



International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 9, September 2015

A Study of Speech Recognition

Kaladharan N

Assistant Professor, Dept. of .Electrical Engineering, Annamalai University, Tamilnadu, India

ABSTRACT: Speech is the superior personality of the human beings gifted by the nature. Speech helps to deliver the thoughts and messages between human. Human are trying to develop an intelligent systems which can recognize and accept the command via speech, which is known as human computer interface. Speech recognition is the process of the computer recognizing human speech to generate a string of words or commands. Sometimes it is known as automatic speech recognition. Speech recognition is becoming more perplexing and difficult task. The speech recognition research is focuses mainly on large vocabulary, continuous speech capabilities and speaker independence. The design of speech recognition requires cautious attention to some issues like speech representation, depiction of various types of speech Classes, techniques, and database and performance evaluation. This paper presents the review of the different speech recognition system and its recent progress.

KEYWORDS:Speech Recognition, Feature Extraction, MFCC, LPC, HMM, Neural Network, DTW.

I. INTRODUCTION

Speech is the most broadly and frequently used method of communication between humans [1]. There are several spoken languages which are used throughout the world. Humans are very much interested to attach with speech to the computer system effectively and friendly. Such type of system is called speech recognition system. Speech recognition is the process of converting a speech signal to a sequence of word or other spoken entity, by means of an pragmatic as computer platform [3][6].The speech recognition system would support many valuable application like dictation, command and control, embedded applications, telephone directory assistance ,spoken data base querying , medical applications, blind peoples, banking etc.,[4][5] . Speech Recognition is a distinct case of pattern recognition. In speech recognition, two phases are contains viz training and testing. In the training phase, data for known classes are fed to the system. In the recognition phase, the system computes the features of pattern for unknown input and identifies the input with the class whose reference pattern matches these features most closely [8][10][11].In a speech recognition system, the parameters like as vocabulary size, speaker dependency, speaker individuality, period for recognition, Speaking styles, accents, social dialects, gender, type of speech and recognition environment condition are mostly affect the accuracy of recognition [11][15].Robustness and noise tolerant are few of the problems which make speech recognition systems difficult to use. This paper delivers an overview of speech recognition system through its approaches and performance evaluation at various stages.

This paper is organized as follows, Section II describe speech recognition system, Section III represent types of speech recognition method, Section IV approaches to speech recognition, Section V deals speech feature extraction techniques, Section VI denotes performance evaluation, and Section VII deals the conclusion.

II. SPEECH RECOGNITION SYSTEM

InSpeech recognition system is to accurately and efficiently convert a speech signal into a text message transcription of the spoken words independent of the speaker, environment or the device used to record the speech [8]. Some of the major applications of such speech recognition are voice-recognized passwords, voice repertory dialers, automated call-type recognition, call distribution by voice commands, directory listing retrieval, credit card sales validation, speech to text processing, automated data entry etc.[11]. Figure 1, shows basic representation of speech recognition system. It comprises four main stages namely speech analyses or preprocessing, feature extraction coding, language conversion and message accepting.

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 9, September 2015

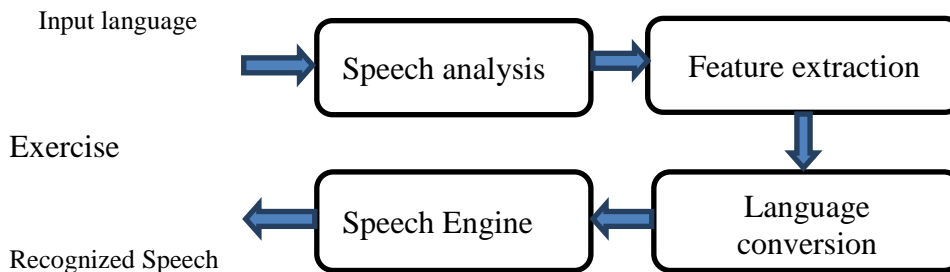


Fig. 1.Representation diagram of Speech Recognition System

Speech analysis block is used to remove the noise, silence and ending point detection from the input speech audio, which are necessary to improve the performance of speech recognition system. Then the feature extraction block is used to remove the unwanted and redundant information and retains only the useful information in type of speaker independent speech recognition. Language conversion unit is used capture the properties of a language and to predict the next word in the speech sequence. Then finally Speech recognition engine block is to convert the input audio into text.

