1861-present)
Psychology	(1879-present)
Computer Engineering	(1940-present)
Control theory and cybernetics	(1948-present)
Linguistics	(1957-present)

IV. SPEAKER INDEPENDENCY

The speech quality varies from person to person because everyone has its own pitch and voice variation sotherefore it is difficult to build an electronic system that recognizes everyone's voice. By limiting the system to the voice of a single person, the system becomes not only simpler but also more reliable. The computer that trained to the voice of a particular individual known as speaker-dependent system.

Speaker independent systems can be used by anybody who can recognize any voice, even though the characteristics vary widely from one speaker to another. Most of these systems arevery costly, complex and they have very limited vocabularies.

It is important to consider the environment in which the speech recognition system has to work. The language (grammar, phrases ,words etc.) used by the speaker should be recognizable and accepted by the AI system. Noise level, Noise type , position of the microphone, and speed and manner of the user's speech are some factors that may affect the quality of speech recognition.

V. ENVIRONMENTAL INFLUENCE

Real time applications demand that the performance of the recognition systems should be unaffected even after there any change in environment encountered. However, it is a fact that when a system is trained and tested under different conditions, the recognition rate drops unacceptably. We need to be concerned about the variability present when different microphones are used in training and testing, and specifically during development of procedures. Such care can significantly improve the accuracy of recognition systems that use desktop microphones.

Acoustical distortions can degrade the accuracy of recognition systems. Obstacles to robustness include additive noise from machinery, competing talkers, reverberation from surface reflections in a room, and spectral shaping by microphones and the vocal tracts of individual speakers.

A number of algorithms for speech enhancement have been proposed. These include the following:

1.   Spectral subtraction of DFT coefficients
2.   MMSE techniques to estimate the DFT coefficients of corrupted speech
3.   Spectral equalization to compensate for convoluted distortions
4.   Spectral subtraction and spectral equalization.

Although relatively successful, all these methods depend on the assumption of independence of the spectral estimates across frequencies. Improved performance can be got with an MMSE estimator in which correlation among frequencies is modeled explicitly.

VI. SPEAKER-SPECIFIC FEATURES

Speaker identity correlates with the physiological and behavioral characteristics of the speaker. These characteristics exist both in the vocal tract characteristics and in the voice source characteristics, as also in the dynamic features spanning several segments.

The most common short-term spectral measurements currently used are the spectral coefficients derived from the Linear Predictive Coding (LPC) and their regression coefficients. A spectral envelope reconstructed from a truncated set of spectral coefficients is much smoother than one reconstructed from LPC coefficients.

Therefore, it provides a more stable representation from one repetition to another of a particular speaker's utterances.

As for the regression coefficients, typically the first and second order coefficients are extracted at every frame period to represent the spectral dynamics.

These coefficients are derivatives of the time function of the spectral coefficients and are called the delta and delta-delta-spectral coefficients respectively.
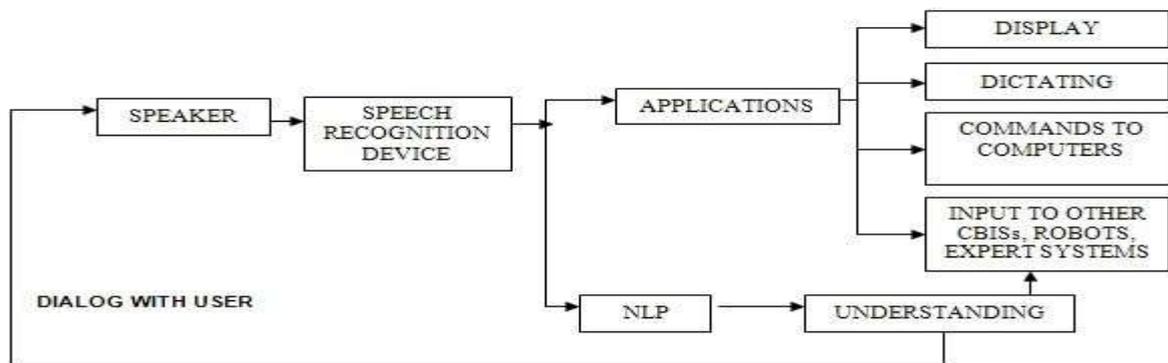
VII. SPEECH RECOGNITION

