International Journal of Trend in Scientific Research and Development (IJTSRD)
Volume 6 Issue 3, March-April 2022 Available Online: www.ijtsrd.com e-ISSN: 2456 – 6470

A Survey on Speech Recognition with Language Specification

Dr. Preeti Savant[1], Lakshmi Sandhya H[2]

[1]Assistant Professor, School of CS & IT, Jain University, Bangalore, India
[2]MCA Department, Jain University, Bangalore, Karnataka, India

ABSTRACT

As a cross-disciplinary, speech recognition is entirely based on the speech as the survey object. Speech recognition allows the machine to convert the speech signal into text or commands via the process of identification and understanding. Speech recognition involves in various fields of physiology, psychology, linguistics, computer science and signal processing, and is even related to the person's body language, and its goal is to achieve natural language communication between man and machine. The speech recognition technology is gradually becoming the key technology of the IT man machine interface. This paper describes the development of speech recognition technology and its basic principles, methods, reviewed the classification of speech recognition systems, speech recognition approaches and voice recognition technology, analyzed the problems faced by the speech recognition.

KEYWORDS: *speech detection, speech recognition, audio to text processing*

How to cite this paper: Dr. Preeti Savant | Lakshmi Sandhya H "A Survey on Speech Recognition with Language Specification" Published in International Journal of Trend in Scientific Research and Development (ijtsrd), ISSN: 2456-6470, Volume-6 | Issue-3, April 2022, pp.343-347, URL: www.ijtsrd.com/papers/ijtsrd49370.pdf

IJTSRD49370

1. INTRODUCTION

Speech recognition is the machine on the command of human voice to identify and understand and react accordingly. It is based on the speech as the survey object, it allows the machine to automatically identify and understand human language through speech signal processing and pattern recognition. The speech recognition technology allows the machine to turn the speech signal into the appropriate text through the process of identification and understanding. It has a close relationship with acoustics, phonetics, linguistics, information theory, pattern recognition theory and neurobiology disciplines. With the development of computer hardware, software and information technology, speech recognition technology is gradually becoming a key technology in the computer information processing technology. Things to develop speech recognition technology is also widely used in voice activated telephone exchange information networks, medical services, bank services, industrial control and people's lives. Many experts believe that speech recognition is one of the 2000-2010 IT technology scientific and technological developments[1].

1.1. Languages of the world -

Counting the number of languages in the world is not an task. An estimate for the total number of languages in the world can be found on the Ethnologue website. They have defined a living language as "one that has at least one speaker for whom it is their first language". That's the reason, extinct languages and spoken languages as a second language are removed from these counts. Based on the definition, Ethnologue lists 6,909 known living languages. This list contains 473 languages that are classified as nearly extinct, i.e. when "only a few elderly speakers are still living". It is important to know that Ethnologue's list have both verbal and visual-kinetic spoken languages. The later ones are known as sign languages, which are used for everyday communication by the deaf; these spoken languages combine hand gestures with lips articulation and facial mimics. Almost all countries over the world declare their own national language[4].

1.2. BETWEEN MACHINE AND HUMAN RECOGNITION

Although there had been many significant researches in this field, but we are still trying to establish a

system for unify the speech. There are some major issues in this field such as the question of relevancy between human-human communication to human-machine communication, design of good architecture for speech process, the importance of individuality and physiological mechanism and question about communicative nature of speech, formalism adaptation quality, constructive nature, random variability of speech are still unconcluded. In fact, as in 1994 Moore presented the 2 themes for better understanding of the concept of speech pattern processing are wants a deep research [7].