III. TYPES OF SPEECH RECOGNITION

3.1 Types of speech

3.1.1. *Text to speech* -Text to speech will operate a string of text into an audio slide. It is also useful for blind persons to be able to use computers.

3.1.2. *Simple voice device/ commands* - This is the basic form of Speech to text application and this modules are designed to recognize a small number of exact, typically one-word commands and then perform an action.

3.1.3. *Complete dictation recognition*-Complete dictation and recognition software permits the user to read full sentences or paragraphs and translates that data into text on the fluster.

3.2. Isolated word

Isolated word pronunciation needs more gap between words, system is able to recognize the word by difference of utterance. It obtains single words or single utterance at a time. The main disadvantage of this type is choosing different limits affects the entire results. Some examples for isolated words are 'hello', 'go', 'Start', etc.

3.3. Connected word

Connected word, similar to isolated word recognition however user can pronounce the words together with fewer gaps in time. Example of this word is, to spell the word by isolated model like 'g..o..o..d'

3.4. Continuous word

Continues speech, system is able to recognize the word though person is speaking naturally. Really, it's computer dictation. It embraces a great pact of "co articulation", where contiguous words run together without pauses or any other apparent division between words. In some examples, linguistically different sequences have very similar or identical acoustic information. Example for the continuous word is e.g., 'Good day' vs. 'Good A'. This sort of speech makes machine understanding much more difficult.



International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 9, September 2015

3.5. Spontaneous word

Spontaneous speech, system is able to understand words during natural speech with all the gaps and stutters. When a person reads usually in a context that has been previously prepared, but when a person uses spontaneous speech, it is difficult to recognize the speech because of the disfluencies. Examples like "uh" and "um", wrong starts, incomplete verdicts, stuttering, coughing, and laughter and limited vocabulary.

3.6. Types of speaker model

3.6.1. *Speaker dependent model*-Speaker dependent systems are designed for a specific speaker. Some speaker dependent systems require only that the user record a subcategory of system vocabulary to make the entire vocabulary recognizable.

3.6.2. *Speaker independent model*-This speaker-independent system does not necessitate any recording earlier to speaker-dependent system requires that the user record an example of the word, sentence, or phrase system use. Speaker independent system is established to operate for any speaker of a particular type. e.g., American English.

3.6.3. *Speaker adaptable*-A speaker adaptive system is established to adapt its operation to the characteristics of new speakers with speaker dependent data to recognize the speech.

3.7. Types of vocabulary

Small words like 'read', 'write', 'give' etc.,-(three to ten words)

Medium words like 'Sit down', 'Come fast' etc., (ten to hundred words)

Large words like 'As I have some work, so I am very busy' etc., (hundred to thousand words)

Out of words like 'wanna', 'gonna' etc., (depends upon the region)

Very huge words are too many sentences (more than thousand words)

IV. APPROACHES TO SPEECH RECOGNITION

- i. The acoustic-phonetic approach-This approach is based upon theory of acoustic phonetics and guesses.
- ii. Pattern Recognition approach-Pattern training and comparison are the steps includes in pattern recognition approach. Using a well formulated mathematical framework and pledges reliable speech pattern for pattern comparison.
- iii. Artificial intelligence approach-Machine works like human being. It is the combination of the pattern recognition approach and acoustic phonetic method.
- iv. Template based approaches- Unknown speech is equaled against a set of pre-recorded words patterns correct to catch the best Match.
- v. Stochastic approach-Stochastic modeling involves the use of probabilistic models to pact with uncertain or inadequate information.
- vi. Knowledge /Rule based approach - The "skilled" knowledge about distinctions in speech is hand-coded into the system. And it uses set of topographies from the speech, then the training system engenders set of production rules automatically from the samples.
- vii. Dynamic time warping- Dynamic Time Warping is an algorithm for calculating resemblance between two classifications which may diverge in time or speed. This technique is also fairly efficient for isolated word recognition and can be adapted to recognize connected word too.
- viii. Learning based approach-Objective of machine learning is study or determine some kind of knowledge from a documents set.
- ix. Neural Network based approach- An extra broad approach spending neural networks is phoneme recognition. It is better than HMMs model results.



