Artificial Intelligence and its Application in Speech Recognition

S.V. Viraktamath, Chaitra P. Shet and Pooja R. Nayak

Abstract--- Artificial intelligence involves studying the thought processes of human beings and also representing those processes via computers and robots like man made machines. Speech recognition system lets user do other works simultaneously so that the user can concentrate on observation and manual operations. With the help of speaker recognition technology we can help physically challenged skilled persons. So that they can do their works without the help of others. This Artificial Speech Recognition technology is also used in various applications. Now days this technology is also used by CID officers in order to trap the criminal activities. This technology is also used in military applications.

Keywords--- NLP, LPC, DFT, MMSE Technique.

I. INTRODUCTION

ARTIFICIAL intelligence involves two basically deals with studying the thought processes of human beings and then representing those processes via computer[1]. There is one artificial intelligence method through which we can communicate with a computer in a natural language like English which is termed as Natural Language Processing(NLP). The main objective of a NLP program is to understand the applied input and then initiate related action.

II. DEFINITION

Artificial Intelligence is the science and engineering of making intelligent machines, especially intelligent computer programs so that it can serve most of user needs[2]. AI implies Artificial Intelligence. Intelligence is the factor which cannot be defined but whereas AI can be defined as branch of computer science which deals with the simulation of machine that exhibits intelligent behavior as the human being. Speech recognition is considered to be one of the applications of the artificial intelligence which mainly deals with the translation of user spoken words into the corresponding text.

III. OBJECTIVES

- 1) Deals with understanding human thinking capabilities: This means to obtain deep knowledge of human memory, Problem solving ability, learning and decision making etc.

- 2) Replaces human in intelligent tasks: Results in building of system in order to help humans think better, faster and deeper.
- 3) Enhance human intelligence: It implies building program to exceed human intelligence.
- 4) Deals with Coherent discourse: Communication with people using natural language involves intelligent dialogue.

IV. SPEAKER INDEPENDENCY

Usually the speech quality varies from one person to another person. So it becomes difficult to design an electronic machine that recognizes everyone's voice. The system is made simpler and also more reliable by designing it in order to recognize single person's voice. The computer is trained to the voice of the particular individual. Such developed system is called speaker-dependent system. Speaker independent systems can be used by anybody, as it recognizes any voice, although the characteristics vary widely from one speaker to another. These speaker independent systems are costly and complex to construct. These systems have got limited vocabularies. There are some factors that may affect the quality of speech recognition. That includes 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 and so on.

V. ENVIRONMENTAL IMPACT

The recognition rate usually drops widely when a system is trained and tested under different conditions. It is necessary that we should be aware of the variations present when different microphones are used in training, testing, and also during development of required procedures. Such that with this the accuracy of recognition systems can be improved to greater extent.

Accuracy of recognition systems are going to be degraded mainly because of Acoustical distortions. Obstacles for robustness include additive noise from machinery, competing talkers, reverberation from surface reflections in a room, and spectral shaping by microphones and also the vocal tracts of individual speakers and so on. The sources of distortions are categorized as additive noise and distortions resulting from the convolution of the speech signal with an unknown linear system. There are various algorithms proposed for speech enhancement. They are given as follows:

1. Spectral subtraction of the obtained DFT coefficients.
2. The DFT coefficients of corrupted speech are estimated by applying MMSE technique Convolved

S.V. Viraktamath.

Chaitra Prakash Shet, Student, Department of Electronics & Communication, Shri Dharmasthala Manjunatheshwara College of Engineering and Technology, India. E-mail:chaitraprakashshet@gmail.com

Pooja Ravindra Nayak, Student, Department of Electronics & Communication, Shri Dharmasthala Manjunatheshwara College of Engineering and Technology, India. E-mail:poojaravindranayak@gmail.com

DOI:10.9756/BIJRCE.8199

Distortions are compensated by spectral equalization method.

3. Spectral subtraction and spectral equalization algorithm.
4. These methods are relatively successful, and all these methods depend on the assumption of independence of the spectral estimates across frequencies. MMSE estimator is used in order to get better performance in which correlation among frequencies is modeled explicitly.

VI. SPEAKER RELATED FEATURES

Physiological and behavioral characteristics of the speaker are mainly considered as speaker identity. These features are found both in the vocal tract characteristics and in the voice source characteristics, as also in the dynamic features spanning the several segments. The most common short-term spectral measurements currently used are the spectral coefficients derived from the Linear Predictive Coding

(LPC)[3].

The reconstruction of spectral envelop from a truncated set of spectral coefficients is found to be much smoother than one which is obtained from LPC coefficients. Hence, it provides a more stable representation of a particular speaker's utterances. In order to represent the spectral dynamics typically the first and second order coefficients are extracted at every frame period. These coefficients which are derivatives of the time function of the spectral coefficients are called the delta and delta-delta-spectral coefficients respectively.

VII. SPEECH IDENTIFICATION

Microphone is found to be the input device through which the user communicates with the application. For the speech processing the Recognizer converts the analog signal into digital signal. As a result of which stream of text is generated[4]. This source-language text acts as the input to the Translation Engine, which converts it to the target language text [5].

