# Current Challenges and Application of Speech Recognition Process using Natural Language Processing: A Survey

Neha Chadha
M.Tech Scholar
BCET,Gurdaspur,Punjab, India

R.C. Gangwar, PhD
Associate Professor
BCET,Gurdaspur Punjab, India

Rajeev Bedi
Assistant Professor
BCET,Gurdaspur Punjab, India

## ABSTRACT
Speech recognition is a vast research field for researchers in modern era. Earlier, the human language was processed by the computer system for speech recognition. Thus, the main objective is to develop recognition system which improves human to human communication by enabling human-machine communication by processing of text or speech. Various applications of speech recognition systems are present and these all includes various research challenges. A critical machine learning based review is defined which addresses the various challenging tasks of speech recognition system in NLP. In the existing systems, the recognition rate is very less and the noise ration during the recognition process creates a problem. Thus in this literature review we try to address such kind of challenges and provides a solution to work further in future.

## Keywords
NLP, GUI, MFCC, LFCC, LPC, KLM, LIF, HMM, DTW, SAE.

## 1. INTRODUCTION
The science that is most directly related to processing of human language is natural language processing. The dealing of this science directly to the natural language makes it different from other processing related activity in the field of application: the human language. NLP and Understanding is the state of art that is quite demanding these days. The research in this field has been started 50 years ago, but because of limitations of resources that are required in processing the speech, it was not implemented then in commercial applications. In today world the computer dependency expand the field of speech processing. In various science fi movies like Star Wars, Sneakers, Star Trek, Red Dwarf, UFO, Blade runner and many more the concept of natural language processing has been used.

## 2. TYPES OF SPEECH RECOGNITION
**A. Connected word system**
The combination of two words forms a one single word. The minimal pause is taken between the utterance of two words and isolated word is formed.

**B. Continuous speech recognizer**
It is also known as computer dictation. It is natural speech of a speaker.

**C. Spontaneous speech system**
The natural sounds like “ums” “aah” “hmm” utterance along with speech.

**D. Voice verification/identification**
The identification and verification of specific speaker’s voice by applying various tools and techniques.

## 3. APPLICATIONS OF SPEECH RECOGNITION
***A. Voice biometrics***
This technology compares the previously stored voice print or template with the utterance and produces score. It works on voice interpretation algorithm. Biometric reduces each spoken word into some frequency segments called formants. This technology is used at various agencies like online banking, online security trading, online information services, computer access security and many more.

***B. Siri technology***
It is an application used by the company Apple Macintosh in their iphones. This application captures the voice from the speaker and performs the function narrated by the speaker. For example, you can ask to call a specific person from your contact list, or send him a message and can narrate the message as well.

***C. Games and toys***
Various voice driven toys and games are available in market. The simulation of these types of products is based on the voice command given to them.

***D. Fighter aircrafts***
The fighter jets are controlled with the voice system and are being given commands from the base. The speech recognition systems are helpful for controlling the hands free weapons system.

***E. Home automation***
Voice activated alarm and control signals helps to provide security at homes.

## 4. Tools and techniques
**A. Dragon dictate**
Dragon Dictate is proprietary speech recognition software for Windows, Android and Mac. Dragon is 3 xs faster than typing and its 99% accurate. The translation is an easy process rather than typing. This is the great benefit of dragon dictate.

**B. SRI dynaspeak**
This speech recognition engine with modest hardware requirement can recognize up to 50k vocabulary.

**C. Pocket sphinx**
It is an embedded speech recognition engine. This uses the less accurate GMM models. One of the big advantage of Pocket sphinx is that it can recognize 10k words with error rate of 20%. It has configurable memory which is used to build any model you require.

**D. Voice navigator**
This was the first voice recognition system for graphical and command line interface. This system includes both software

and hardware. The first voice navigator product was with the name voice navigator with fax modem capability. After that many products like voice navigator II, voice navigator SW, voice navigator SDK has been issued. This was originally designed for Apple Macintosh Plus and after succeeding many versions was created for windows based OS. Today's world various android apps have been created for voice navigation.

**E. Transcriber**

This tool is used for segmentation of long duration voice signals, labeling and transcribing speech. The Transcriber Pro is most convenient tool which is used by professionals to convert the speech into text. Transcriber works on various platforms like Windows XP/2k, Mac OS X and Linux.

