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, India

---***---

Abstract - The voice is most important and one of the natural forms of communication among from livelihood. The voice is a signal of countless information. In this field we can collect infinite data and information in the voice signal. There are many aspects in this field of research like voice recognition, voice verification, voice synthesis, speaker recognition; speaker identification language recognition etc. Voice recognition is a dominant tool of the information interchange using the acoustic signal. Due to these reasons the voice signal is the leading subject of research from many decades. With the help of microphone, we can easily store or capture voice of any speaker. All captured words are later-on recognized by voice recognizer, and in the end, system outputs the recognized words. Voice recognition is basically the science of talking with the computer, and having it correctly recognized.

Key Words*: speech detection, speech recognition, audio to text processing, voice recognition, *speech detection, speech recognition, audio to text processing

1. INTRODUCTION

Speech recognition involves many 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. The paper describes the development of speech recognition technology and its basic principles, methods, reviewed the classification of speech recognition systems and voice recognition technology, analyzed the problems faced by the speech recognition.

Speech recognition is the machine on the command of human voice to identify and understand and react accordingly. The speech recognition technology allows the machine to turn the speech signal into the appropriate text through 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. 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[1].

2. OBJECTIVES

The objectives of this project are as follows: -

This project is completely a scripted language. We are using multiple languages to produce the output.

The advantage is that, we can have the voice over for the translated sentence for better understanding.

3. LITERATURE REVIEW

The author provided the overview of the speech recognition pattern in [1] - "Overview of the Speech Recognition Technology," which was published in 2012. He also explained how speech recognition technology uses the Hidden Markov Model and Artificial Neural Network.

Forsberg, Markus, explained in [2] - "Why is speech recognition tough" that voice recognition is one of the key things happening, where we find the most helpful information. Every aspect has its own set of benefits and drawbacks. However, this study succinctly explains the difficulties and complications that arise when we target more than one usage.

The authors of [3] – "A study on voice recognition system: a literature review" by Gupta, Shikha, A. Pathak, and A. Saraf – have attempted to describe speech recognition techniques and modelling methodologies. This document also includes a list of strategies for feature extraction and feature matching, as well as their attributes. This review research discovered that MFCC is commonly utilised for feature extraction and that VQ outperforms DTW.

In [4] - "Automatic voice recognition for under-resourced languages: A survey," published in 2014 by Besacier, Laurent, et al., this work shows that speech processing for under-resourced languages is an active field of research that has made great progress over the last decade. Although the technical developments outlined have accounted for most of recent progress. It is also evident that the changes will be required to resolve many of the pertinent issues.

In the year 2020, N. M. and A. S. Ponraj published [5] - "Speech Recognition with Gender Identification and Speaker Diarization." The focus of the research is on speech analysis, with the MFCC and GMM being utilised to create model parameters. During training, the model had a 92 percent accuracy rate. Gender identification with speaker identification is trained using real-time data from the dataset. The duration and overlapping of the speakers in the audio samples are used to determine speaker diarization. The audio processing performed with this model can be used to predict gender-based class.

This work was published by Zwass, Vladimir in the year 2016 in [6] - "voice recognition." As we all know, research is currently focused on inventing and constructing systems that are far more resilient to changes in the acoustic environment, speaker characteristics, language characteristics, external noise sources, and so on. The author discovered that HMM is the most effective method for creating language models.

A. P. Singh, R. Nath, and S. Kumar published [7] - "A Survey: Speech Recognition Approaches and Techniques" in 2018. The essential strategies and approaches for speech recognition are presented in this study. The many options available for constructing an ASR system are described in detail, together with their benefits and drawbacks. The performance of an ASR system is determined by two factors: first, the feature extraction techniques used, and second, the speech recognition approach used for the specific language.