Figure 1: Speech Recognition

Salient Features

- 1) Modes of applying input
 - a) Using Speech Engine
 - b) Using soft copy
- 2) Feature concerned with Interactive Graphical User Interface
- 3) Format Retention
- 4) Standard and quick translation
- 5) Comprising of various Interactive Pre-processing tool
 1. Spell checker.
 2. Phrase marker
 3. Proper noun, date and other package specific identifier Input Format
 4. Input Format : txt, .doc .rtf Selection of multiple output which is user friendly
5. contextual selection of appropriate synonym using Online thesaurus
6. Word addition, grammar creation and updating facility through online
7. Includes Personal account creation and also inbox Management.

VIII. VOICE IDENTIFICATION

The field of computer science is made to design computer systems that can recognize spoken words by the user. Voice recognition implies identifying particular voice. A number of voice recognition systems are available on the market most of which require a training session during which the computer system is trained to identify particular voice and accent. Such systems are known as speaker dependent.

In most of the discrete speech systems it requires that the speaker speak slowly and distinctly and separate each word

with a short pause so that identification becomes easier.

Figure 2: Voice Recognition

IX. VOICE PROCESSING

The handling of voice through computer involves, storing of voice, forwarding it, voice response, voice recognition and

text to speech technologies. After voice is recognized processing of voice is done. In order to facilitate voice processing by the system analog signal is converted to digital signal

Figure 3: Voice Processing

X. APPLICATIONS

Speech recognition has got many applications among which it lets user do other works simultaneously. So that user can concentrate on observation and manual operations, and still control the machinery by voice input commands. In the field of military also speech recognition has got its application [6]. The best example for reliable speech recognition equipment is Voice control of weapons, computers take input as simply spoken words for example the commands given by the pilots through their microphones such that they don't have

to use their hands for this purpose[7].

A radiologist usually scans hundreds of X-rays, ultra sonograms, CT scans and simultaneously dictating conclusions to a speech recognition system which is connected to word processors [8]. With the help of this radiologist can focus his attention on the images rather than writing the text. For making airline and hotel reservations also voice recognition is used. [9]. With the help of this application a user requires simply stating his needs, in order to make reservation, cancel a reservation, or making any enquiries about schedule.

Figure 4: Fields of AI

AI is (also known as intelligent system) mainly based upon mathematics (particularly logic, combinatory, statistics, probability and optimization theory), psychology, linguistics, neuroscience and philosophy [10]. But it is primarily a branch of computer science and it has borrowed a lot of concepts and ideas from the above mentioned fields.

XI. ULTIMATE GOAL

The ultimate goal of the Artificial Intelligence is to help user do his work simultaneously which in turn is going to reduce the time consumption and also reduce human work to a greater extent.

XII. CONCLUSION

This speaker recognition technology helps physically challenged skilled persons. These people can do their works by using this technology by consuming very less time. This ASR technology is used in military weapons and in Research centers. This technology is also used by CID officers in order to trap the criminal activities.

REFERENCES

- [1] J.L. Barrett and F.C. Keil, "Conceptualizing a non-natural space entity", *Anthropomorphism in God concepts Cognitive Psychology*, Vol.31, No.3, Pp.219:247, 1996.
- [2] Himanshu and S. Kaur, "Literature Survey on Automatic Speech Recognition System", Vol.4, No.7, 2014.
- [3] H.H. Ammar, W. Abdelmoez and M.S. Hamdi, "Software engineering using artificial intelligence techniques: Current state and open problems", In *Proceedings of the First Taibah University International Conference on Computing and Information Technology*, Al-Madinah Al-Munawwarah, Saudi Arabia, Pp.52, 2012.
- [4] J.T. Chien and S. Furui, "Predictive hidden Markov model selection for speech recognition", *IEEE Transactions on Speech and Audio Processing*, Vol.13, No.3, Pp.377-387, 2005.
- [5] A.M. Anusuya and K.S. Katti, "Speech Recognition By Machine: A Review", *International Journal of Computer Science and Information Security*, 2009.
- [6] W. Dai and P. Wang, "Application of pattern recognition and artificial neural network to load forecasting in electric power system", In *Third International Conference on Natural Computation*, Vol.1, Pp.381-385, 2007.
- [7] A. Choudhary and R. Kshirsagar, "Process Speech Recognition System using Artificial Intelligence Technique", *International Journal of Soft Computing and Engineering*, 2012.
- [8] S. Rawat, P. Gupta and P. Kumar, "Digital life assistant using automated speech recognition", *2014 Innovative Applications of Computational Intelligence on Power, Energy and Controls with their impact on Humanity (CIPECH)*, Pp. 43-47, 2014.

- [1] P. Saini and P. Kaur, "Automatic speech recognition: A review", *International journal of Engineering Trends & Technology*, Pp.132-136, 2013.
- [2] L. Deng and X. Li, "Machine learning paradigms for speech recognition: An overview", *IEEE Transactions on Audio, Speech, and Language Processing*, Vol.21, No.5, Pp.1060-1089, 2013.

ABOUT THE AUTHOR

Chaitra Prakash Shet (30-08-1995) is an under graduating student pursuing her final year degree in Electronics and Communication department of SDMCE, Dharwad, Karnataka. She is interested in VLSI, Control Systems and HDL. (E-mail: chaitraprakashshet@gmail.com)

Pooja Ravindra Nayak (17-05-1996) is an under graduating student pursuing her final year degree in Electronics and Communication department of SDMCE, Dharwad, Karnataka. She is interested in VLSI, Control Systems and HDL. (E-mail: poojaravindranayak@gmail.com)