International Journal of Electronics Engineering (ISSN: 0973-7383
Volume 11 • Issue 1 pp. 971-976 Jan 2019-June 2019 www.csjournals.com

Speech Recognition Using Machine Learning: A Review

Sunila Godara
Guru Jambheshwar University of Science & Technology, Hisar, Haryana.

Abstract: Speech it s a pressure wave that travel through the air created by the vibration of the larynx by the opening and the closing of mouth and human experiments on speech recognition programs using machine learning techniques. In this paper various speech types and speech recognition system using machine learning techniques are reviewed.

I. INTRODUCTION

Program designers notice human experiments on speech recognition programs; they get valuable insight into technological issues and barriers that they can never see. Testing speech recognition products for the general use is again an important step when looking for a product to find a viable solution for customers later. While talking to the microphone, and the computer transforms your voice into a text that will use your word processor or other software on your computer. The computer can repeat what you just said or you are prompted to anticipate what you expect afterwards. This is the key promise of interactive speech recognition. In past staccato style speech recognition programs were made claiming that they leave the gap between each two can think as humans. New voice recognition systems, of course, are much easier to use. Everyone can speak at a naturally without breaking words between a clear break. However, you cannot really use the word "natural" as producers demand. You need to speak clearly as you do when you talk to one of the dictations or phone. Speech it s a pressure wave that travel through the air created by the vibration of the larynx by the opening and the closing of mouth.

Types of speech

This system is used to recognition may be divide in the many parts of capability by know tha the list of words we have. Many parts of speech recognition are described here:

Isolated Speech

Isolated speech generally take a gap between two words with a special rule by which the words differ from one another.

Connected Speech

Connected speech or connected discourse, is a continuous sequence of sounds forming conversations in spoken language

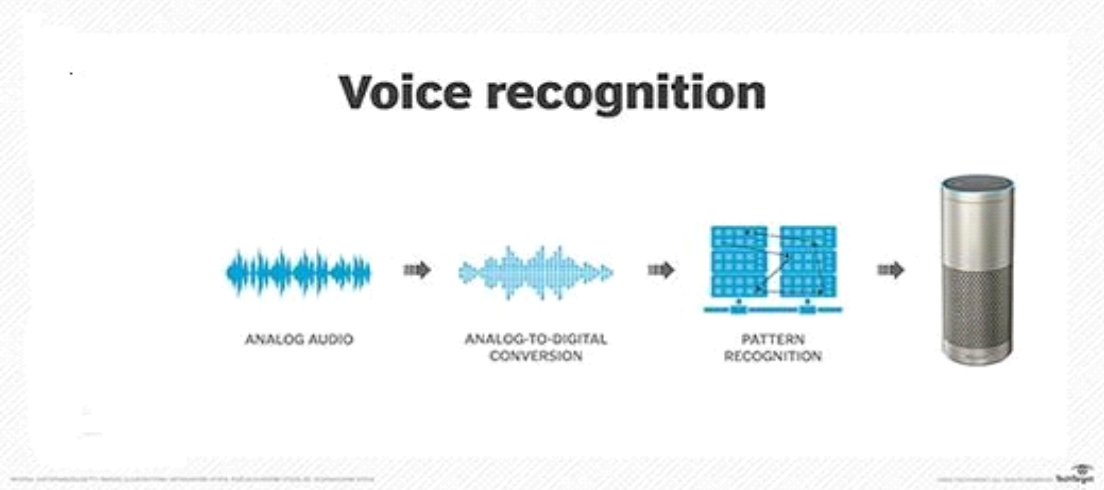
Continuous speech

Continuous speech is a system operates on speech by which words are connected together i.e. not separated by pauses

Speech Recognition:

It is a growing technology that impact the telephony and television computing and other devices. This technology available many years. But has not so practical the real of world due to the high price of computing resources of lack of simple standards to integrate SR with software applications. The business field was not fully depend on SR the voice-totext application only generated $48 million of revenue n the US during 1996. According to William this technology not impact

- quality - the ability of understand the simple spoken word and the ability of understand the difference between the words like
- price of application and the resources
- lack of knowledge between SR software, operating systems, and applications.



Acoustic Model

This model designed by basic feature of the machine recordings of speech and their text changing it has a very small representation in the field this system is work with the sounds and it is change with the outside so is basically depends on the outside

Language Model

Language iused by the simple language and it have a very simple application. Such as it can perdict the next word according to the previous sequence of speech. This software is use the way of dictionary in language pattern.

Speech engine

The problem of speech reputation and the audios text-sharing, which was used to access the data and the software algorithms and data the first operation was an appropriate way for digitalization, that transformation. The sound signal take the relevant format looking for the game and it wants to get you know the signals by looking at the the data.