2. SPEECH RECOGNITION APPROACHES

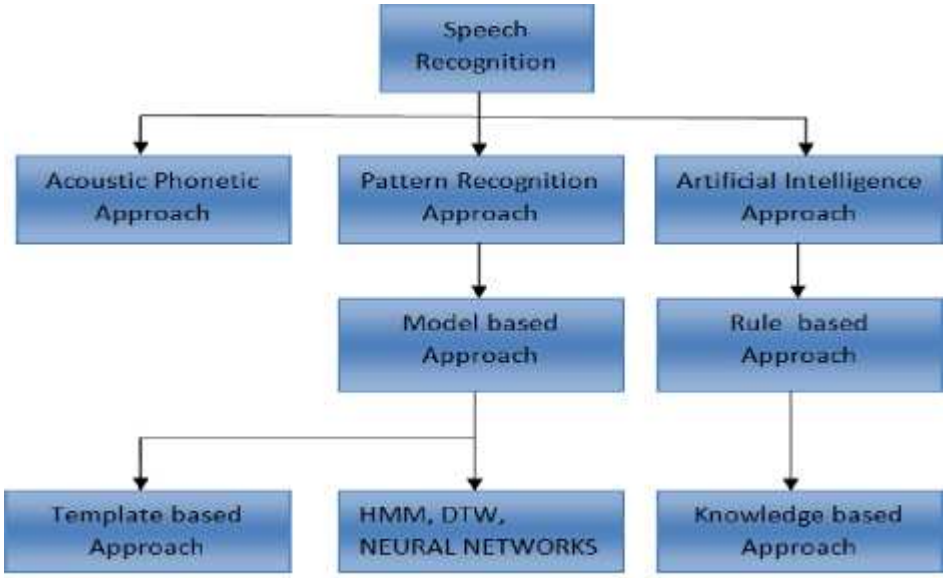

Fig 2.1: Speech recognition approaches[7]

A. Acoustic phonetic approach:

It is a traditional method of speech recognition, it claims that the spoken language contains finite characteristic phonetic units, also known as phonemes and these units are generally considered by a set of acoustics properties that are established in the speech signal over time. The acoustic properties of phonetic units are highly variable, both with speakers and with neighboring sounds, it is also known as vocalization effect, it is assumed in the acoustic-phonetic approach that the rules leading the unpredictability are straightforward and can be willingly learned by a system[7].

B. Pattern Recognition Approach:

The approach depends on the classification of input data into modules via the extraction of significant features or attributes of the data from a contextual of irrelevant content. This method collects the raw data and analyses it statistically. And then generate the pattern which depends on statistical feature. It produces same results in same fundamental structural pattern. If there are two different situations, then system should be developed or trained for desired result.

Thus with the nature of problems the speech approach system can provide various mathematical and statistical techniques to find out the desired solutions or the results. There are best known approaches for voice recognition pattern. Template based approach has a collection of typical speech patterns.

This approach works on reference system. First we maintain a dictionary of refereed speech pattern.

This approach works on reference system. First we maintain a dictionary of refereed speech pattern.

There are many drawbacks of this approach, in this paper we are going to discuss three most common and important are as follows -

1. Dependent on the lexicon size of templets.
2. Nonlinear time alignment is important factor; Same word is spoken by the same person in different rates).
3. Reliability determines the word limitations.

In Statistical or Stochastic Based approach, modeling of speech on the basis of variation of speech, all variation factors managed statistically. for modelling we can use HMM.

This approach is depending on previous assumptions. May be incorrect assumption or previous modeling restrict the system's performance[7].

C. Artificial Intelligence approach:

AI Technique for Speech Recognition is based on Neural Networks. Creation of natural human sources to communicate with the computer is currently one of the greatest challenges of modern science. Computer Simulation proves the effectiveness of results. There are three approaches to create artificial neural networks[7].

3. CLASSIFICATION OF SPEECH RECOGNITION

Classification of speaker recognition is shown below in Fig.3.1.

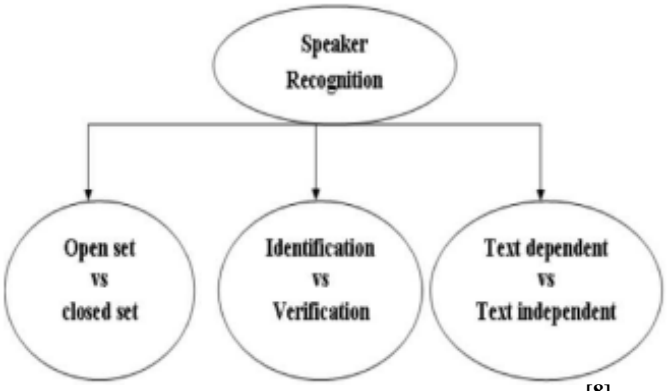

Fig 3.1: Block diagram of classification[8]

A. Open Set Vs Closed Set –

This type of classification is based on the set of trained speakers available in a system.