"A Review Article on Speaker Recognition with Feature Extraction" is found in [8]. Parvati J. Chaudhary and Kinjal M. Vagadia published the book in 2015. This survey study discusses speaker recognition classification, which can be utilised for a variety of speech processing applications, including security and authentication. The most popular feature extraction techniques are detailed here, with MFCC being the most popular.

4. SYSTEM ANALYSIS

In this paper, Speech Recognition is discussed. Speech Recognition is the procedure and the related innovation for changing over the discourse signal into its comparing grouping of words or other semantic substances.

i. Automatic speech recognition -

Automatic Speech Recognition is basically used to change over expressed words into computer content. Also, Automatic Speech Recognition is utilized for validating clients by means of their voice (biometric confirmation) and playing out an activity dependent on the directions characterized by the human. Generally, Automatic Speech Recognition requires preconfigure for voices of the essential clients. Human needs to prepare the Automatic Speech Recognition framework by putting away discourse examples of them into the framework[5].

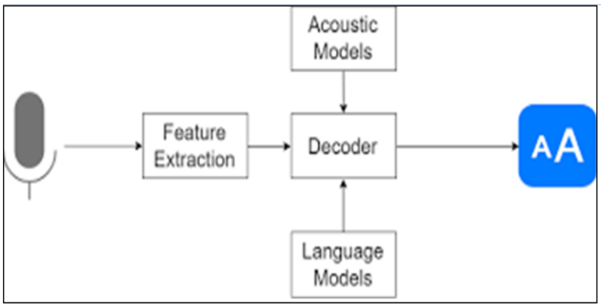


Fig-1: Automatic Speech Recognition

ii. Speech to text conversation –

Speech is an uncommonly methodology for human computer collaboration: it is "without hands"; it requires just humble equipment for obtaining (top-notch mouthpiece or amplifiers); and it shows up at an unassuming piece rate. Perceiving human speech, particularly persistent (associated) speech, without troublesome preparing (speaker autonomous), for a vocabulary adequate multifaceted nature (60,000 words) is exceptionally hard. Be that as it may, with present day forms, stream outline, calculations, while developing the framework speech recognition is being made using the decoder. In this framework, speech-to-text convertor is built with ASR[5].

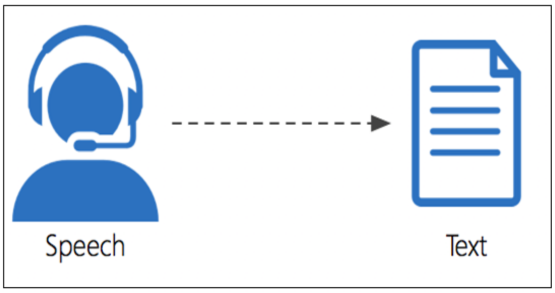


Fig-2: Speech to text

4.1 Types of speech –

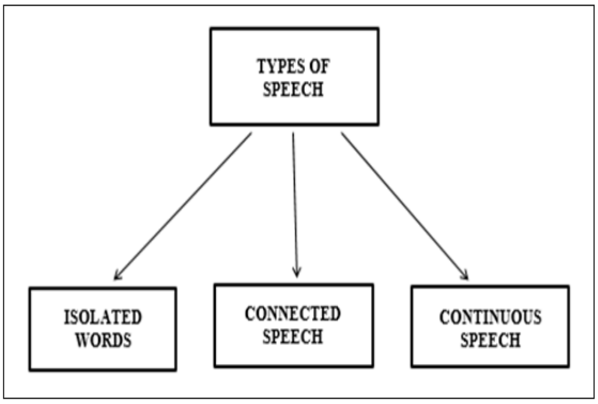


Fig-3: Types of Speech

Speech can be generally classified into following categories:

i. Isolated words:

Isolated word accepts single word at a time. These method have "Listen/Not-Listen" states, where they require the speaker to wait between utterances. Isolated Utterance might be the better name for this type[3][7].

ii. Connected words:

