



SPEECH RECOGNITION USING PYTHON

¹ Manthan Mohile, ² Aishwarya Patil, ³ Tanvi Pawar , ⁴ Meena Perla
^{1,2,3,4}(EXTC, VIVA INSTITUTE OF TECHNOLOGY/ MUMBAI UNIVERSITY , INDIA)

Abstract : Speech recognition technology is one from the fast growing engineering technologies. It has a number of applications in different areas and provides potential benefits. Nearly 20% people of the world are suffering from various disabilities; many of them are blind or unable to use their hands effectively. The speech recognition systems in those particular cases provide a significant help to them, so that they can share information with people by operating computer through voice input. This project is designed and developed keeping that factor into mind, and a little effort is made to achieve this aim. Our project is capable to recognize the speech and convert the input audio into text; it also enables a user to perform operations such as “save, open, exit” a file by providing voice input. It also helps the user to open different system software such as opening Ms-paint, notepad and calculator. At the initial level effort is made to provide help for basic operations as discussed above, but the software can further be updated and enhanced in order to cover more operations.

Keywords - Python , speech , speech processing

I. INTRODUCTION

Designing a machine that converse with human, particularly responding properly to spoken language intrigued engineers and scientists for centuries. Speech Recognition System(SRS) is also known as Automatic Speech Recognition (ASR) or computer speech recognition which is the process of converting a speech signal to a sequence of words by means of an algorithm implemented as a computer program. It has the potential of being an important mode of interaction between humans and computers . Today speech technology enabled applications are commercially available for a limited but interesting range of tasks. Very useful and valuable services are provided by these technology enabled machines, by responding correctly and reliably to human voices. In order to bring us closer to the “Holy Grail” of machines that recognize and understand fluently spoken speech, many important scientific and Technological advances have been took place, but still we are far from having a machine that mimics human behavior. Speech recognition technology has become a topic of great interest to general population, through many block buster movies of 1960's and 1970's. The anthropomorphism of "HAL", a famous character in Stanley Kubrick's movie “2001: A Space Odyssey”, made the general public aware of the potential of intelligent machines. In this movie, an intelligent computer named “HAL” spoke in a natural sounding voice and was able to recognize and understand fluently spoken speech, and respond accordingly. George Lucas, in the famous Star Wars saga, extended the abilities of intelligent machines by making them intelligent and mobile Droids like R2D2 and C3PO were able to speak naturally, recognize and understand fluent speech, move around and interact with their environment, with other droids, and with the human population. Apple Computers in the year of 1988, created a vision of speech technology and computers for the year 2011, titled

“Knowledge Navigator”, Navigator”, which defined the concepts of a Speech User Interface (SUI) and a Multimodal User Interface (MUI) along with the theme of intelligent voice- enabled agents. This video had a dramatic effect in the technical community and focused technology efforts, especially in the area of visual talking agents. Languages, on which so far automatic speech recognition systems have been developed, are just a

fraction of the total around 7300 languages. Chinese, English, Russian, Portuguese, Vietnamese, Japan, Spanish, Filipino, Arabic, Bengali, Tamil, Malayalam, Sinhala and Hindi are prominent among them.

A. Types of Speech Recognition. In Speech Recognition System the ability to recognize the speech signal can be subdivided into different classes,

a. **Isolated Words:** In this type, system accepts single utterance at a time. And usually requires each utterance to have quiet on both side of sample window and require a speaker to wait between words. Its response will be better for single word but give poor result for multiple words input.

b. **Connected Words:** In this type, multiple words given to the system which runs separately as isolated words and having small duration of time between them.

c. **Continuous Speech:** In this type, natural speech is spoken by the user that is detectable by the machine. Continuous speech recognition is difficult to create because they utilize special method for implementation.

d. **Spontaneous Speech:** In this type natural and spontaneous word has the ability to handle a variety of natural features such as words run together including mispronunciations, non-word and false statements, which are difficult to read.

B. Types of Speaker Model Every speaker has unique properties which affects the voice. On the basis of these properties system is divided into two main classes.

a. **Speaker Dependent Model:** Speaker dependent model depends on specific speaker. These models are easier to implement and less expensive. It gives more accurate result for specific speaker and less accurate result for other speakers.

b. **Speaker Independent Model:** Speaker independent models depend upon many speakers. These models are difficult to implement and more expensive. It gives more accurate result for many speakers and less accurate result for specific speaker .

II. DESIGN METHODOLOGY

3.1 Methodology

As an emerging technology, not all developers are familiar with speech recognition technology. While the basic functions of both speech synthesis and speech recognition takes only few minutes to understand (after all, most people learn to speak and listen by age two), there are subtle and powerful capabilities provided by computerized speech that developers will want to understand and utilize. Despite very substantial investment in speech technology research over the last 40 years, speech synthesis and speech recognition technologies still have significant limitations. Most importantly, speech technology does not always meet the high expectations of users familiar with natural human-to-human speech communication. Understanding the limitations - as well as the strengths - is important for effective use of speech input and output in a user interface and for understanding some of the advanced features of the Java Speech API. An understanding of the capabilities and limitations of speech technology is also important for developers in making decisions about whether a particular application will benefit from the use of speech input and output.