1. Open Set: An open set system can have n number of trained speakers. We have an open set of speakers and the number of speakers is always greater than one.
2. Closed Set: A closed set system has only a specified (fixed) number of users registered to the system.

B. Identification Vs Verification –

Speaker recognition is a biometric system which takes the speech samples, extracts the characteristic features and performs the computing task of validating a user"s claimed identity. As shown in Fig.3.2, it is performed in two parts: Identification and verification.

Verification executes a binary decision which consists of determining whether the person speaking is the same person either he or she claims to be or to put it in other words verifying their identity. On the other hand „Identification" does the job of matching (comparing) the voice of the speaker with a database of reference templates in an attempt to identify the speaker [8].

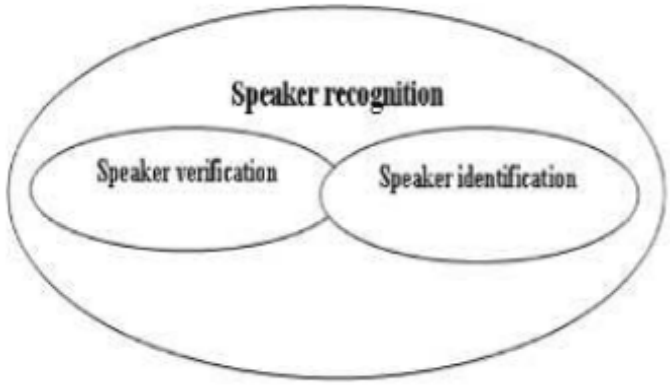

Fig 3.2: Speech recognition[8]

Speaker identification and verification are mostly considered to be the most natural and economical methods for avoiding unauthorized access to physical locations or computer systems.

Speaker identification: The process of determining where the registered speaker provides a given utterance.

Speaker verification: The process of accepting or rejecting the identity claim of a speaker. Both the figures depict the differences between ASI (Automatic Speaker Identification) and ASV (Automatic Speaker Verification) systems [8].

C. Text-Dependent Vs Text-Independent –

This type is based on the text uttered by the speaker during the identification process.

1. Text-Dependent: The test utterance is the same to the text used in the training phase. The test speaker has prior knowledge of the system.
2. Text-Independent: The test speaker doesn't have prior knowledge about the contents of the training phase and can speak anything [8].

4. LITERATURE REVIEW

Speech recognition is the statement or command of human speech. It is based on the voice as the survey object, it allows the machine to automatically identify and understand human spoken language through speech signal processing and pattern recognition. The speech recognition technology allows the machine to turn the speech signal into the appropriate text or command via the process of identification and understanding. Speech recognition involves a wide range. It has a close relationship with acoustics, phonetics, linguistics, information theory, pattern recognition theory and neurobiology disciplines. With the development of computer hardware, software and information technology, speech recognition technology is gradually becoming a key technology in the computer information processing technology[5] [7].

In [1] - "Overview of the Speech Recognition Technology" which got published in the year 2012, the author has explained about the overview of the speech recognition pattern.

Also he has explained about how the Hidden Markov Model and Artificial Neural Network is used for speech recognition technology.

In [2] - "Why is speech recognition difficult" by Forsberg, Markus, explained that speech recognition is one of the major thing happening, where we find most useful things.

Every aspect has its own working advantages and difficulties.

But the difficulties and problems that occur when we target more than a single use is neatly explained in this paper.

In [3] – "A study on speech recognition system: a literature review" by Gupta, Shikha, A. Pathak, and A. Saraf, the authors have tried to explain about the speech recognition techniques and the modelling techniques of speech recognition.

This paper also present the list of techniques with their properties of Feature extraction and Feature matching.

Through this review paper it is found that MFCC is widely used for feature Extraction and VQ is better over DTW.

In [4] - "Automatic speech recognition for under-resourced languages: A survey." which was published in the year 2014 by Besacier, Laurent, et al., where this paper demonstrate that speech processing for under resourced languages is an active field of research, which has experienced significant progress during the past decade.

Although much of the recent progress has been the result of the technical developments summarized.