Connected word speech recognition is the type where the words are separated by pauses. Connected word speech recognition is a type of fluent speech strings where the set of strings is derived from small-to-moderate size vocabulary such as digit strings, spelled letter sequences, combination of alphanumeric. Like isolated word speech recognition, this set too has a property that the basic speech-recognition unit is the word/phrase to much extent[3][7].

iii. Continuous speech:

Continuous speech recognition deals with the speech where words are connected together instead of separated by pauses. As a result of information about words, co-articulation, production of surrounding phonemes and rate of speech effect the performance of continuous speech recognition. Recognizer with continuous speech capabilities are some of the most difficult to create because they utilize special methods to determine utterance boundaries[2][7]. It allows user to speak naturally, where the computer will examine the content there are special methods used to determine utterance boundaries and various difficulties occurred in it[3].

4.2 Difficulties of Speech –

Following are the some of the difficulties/problems faced by speech recognition:

i. Spoken language is not equal to written language:

Spoken language for many years been viewed just as a less complicated version of written language, with the main difference that spoken language is grammatically less complex and humans make more errors while speaking. Anyways, it has become clear in the last few years that spoken language is different from written language. In speech recognition system, we have to identify and note down these differences. In spoken language, there is often a radical reduction of morphemes and words in pronunciation. The frequencies of words, collocations and grammatical constructions are highly different between spoken and written language[2].

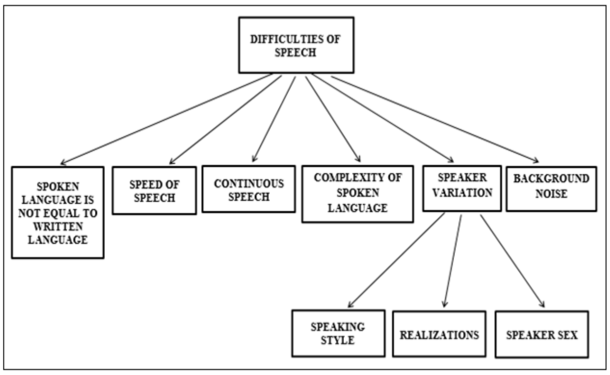


Fig-4: Difficulties of speech

ii. Continuous speech

Speech has no natural pauses between the word boundaries, the pauses mainly appear on a syntactic level, such as after a phrase or a sentence. It introduces a difficult problems for speech recognition — how should we translate a waveform into a sequence of words? After a first stage of recognition of phones and phone categories, we have to group them into words. Even if we disregard word boundary ambiguity, this is still a difficult problem. The only way to simplify this process is to give clear pauses between the words. This works for short command-like communication, but as the possible length of utterances increases, clear pauses get inefficient[2].

iii. Speed of speech

We speak in different speed, at different times. For example, if we are stressed, we tend to speak faster, and if we are tired, the speed tends to decrease. We also speak in different speeds if we talk about something known or unknown[2].

iv. Complexity of the spoken language

In English, few words have different meanings which depends on the context – for example, “red” and “read” sounds the same but have different meanings[7].

v. Background noise

Identifying the speech from the background noise is very difficult. This is specifically true if the background noise is also speech (say at a party) [7].

vi. Speaker variation:

Human speak distinctively. The voices are not only different between speakers but also wide variation within one specific speaker. Some voice variations are given below -

vi.i. Speaking style

All speakers speak distinctively due to their unique personality. They have a different ways to pronounce words. The way of speaking differs in different situations. We don't speak the same way in public area, with our teachers or friends. Humans also express emotions while speaking i.e happy, sad, fear, surprise or anger[2].

vi.ii Realization

If same words were pronounced again and again, the result of speech signal won't be same. There will be small differences in the acoustic wave produced. Speech realization changes over time[2].

vi.iii Speaker sex

Male and Female have different tone. Female have short vocal tract than male. That's why, generally they say, the pitch of female voice is roughly twice than male[2].