International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 9, September 2015

- x. Statistical based approach- Here the differences in speech are modeled statistically, by the learning events. This approach signifies the current state of the skill.
- xi. Hidden markov model-Hidden Markov Model are very much used because, they can be trained automatically. Hence the HMM is simple and computationally possible to use.
- xii. Vector quantization - Vector Quantization (VQ) is frequently applied in speech recognition and it is valuable for speech coders, that is., efficient data lessening.
- xiii. Support vector machine- SVMs use linear and nonlinear separating hyper-planes for data classification.
- xiv. Bayesian

V. SPEECH FEATURE EXTRACTION TECHNIQUES

5.1. Linear Predictive Coding

LPC is used in audio, speech and signal processing to indicates the spectral surround of a speech digital signal in compacted way. This LPC is very much useful for training good quality of speech with low bit rate and provide accurate speech. LPC analyses the speech signal by estimating the formants, removing their chattels from the speech signal.

5.2. Mel Frequency Cepstral Coefficients (MFCC)

MFCC are the coefficients obtained from the Mel-frequency Cepstrum account. The Mel Frequency Cepstrum (MFC) can be defined by the short time power continuum of speech signal which is calculated as the linear cosine transform of the log power spectrum on a nonlinear Mel scale of frequency(18)(15). The Mel-frequency Cepstrum Coefficient (MFCC) technique is often used to create the fingerprint of the sound files. MFCC are based on the known dissimilarity of the human ear's vital bandwidth frequencies with filters output.. MFCC is also ever more detection uses in music information such as genre classification, audio similarity quantify and so on. It gives high accuracy results for clean speech.

5.3. Signal Subspace Method

The approach engages the use of a signal dependent convert to decompose a noisy signal into two split subspaces, the signal in addition noise subspace, and the noise-only subspace (5).The change employed to perform by using the Karhuenen-Loeve transform (KLT) method. Speech can only period the signal plus noise subspace, called the signal subspace, while noise can extent the entire Euclidean gap. Only the signal subspace is used to estimating the clean signal.

VI. PERFORMANCE OF METHODS

The performance of speech recognition system is usually specified in terms of accuracy and speed. Accuracy is computed by word error rate, whereas speed is measured with the real time factor. Word error rate (WER) is a common metric of the performance of a speech recognition or machine translation system. WER can then be computed by the

$$WER = S+D+I/N$$

Where S is the no. of substitutions, D is the no. of deletions, I is the no. of insertions, and N is the no. of words in the reference. While reporting the performance of a speech recognition system, sometimes word recognition rate (WRR) is by

$$WRR = 1-WER = N-S-D-I/N = H-I/N$$

Where H is N-(S+D) the no. of properly recognized words. Further techniques of accuracy include Command Success Rate (CSR) and Single Word Error Rate (SWER).

VII. CONCLUSION AND FUTURE WORK

This review paper discussed the speech recognition process which includes the feature extraction, types of speech recognition approaches and speech classification techniques. Every method has a different recognition rate with altered classification process. Therefore this research effort is to understand the process of speech recognition and to develop in future an efficient speech recognition system for the people with unlimited accuracy.