**F. Praat**

Praat software package is designed for linguists to use in speech of analyzing .It is free with open source code. Praat is compatible with Mac OS X, Windows and Linux. The first version of Praat 1.0 was created on 10th July 2011; the 19th and latest version of Praat 1.6 was released on 5th July 2015.

## 5. LITERATURE SURVEY

Anupam Choudhary et. al. (2012) [1] described the speech recognition process using the approach of AI. The recognition method used is language mode, trigram model and acoustic model. No GUI is used, acoustic model interface with the telephony system to manage spoken dialogues by the speaker.

Alexandre Trilla (2012) [2] worked on the approach of Automatic Speech Recognition using NLP technique. It depicts the production of sound from the text i.e. text to speech synthesis and vice versa i.e. known as automatic speech recognition.

D D Doye et. al. (2015) [3] they worked on the approach of new non linear time alignment model rather than DTW algorithm. They worked for finding suitable time alignment algorithm for the Marathi language. They took 46 monosyllabic confusing alphabets and 46 confusing names for their work. They main feature used in this research were Mel Frequency Cepstral Coefficients (MFCC), Linear Frequency Cepstral Coefficients (LFCC) and Linear Prediction Coefficient (LPC)

Dr. Kavita R. et. al. (2014) [4] They proposed a work on digitizing the audio into samples by using the concept of sampling. The MFCC feature is used for extraction process. These coefficients are used for matching the Tamil database through the DTW approach. This main focus of this paper is security of extracting and matching by using the DTW and mathematical approaches.

Elyes et. al. (2014) [5] The hybrid approach of SVM/HMM is used for Arabic ASR on triphones modeling. They used the Arabic speech recognition system that is based on triphones are 64.68 % with HMMs, 72.39 % with MLP/HMM and 74.01 % for SVM/HMM hybrid model.

Fook C.Y et.al. (2012) [6] The main aim of this research is to compare and summarize the well known speech recognition methods used by various researchers.

Jayashree Padmanabhan et.al. (2015) [7] The automatic speech recognition along with Gaussian mixture model, machine learning and HMM is reviewed. The scanning, preprocessing, extraction and classification of input are done by using the feature of acoustic, bottleneck and MLP.

Kenji Sagae et.al. (2009) [8] researched on the incremental system which complete the speech by using incremental language processing capabilities before the completion of speech utterance.

Mohammad et.al. (2014) [9] They proposed an emotion recognition system based on speech analysis. The same data is repeated by user six times and recorded. The emotion is calculated by evaluating pitch, frequency and intensity like parameters. A freeware software"Praat" is used in the process of emotion recognizing.

Poonam.S.Shetake et. al. (2014) [10] This paper reviewed many techniques which are used for TTS conversion. Various applications are based on TTS. This paper many focused on character recognition and the TTS conversion approaches.

Qirong Mao et.al. (2014) [11] They proposed an SER using Convolution Neural Network. They worked in two steps; firstly they use unlabelled sample to learn local invariant factor (LIF) using a variant of sparse auto-encoder (SAE) with reconstruction penalization. After that, in the second step, they use LIF which act as an input for extracting feature

Siva Prasad Nandyala et. al. (2014) [12] The new approach of hybrid HMM/DTW by using kernel adaptive filters for speech analysis and recognition is used. The noise removal or filtration of conversations like over the telephone is very important in speech recognition. Their approach gave better experimental results as compare to traditional results.

Xiang-Lilan et. al.(2014) [13] In this paper they introduced a new merged-weight dynamic time wrapping algorithm(MWDTW). This method defines a template confidence index for measuring the similarities between training and testing data, by using the DTW approach. By using the merge approach of SD speech recognition datasets, HMM and DTW on merged data sets, resulted six times better than DTW overall.

Zue, V. et. al. (2011) [14] defined an approach for the audio dialogues, text, icons and graphics. The SR and understanding system for urban penetration. The authors produced a language of word-pair which helps in searching and navigation. N word string matching and filtering for components is also implemented.

