

Artificial Intelligence Applications for Speech Recognition

Raghendra Priyam

*Team Leader (South 1)
TATA Motors CRM DMS Project
Bangalore
Email: priyambabu@gmail.com*

Rashmi Kumari

*TATA Motors CRM DMS Team
Bangalore
Email: rashmipriyamraj@gmail.com*

Dr. Prof Videh Kishori Thakur

*HOD, Physics
BRA Bihar University, Muzaffarpur
Email: dr.videh@gmail.com*

Abstract

Artificial intelligence involves two basic ideas. First, it involves studying the thought processes of human beings. Second, it deals with representing those processes via machines (like computers, robots, etc.). 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.

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.

By using this speaker recognition technology we can achieve many uses. This technology helps physically challenged skilled persons. These people can do their works by using this technology with out pushing any buttons. This ASR technology is also used in military weapons and in Research centers. Now a day this technology was also used by CID officers. They used this to trap the criminal activities.

Introduction:

Artificial intelligence involves two basic ideas. First, it involves studying the thought processes of human beings. Second, it deals with representing those processes via machines (like computers, robots, etc.).

AI is behavior of a machine, which, if performed by a human being, would be called intelligent. It makes machines smarter and more useful, and is less expensive than natural intelligence.

Natural language processing (NLP) refers to artificial intelligence methods of communicating with a

computer in a natural language like English. The main objective of a NLP program is to understand input and initiate action.

Definition:

It is the science and engineering of making intelligent machines, especially intelligent computer programs.

AI means Artificial Intelligence. Intelligence” however cannot be defined but AI can be described as branch of computer science dealing with the simulation of machine exhibiting intelligent behavior.

Speaker independency:

The speech quality varies from person to person. It is therefore 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 must be trained to the voice of that particular individual. Such a system is called speaker-dependent system.

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

It is important to consider the environment in which the speech recognition system has to work. The grammar used by the speaker and accepted by the 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.

Environmental influence:

Real applications demand that the performance of the recognition system be unaffected by changes in the environment. 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. These sources of distortions fall into two complementary classes; additive noise and distortions resulting from the convolution of the speech signal with an unknown linear system.

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.

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.

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:

- (1) Input Modes
 - Through Speech Engine
 - Through soft copy
- (2) Interactive Graphical User Interface
- (3) Format Retention
- (4) Fast and standard translation

(5) Interactive Pre-processing tool

- Spell checker.
- Phrase marker
- Proper noun, date and other package specific identifier Input Format
- Input Format : txt, .doc .rtf
- 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

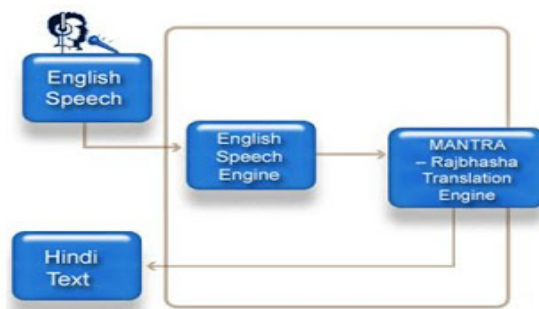


Figure 1 Method for Speech Recognition

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 stating his needs, to make reservation, cancel a reservation, or making enquiries about schedule.

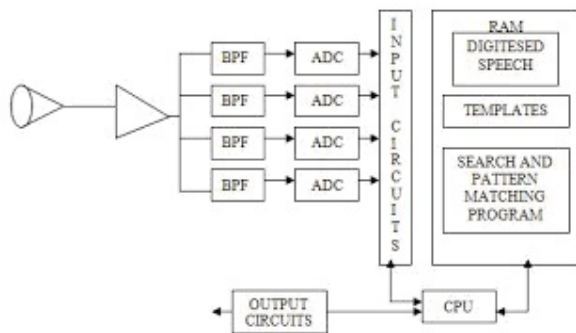


Figure 2 Voice Recognition

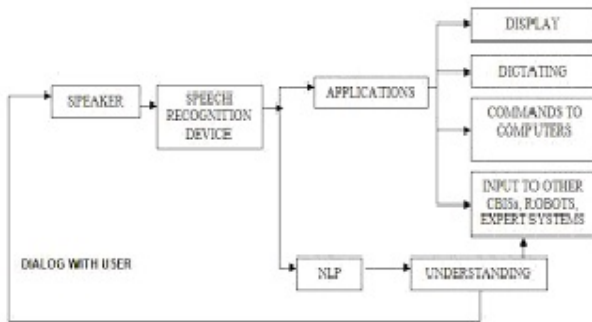


Figure 3 Voice Processing

Ultimate Goal:

The ultimate goal of the Artificial Intelligence is to build a person, or, more humbly, an animal.

Conclusion:

By using this speaker recognition technology we can achieve many uses. This technology helps physically challenged skilled persons. These people can do their works by using this technology without pushing any buttons. This ASR technology is also used in military weapons and in Research centers. Now a day this technology was also used by CID officers. They used this to trap the criminal activities.

References

1. [www.google.co.in/Artificial intelligence for speech recognition](http://www.google.co.in/Artificial_intelligence_for_speech_recognition)
2. www.google.com
3. www.howstuffworks.com
4. www.ieeexplore.ieee.org
5. www.wikipedia.org
6. Developing an Artificial Intelligence Engine (Michael van Lent and John Laird)