Speech Synthesis

Text-to-phone convert and every word of phonemes. Processings are the fundamental unit of sound and the language. U.S english hasiaround 47 phonemes with the consonant some vowel sounds. Different languages have separate sets of sounds For example Japan has fewer phonemes sounds not found in English such as "ts" n "tsunami".
A speech tool speech engine that translates the basic text. for speech synthesis as well as a number of basics classes and interfaces. As a word system, many vocabulary functions are inherited from the nterface of the java.x for this system.

International Journal of Electronics Engineering (ISSN: 0973-7383
Volume 11 • Issue 1 pp. 971-976 Jan 2019-June 2019 www.csjournals.com

Synthesizer as an Engine

A speech synthes is the engine that will converts text into speech. The speech synthesizer package give information about the word we are using in the system it will convert the text into the words on the screen by the input we are given to the system this system will make the interface to the user the interface is used to support the system for better usement and the better inerface to support speech synthesis plus it has the additional supporting class that we are using.

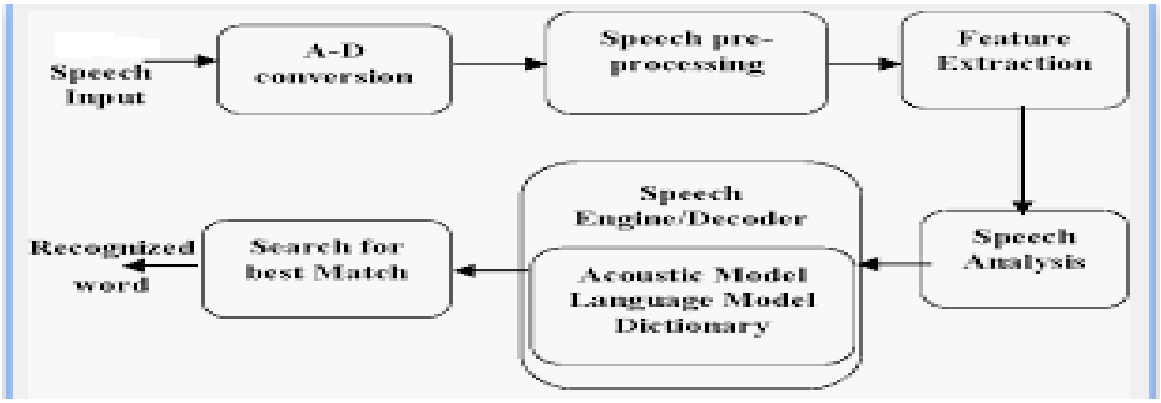


Fig 2: Steps of speech recognition system

Applications:-

1. In the work place:-

We can use this system at work staions like when we search for documents on your computer

We can creat a Graph or tables using data of audio type

We can dictate the information you want to be in corparated into a document

We can print Documents on request

We can start video confrences

2. Call Steering:-

waiting in queue to get through an operater can be very frustrating the customers and by the speech recognition system the customer can call the the department individual and can reach to the right department

3. Telephony:-

Some Mail system is allow the user to call the name of particular mail id by which it can directly call or contact to the specific system

4. Medical Disability:-

Many features in the system we are using in the medcal field the person who are physical disable and the about physical and more. For example,the people who are hearing problem can take speaker text convert into the voice

Weakness of speech Recognition system

With all these benefits and benefits, 100% recognition system can not develop. it has the of point that remove the accuracy and and the of the system. It's easy to dentify people, but t's hard on the car. Comparison human speech recognition programs s relatively ntelligent, because my mind To conquer God's power and philosophy. Understanding and nteractivity s normal and should first be understood n terms of the position of the computer program and should create a proper balance words, noise and distances. There's a manSpeech Speech build ability Although the device needs training, the computer should also help with other sounds. Voice input Enter up to 150 words or more words long f spoken faster. This perspective of the so-called recognition programs creates digitally digitally easy text and millions of words in a short of time

The factors which are considerable

Homonyms: is the words different and meaningful, but the same meaning, for example, "there", "they" and "to become"? it is very difficult to distinguish words from a computer.

Overlapping speech The second argument of this process to understand the conversation of different users, making difficult for existing users to talk to multiple users simultaneously.

Noise factor The program should be clear and clear. Any additional sound may cause interference, first and foremost, the device should be precisely dialed out of a noisy environment and mixing words and mixing words

II. RELATED WORK

Thiang et al. (2011) presented speech recognition using Linear Predictive Coding (LPC) and Artificial Neural Network (ANN) to control the movements of mobile robots. The input signals were taken directly from the microphone and therefore the extraction of LPC and ANN was carried out [39].

Vimala.C and Dr. V. Radha (2012) proposed a speaker-independent, isolated speech recognition system for the Tamil language. Feature extraction, acoustic model, pronunciation dictionary and language model were implemented using HMM, giving an accuracy of 88% in 2500 words [29].