5. SYSTEM ARCHITECTURE

It is one of the significant part of speech recognition systems it detects and analyze the given sample data. It improves the speech signal to enable to get cataloging of sounds. The raw speech data obtains a lot of facts besides the linguistics information. Sometime these facts are very bulk in quantity and variable in nature. It becomes very difficult to categories and analyze that raw data. The feature extraction very important to extract silent features with nominal variability from this raw data That is important for classification of sounds, it gives the direction to verify the jest of information. Many feature extraction techniques are present but voice gives better result[7].

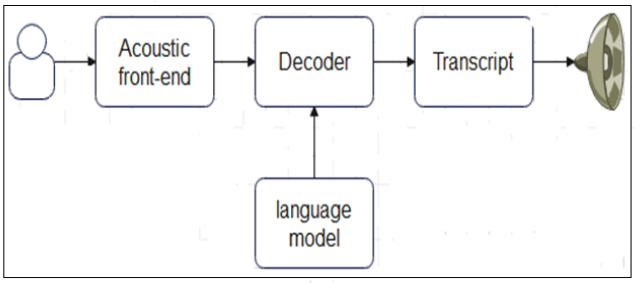


Fig 5: System architecture

At the initial state, the user inputs the audio through mic.

Then, the system recognizes the audio format.

If the system doesn’t recognizes or any error, it’ll take you to the final state

After audio recognition, language specification is given to which specific language it should be converted.

Converting and translating process happens using the language specification and gives the output.

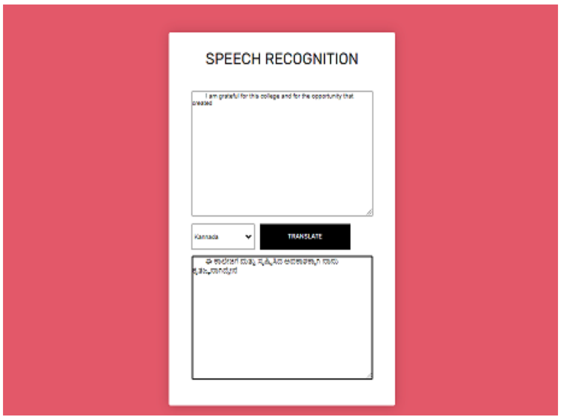
The converted text will be called transcript here.

At the final state, the user can view the translated transcript and also have the feature to listen to the transcript through the player.

6. RESULTS

Speech Recognition is the procedure and the related innovation for changing over the discourse signal into its comparing grouping of words or other semantic substances.

As a result of speech recognition in this project, we can able to say the sentence which is needed to be translated to the specified language and display the converted text as the output. Also we can listen to the converted text for the proper pronunciation in that language specified.


SPEECH RECOGNITION

I am grateful for this college and for the opportunity that created

Kannada

[illegible]

Fig 6:

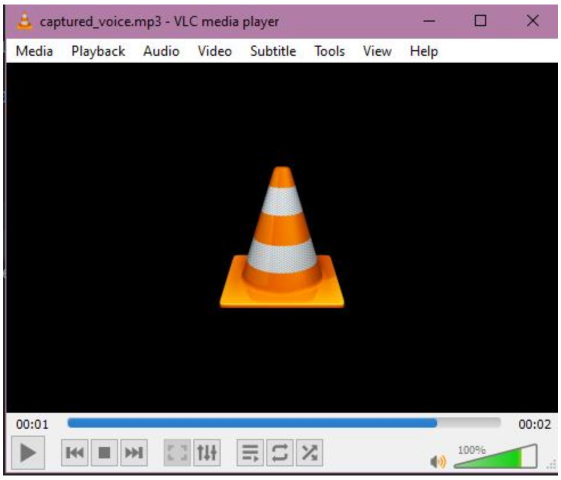


Fig 7

7. CONCLUSIONS

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[1].

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).

[7] 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.

[8] 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.