The user communicates with the application through the appropriate input device i.e. a microphone. The Recognizer converts the analog signal into digital signal for the speech processing. A stream of text is generated after the processing. This source-language text becomes input to the Translation Engine, which converts it to the target language text.
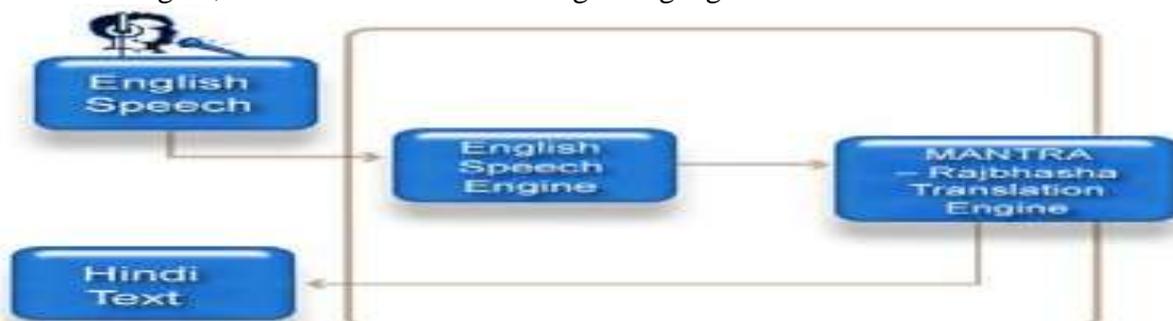
Salient Features
- Input Modes
  - ✓ Through Speech Engine
  - ✓ Through soft copy
- Interactive Graphical User Interface
- Format Retention
- Fast and standard translation

Interactive Preprocessing tool
  - ✓ Spell checker.
  - ✓ Phrase marker
  - ✓ Proper noun, date and other package specific identifier

User friendly selection of multiple output.
Online thesaurus for selection of contextually appropriate synonym.
Online word addition, grammar creation and updating facility.
Personal account creation and inbox management.

VIII. APPLICATIONS

One of the main benefits of speech recognition system is that it lets user do other works simultaneously. The user can concentrate on observation and manual operations, and still control the machinery by voice input commands.

Another major application of speech processing is in military operations. Voice control of weapons is an example. With reliable speech recognition equipment, pilots can give commands and information to the computers by simply speaking into their microphones - they don't have to use their hands for this purpose.

Another good example is a radiologist scanning hundreds of X-rays, ultra sonograms, CT scans and simultaneously dictating conclusions to a speech recognition system connected to word processors. The radiologist can focus his attention on the images rather than writing the text.

Voice recognition could also be used on computers for making airline and hotel reservations. A user requires simply to state his needs, to make reservation, cancel a reservation, or make enquiries about schedule.

IX. ULTIMATE GOAL

The ultimate goal of the Artificial Intelligence is to build a system who can act as intelligently as human being are.
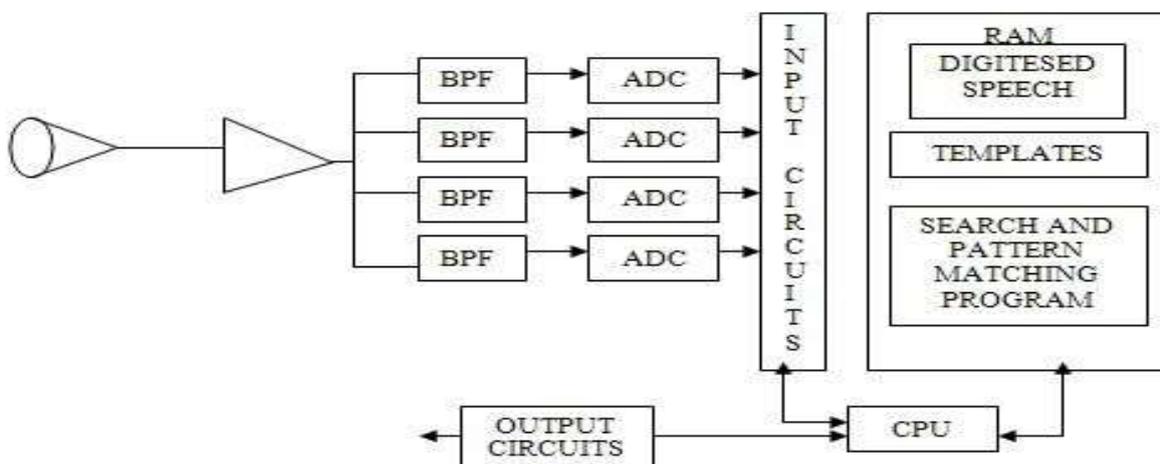
X. CONCLUSION

Speech Recognition will revolutionize the way people conduct business over the web and will,ultimately,differentiate world-class e-businesses. VoiceXML ties speech recognition and telephony together and provides the technology with which businesses can develop and deploy voice-enabled Web solutions TODAY! These solutions can greatly expand the accessibility of web-based services transactions to customers who would otherwise not have access.In near future people will be using their home and business computers by speech not by Keyboard or mouse. Home automation will be completely based on speech recognition.

REFERENCES

[1] www.google.co.in/Artificial intelligence for speech recognition
[2] www.google.com
[3] www.howstuffworks.com
[4] www.ieeexplore.ieee.org
[5] www.slideshare.net
[6] http://aitopics.org
[7] http://www.pcworld.com