**Table 1: Summary of Literature Review**

| Name of Author | Recognition Method | Efficiency | Time Taken | Features | Error rate |
|---|---|---|---|---|---|
| Anupam Choudhary | Speech Recognition (AI, Language Model, Trigram) | 94% | 2-3 times the real time | Frequency response resonance & anti resonance | 5-6% |
| Alexandre Trilla | ASR using NLP Technique & Text to Speech | 96% | 6-7 times the real | TTS synthesis | 4-5% |
| Daniel Jurafsky | Speech & language processing | 97-98% | Real time | Morphological & lexical features | 2-3% |
| D D Doye | Speech Recognition System | 89.13% | 1 (relative time) | MFCC, LFCC & LPC | 8-9% |
| Jim Glass | Automatic speech Recognition | 98.00% | Real time | Acoustic Modeling, HMM | 0.30% |
| Kenji Sagae | Partial Speech Recognition ( PSR) | 96-97% | Run/Real time Dialogue | Consecutive words | 0.54% |
| Mohammad Rabiei | Human Emotion Recognition | 78.54% | Real time as computational load is very low | Sound Waves & Speech Energy | 10-11% |
| Jayashree Padmanabhan | ASR(Gaussian mixture models,HMM,Machine learning) | 68% | Scanning, Pre-processing, Extraction & Classification | Accoustic, Bottle neck & MLP Features | 5.80% |
| Siva Prasad Nandyala | Adaptive Filters, Dynamic time warping (DTW) | 86.96 % | Real time | KLMS & DTW | 7-8% |
| Victor Zue | NLP(TINA) | 51.70% | 3 to 5 times the real time | Summit system, TINA system | 48.30% |
| Suma Swamy | Speech Recognition System | 98% | Real time | MFCC & Distance Minimum technique | 2-3% |

## 6. COMPARISON OF EXISTING APPROACHES

**Table 2: Comparison of Existing Approaches**

| OCR | NLP | SR | PR | DTW | HMM |
|---|---|---|---|---|---|
| Optical Character Recognition is conversion software used to convert and extract text from image | Natural language processing (NLP) is a branch of computer science which deals which the human (natural ) language with computer | Speech recognition (SR) is to translate an audio or voice into the text. | Pattern recognition is a branch of pattern recognition for machine learning. | Dynamic time warping (DTW) is an algorithm for used for pattern matching between the two who may vary in speed or time | A hidden Markov model (HMM) is a model which is to use to study the hidden or unobserved states. |
| It is mainly used in the passport documents, invoices, bank statements, computerized receipts and business cards. | It is used widely in the development of Artificial intelligence and is the key feature to solve the central artificial intelligence problem | It is widely used in the car systems, therapeutic use , military programs, education, home automation etc | It is used in license plate recognition, fingerprint analysis and face detection/verification | DTW has been applied to temporal sequences of video, audio, and graphics data. | HMM is widely used in the application areas of speech, handwriting, gesture, POS tagging |
| Its method is used in digitized printed texts so that it can be electronically edited, searched, stored more compactly | Its method is used of converting human language into another, like machine understandable language etc. | Its method is used in the telephony has shown benefits to short-term-memory re-strengthening in brain | Its method is commonly used in face recognition, to check spam or non spam email messages | The most common applications include speaker recognition and online signature recognition. Also it can be used in partial shape matching application | HMM provide a framework for modeling using mathematical computations. |
| OCR is a field of research area of AI and computer vision | It enables computers to derive to analyze all human language | It enables the humans to give verbal command to a machine to perform the particular task instead on controlling it manually. | PR main aim is to give useful and valid result on all possible likely matching inputs | Used in shape analysis and geometric shapes use in computers. | HMM pair is widely used in finding pair wise alignment of protein and DNA |

## 7. CURRENT CHALLENGES IN SPEECH RECOGNITION

The performance of the audio input system degrades due to noise from the outer sources. Accuracy and reliability of the system is affected by the unwanted input and low output result. The fault tolerance capacity lacks in this case. User responsiveness is also one of the challenges, it happens when the resources are not ready and user starts to speak the command and then it leads to problem of synchronizing the data with multiple applications (media, phone, navigation)

## 8. CONCLUSION

A good way and process for the recognition of speech is to find a best way which can minimize the error rate during recognition. This paper defined the various recognition techniques and methods used in the current era with their pros and cons. Thus our literature indicated that efforts can be made to propose a novel approach for the recognition process which will produce better results as compare to the existing methodologies. For this better results, database of the speech signals should be last so that texting can be performed on large database. Furthermore in future research can be made when people interact with complex media indicate that speech and language processing tools and techniques will be critical in development.