Cini Kurian and Kannan Balakrishnan (2012) discovered the development and evaluation of various acoustic models for the continuous vocal recognition of Malayalam. In this document, the HMM is used to compare and evaluate context-dependent (CD), context-independent (CI) and context-dependent (CD-related) models of this 21% CI model. The database consists of 21 speakers, including 10 men and 11 women.

Suma Swamy et al. (2013) introduced an efficient speech recognition system that was tested with cepstrum coefficients of the Mel frequency (MFCC), vector quantization (VQ) and HMM and recognized speech with an accuracy of 98%. The database consists of five words spoken ten times by four speakers.

Annu Choudhary et al. (2013) proposed an automatic speech recognition system for isolated and connected words in Hindi using the Hidden Markov Model Toolkit (HTK). Hindi words are used for the data set extracted from the MFCC, and the recognition system has achieved an accuracy of 95% for isolated words and an accuracy of 90% for related words [3].

Preeti Saini et al. (2013) proposed automatic Hindi speech recognition with HTK. Isolated words are used to identify 10-state speech in the HMM topology, which gave 96.61% [31].

Md. Akkas Ali et al. (2013) presented the automatic speech recognition technology for Bengali words. The features were extracted using Linear Predictive Coding (LPC) and Gaussian Mix Model (GMM). 100 words were recorded 1000 times with an accuracy of 84% [25].

Maya Moneykumar et al. (2014) developed the identification of words in Malayalam for the speech recognition system. The proposed work was performed using syllable-based segmentation using HMM in MFCC for feature extraction [24].

Jitendra Singh Pokhariya and Dr. Sanjay Mathur (2014) introduced Sanskrit speech recognition via HTK. MFCC and two HMM states were used for the extraction, which give an accuracy of 95.2% to 97.2% [16].

Geeta Nijhawan et al., 2014, developed a real-time speaker recognition system for Hindi words. Characteristic extraction performed with MFCC using the Linde, Diver and Gray quantization algorithm (VQLBG). The Voice Activity Detector (VAC) has been proposed to eliminate silence [10].

III. CONCLUSION

This paper reviewed the work done on speech recognition using machine learning. It is growing technology and has great impact in the future .it is widely used in various fields and the largest companies of the world are working in field of speech recognition to see a better future.

REFERENCES

[1] A. Ganapathiraju, J. Hamaker, and J. Picone, "Hybrid SVM/HMM architectures for speech recognition," in Speech Transcription Workshop, 2000.
[2] A. J. Robinson, L. Almeida, J. Boite, H. Bourlard, F. Fallside, M. Hochberg, D. Kershaw, P. Kohn, Y. Konig, N. Morgan, J. P. Neto, S. Renals, M. Saerens, C. Wooters, H. Speechproducts, and H. Speechproducts, "A neural network based, speaker independent, large vocabulary, continuous speech recognition system: The wernicke project," in Proc. EUROSPEECH'93, 1993, pp. 1941– 1944.
[3] L. R. Rabiner, "A tutorial on hidden markov models and selected applications in speech recognition," in Proceedings of the IEEE, 1989, pp. 257–286.
[4] B. P. Bogert, M. J. R. Healy, and J. W. Tukey, "The quefrency analysis of time series for echoes: Cepstrum, pseudo autocovariance, cross-cepstrum and saphe cracking," in Proceedings of the Symposium on Time Series Analysis M. Rosenblatt, Ed. New York Wiley, 1963, ch. 15, pp. 209–243.
[5] Hermansky, "Perceptual linear predictive (PLP) analysis for speech," Journal of Acoustical Society of America, vol. 87, pp. 1738–1753, 1990.
[6] ETSI, "Speech processing, transmission and quality aspects (STQ); distributed speech recognition; advanced front-end feature extraction algorithm; compres- sion algorithms," in ETSI ES 202 050 Ver.1.1.3, Nov. 2002.
[7] P. Viola and M. Jones, "Rapid object detection using a boosted cascade of simple features," in Proc. of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2001, pp. 511–518.
[8] M. B. Stegmann, R. Fisker, B. K. Ersbøll, H. H. Thodberg, L. Hyldstrup, Active appearance models: Theory and Cases, Proc. 9th Danish Conference on Pattern Recognition and Image Analysis,vol. 1, pp. 49-57, AUC, 2000

[9]H. McGurk and J. MacDonald, “Hearing lips and seeing voices,” Nature, vol. 264, pp. 746–748, 1976.
[10] A. Nefian, L. Liang, X. Pi, X. Liu, and K. Murphy, “Dynamic bayesian networks for audio-visual speech recognition,” EURASIP Journal on Applied Signal Pro- cessing, pp. 1–5, Nov. 2002.
[11]H. Hermansky, D. P. W. Ellis, and S. Sharma, “Tandem connectionist feature extraction for conventional HMM systems,” in Proc. ICASSP, 2000, pp. 1635– 1638.
[12] David H. Wolpert, “Stacked generalization” in Neural Networks, v.5 n.2, p.241-259, 1992