International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 9, September 2015

REFERENCES

1. Pratik K. Kurzekar.et.al “Continuous Speech Recognition System: A Review”, Asian Journal of Computer Science and Information Technology. 4: 6 (2014) 62 - 66.
2. Aseem Saxena.et.al “Speech Recognition Using Mat lab”, International Journal of Advances in Computer Science and Cloud Computing. Volume- 1, Issue- 2, Nov-2013.
3. Ms. Vrinda, Mr. Chander Shekhar “speech recognition system for English language”, international journal of advanced research in computer and communication engineering, Vol. 2, issue, January 2013.
4. Vimala.C et.al, “ A Review on Speech Recognition Challenges and Approaches”, World of Computer Science and Information Technology Journal (WCSIT), ISSN: 2221-0741 Vol. 2, No. 1, 1-7, 2012
5. M.A.Anusuya “Speech Recognition by Machine: A Review”,International Journal of Computer Science and Information Security, Vol. 6, No. 3,2009
6. Nnamdi Okomba et.al, “Survey of Technical Progress in Speech Recognition by Machine over Few Years of Research”, IOSR Journal of Electronics and Communication Engineering (IOSR-JECE) e-ISSN: 2278-2834, p- ISSN: 2278-8735.Volume 10, Issue 4, Ver. I (Jul - Aug .2015), PP 61-67
7. Kavita S.Yadav , M.M.Mukhedkar, “Review on Speech Recognition”, International Journal of Science and Engineering, vol 1,number 2,2013 pp-61-70
8. Anchal Katyal, Amanpreet Kaur, Jasmeen Gill, “ Automatic Speech Recognition: A Review”,International Journal of Engineering and Advanced Technology (IJEAT) ISSN: 2249 – 8958, Volume-3, Issue-3, February 2014
9. Yogesh K. Gedam.et.al , “Development of Automatic Speech Recognition of Marathi Numerals - A Review”, International Journal of Engineering and Innovative Technology (IJEIT) Volume 3, Issue 9, March 2014.
10. Asm SAYEM, “Speech Analysis for Alphabets in Bangla Language: Automatic Speech Recognition”, International Journal of Engineering Research (ISSN: 2319-6890)(online),2347-5013(print),Volume No.3, Issue No.2, pp : 88-93 01 Feb. 2014.
11. Mayur R Gamit. et.al, “Classification Techniques for Speech Recognition: A Review”, International Journal of Emerging Technology and Advanced Engineering, Volume 5, Issue 2, February 2015.
12. Akanksha Singh Thakur et.al, “Speech Recognition Using Euclidean Distance”, International Journal of Emerging Technology and Advanced Engineering, Volume 3, Issue 3, March 2013
13. Preeti Saini et.al, “Automatic Speech Recognition: A Review”, International Journal of Engineering Trends and Technology- Volume4, Issue2-2013.
14. Rashmi C R, “Review of Algorithms and Applications in Speech Recognition System”, International Journal of Computer Science and Information Technologies, Vol. 5 (4), 2014, 5258-5262.
15. Shanti Therese S.et.al, “Review of Feature Extraction Techniques in Automatic Speech Recognition”, International Journal of Scientific Engineering and Technology, Volume No.2, Issue No.6, pp.: 479-484 1 June 2013.
16. Santosh K.Gaikwad.et.al “A Review on Speech Recognition Technique”, International Journal of Computer Applications (0975 – 8887), Volume 10– No.3, November 2010.
17. Kishori R. Ghule, R. R. Deshmukh, “Feature Extraction Techniques for Speech Recognition: A Review”, International Journal of Scientific & Engineering Research, Volume 6, Issue 5, May-2015.
18. Harpreet Singh.et.al, “ A Survey on Speech Recognition”, International Journal of Advanced Research in Computer Engineering & Technology (IJARCET) Volume No. 2, Issue No. 6, June 2013.
19. Hooman Heidari , Suresh Gobe, “Isolated Word Command Recognition for Robot Navigation” International Symposium on Robotics and Intelligent Sensors 2012.
20. Saroj B. Jadhav.et.al, “Speech Recognition in Marathi Language onAndroid O.S.”, International Journal of Research in Computer and Communication Technology, Vol 3, Issue 8, August – 2014.
21. Swati Atame.et.al, “A Survey On: Continuous Voice Recognition techniques”, International Journal of Emerging Trends & Technology in Computer Science (IJETTCS),Volume 4, Issue 3, May-June 2015.