## 9. REFERENCES

[1] Anupam Choudhary, Ravi Kshirsagar, 2012 Process Speech Recognition System using Artificial Intelligence Technique In International Journal of Soft Computing and Engineering (IJSCE) ISSN: 2231-2307, Volume-2, Issue-5.

[2] Alexandre Trilla, 2012 Natural Language Processing in Text to Speech synthesis and Automatic Speech Recognition In IEEE, VOL.4

[3] Daniel Jurafsky, James H. Martin, 2000 Speech and Language Processing In Pearson Education, pp-1-975.

[4] D D Doye, T R Sontakke & Smita Nagtode, 2015 The Nonlinear Time Alignment Model for Speech In IETE Journal of Research, Taylor & Francis, pp 1-6.

[5] Dr. Kavitha, Nachammai, Ranjani, Shifali., 2014 Speech Based Voice Recognition System for Natural Language Processing In International Journal of Computer Science and Information Technologies, Vol. 5

[6] Elyes Zarrouk, Yassine Ben Ayed, Faiez Gargouri, 2014 Hybrid continuous speech recognition systems by HMM, MLP and SVM: a comparative study, International Journal of Speech Technology ,Volume 17, Issue 3, pp 223-233.

[7] Fook, C.Y. ; Sch. of Mechatron. Eng., Univ. Malaysia Perlis , Arau, Malaysia ; Hariharan, M. ; Yaacob, S. ; Adom, A., 2012 A review: Malay speech recognition and audio visual speech recognition In Biomedical Engineering (ICoBE), International Conference.

[8] Jayashree Padmanabhan and Melvin Jose Johnson Prem kumar, 2015 Machine Learning in Automatic Speech Recognition: A Survey, IETE Technical Review, Taylor & Francis, pp-1-13.

[9] Kenji Sagae and Gwen Christian and David DeVault and David R. Traum, 2009 Towards Natural Language Understanding of Partial Speech Recognition Results in Dialogue System In Proceedings of NAACL HLT 2009: Short Papers, pages 53–56, Boulder, Colorado M.A. Anusuya , S.K.Katti 2009 Speech Recognition by Machine: A Review In international Journal of Computer Science and Information Security, Vol. 6, No. 3.

[10] Mohammad , Alessandro Gasparetto, 2014 A system for feature classification of emotions based on Speech Analysis; Applications to Human-Robot Interaction In Proceeding of the 2nd RSI/ISM International Conference on Robotics and Mechatronics , Tehran, Iran.

[11] Poonam.S.Shetake, S.A.Patil, P. M Jadhav, 2014 Review of text to speech conversion methods, international Journal of Industrial Electronics and Electrical Engineering, Volume-2, Issue-8, pp-29-35.

[12] Qirong Mao, Ming Dong, Zhengwei Huang, and Yongzhao ZhanLearning, 2014 Salient Features for Speech Emotion Recognition Using Convolutional Neural Networks, In IEEE transactions on multimedia, vol. 16, no. 8.

[13] Siva Prasad Nandyala and T. Kishore Kumar,2014 Hybrid HMM/DTW based Speech Recognition with Kernel Adaptive Filtering Method In International Journal on Computational Sciences & Applications (IJCSA) Vol.4, No.1, [15] Suma Swamy and K.V Ramakrishnan 2013 An efficient speech recognition system, Computer Science & Engineering: An International Journal (CSEIJ), Vol. 3, No. 4.

[14] Xiang-Lilan, Zhang, Zhi-Gang, Luo,Ming Li, 2014 Merge-Weighted Dynamic Time Warping for Speech Recognition In Journal of Computer Science and Technology ,Volume 29, Issue 6, pp 1072-1082.

[15] Zue V, Glass, J., Goodine, D., Leung, H., Phillips, M, Polifroni, J., Seneff, S,2011 Integration of speech recognition and natural language processing in the mit voyager system, IEEE.