It is also clear that the developments will be necessary to clear many of the relevant issues.

In [5] - "Speech Recognition with Gender Identification and Speaker Diarization," published by N. M and A. S. Ponraj in the year 2020.

The research is concentrated on the speech analysis, where the MFCC and GMM is being used to derive the parameters of the model.

The model which gives accuracy of 92% while training. Real time data is taken with dataset to train gender identification with speaker identification.

Speaker diarization is being calculated with the duration and the overlapping of the speaker in the audio samples. Audio processing carried out with this model can be used in prediction of class based on gender.

In [6] - "speech recognition", this paper was published by Zwass, Vladimir in the year 2016.

As we know, currently research is focusing on creating and developing systems that would be much more robust against variability and shift in acoustic environment, speaker characteristics, language characteristics, external noise sources etc.

The author has found that HMM is the best technique in developing language model.

In [7] - "A Survey: Speech Recognition Approaches and Techniques" published in the year 2018 by A. P. Singh, R. Nath and S. Kumar.

In this paper, the fundamentals techniques and methods for the speech recognition are discussed. The various approaches available for developing an ASR system that are clearly explained with its merits and demerits.

ASR system performance is depending on two factors first is adopted feature extraction techniques, and second is speech recognition approach for the particular language.

In [8] - "A Review Article on Speaker Recognition with Feature Extraction." Published by Chaudhary, Parvati J., and Kinjal M. Vagadia in the year 2015.

In this survey paper, there is a discussion on classification of speaker recognition that can be used for many speech processing applications especially security and authentication.

The most commonly used feature extraction techniques are discussed here among which MFCC is the commonly used.

5. CONCLUSION

From the problems faced by the speech recognition, speech recognition systems in order to be widely used still have a lot of areas for improvement. However, it is foreseeable in the near future that, with the voice recognition technology continues to progress, the speech recognition system will be more in-depth, the application of speech recognition systems will be more extensive.

A variety of speech recognition systems will appear in the market, people will adjust their speech patterns to adapt to a variety of recognition system Human beings in the short term is also impossible to create a people comparable to the speech recognition system, to build similar system is still a challenge towards humanity, we can only forward step by step direction to improve the speech recognition system.

REFERENCES

[1] J. Meng, J. Zhang and H. Zhao, "Overview of the Speech Recognition Technology," 2012 Fourth International Conference on Computational and Information Sciences, 2012, pp. 199-202, doi: 10.1109/ICCIS.2012.202.M. M. Zoltán Balogh, "Motion Detection and Face Recognition using Raspberry Pi, as a Part of, the Internet of Things," Acta Polytechnica Hungarica, vol. 16, no. 3, 2019, pp.112-120.

[2] Forsberg, Markus. (2003). Why is speech recognition difficult.

[3] Gupta, Shikha, A. Pathak, and A. Saraf. "A study on speech recognition system: a literature review." *International Journal of Science, Engineering and Technology Research* 3.8 (2014): 2192-2196.

[4] Besacier, Laurent, et al. "Automatic speech recognition for under-resourced languages: A survey." *Speech communication* 56 (2014): 85-100.

[5] N. M and A. S. Ponraj, "Speech Recognition with Gender Identification and Speaker Diarization," 2020 IEEE International Conference for Innovation in Technology (INOCON), 2020, pp. 1-4, doi: 10.1109/INOCON50539.2020.9298241.

[6] Zwass, Vladimir. "speech recognition". Encyclopedia Britannica, 10 Feb. 2016,

[7] Arora, Shipra J., and Rishi Pal Singh. "Automatic speech recognition: a review." *International Journal of Computer Applications* 60.9 (2012).

[8] A. P. Singh, R. Nath and S. Kumar, "A Survey: Speech Recognition Approaches and Techniques," 2018 5th IEEE Uttar Pradesh Section International Conference on Electrical, Electronics and Computer Engineering (UPCON), 2018, pp. 1-4, doi: 10.1109/UPCON.2018.8596954.

[9] Chaudhary, Parvati J., and Kinjal M. Vagadia. "A Review Article on Speaker Recognition with Feature Extraction." *International Journal of Emerging Technology and Advanced Engineering* 5, no. 2 (2015): 94-97.