3.2 Flow charts

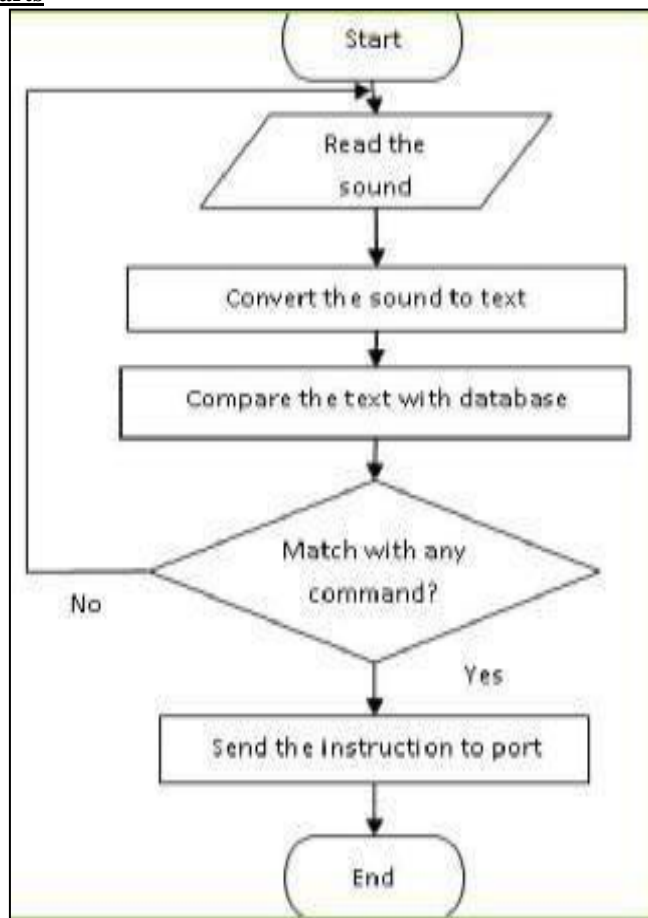


Fig 1.1

The flow chart shows that when person using the interface will speak and sound input will be given

The system will read the sound then the sound will be converted into text format.

This text format compared with the database if command does not matched with the database it will again call the read sound command. And if the command is matched then instruction will be send to port and the process will end with the text on screen

Speech recognition helps us to save time by speaking instead of typing it also gives us the power to communicate with our devices without even writing one line code .

This makes technological device more accessible and easier to use to decode the Speech into text ,the groups of vectors matched to one or more phenomena of fundamental unit of Speech.

This will be improve typing accuracy,speed and this will also make it easy to use system. For this improvisation and overcoming the difficulties we are introducing this system

Recognition Flowchart presentation

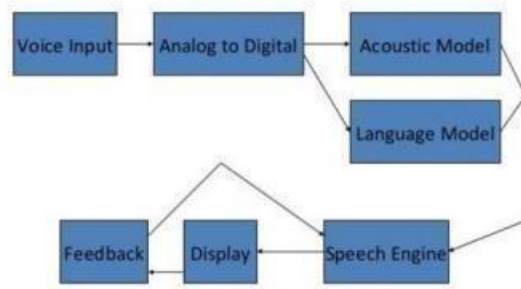


Fig.1.2

III. RESULT

```
Terminal: Local x + v
Try the new cross-platform PowerShell https://aka.ms/pscore6

PS C:\Users\mital\PycharmProjects\pythonProject> PIP INSTALL SPEECHRECOGNITION
ERROR: unknown command "INSTALL" - maybe you meant "install"
PS C:\Users\mital\PycharmProjects\pythonProject> pip install speechrecognition
Collecting speechrecognition
  Using cached SpeechRecognition-3.8.1-py2.py3-none-any.whl (32.8 MB)
Installing collected packages: speechrecognition
Successfully installed speechrecognition-3.8.1
```

```
Run: main x
C:\Users\mital\PycharmProjects\pythonProject1\venv\Scripts\python.exe C:/Users/mital/PycharmProjects/pythonProject1/main.py
Speak:
You said hello my name is Manthan and this is my final year project speech recognition

Process finished with exit code 0
```

Fig.2.1



Fig.2.2

Fig 2.1 Shows the initial stage of output where we have installed various libraries and installed different packages for the various process and requirements the execution is done on the PyCharm platform which is the main software used in our project. Here, we can even see the output on PyCharm with the help of a microphone.

Fig 2.2 Represents the Second stage of the project in this stage we have made additional changes to make it better as compared to fig 2.1 here we can see the GUI representation for the output and we have even added the different types of languages spoken greatly for all kind of people in to be able to get output languages included other than English are Hindi and Marathi.

Whereas Fig 2.3 and 2.4 are moreover, the application which is at its embryonic stage consists of speech to text and text to speech. However, here we have executed only a single language but it is an android app so it will be accessible for more smartphone users just like many other apps available.

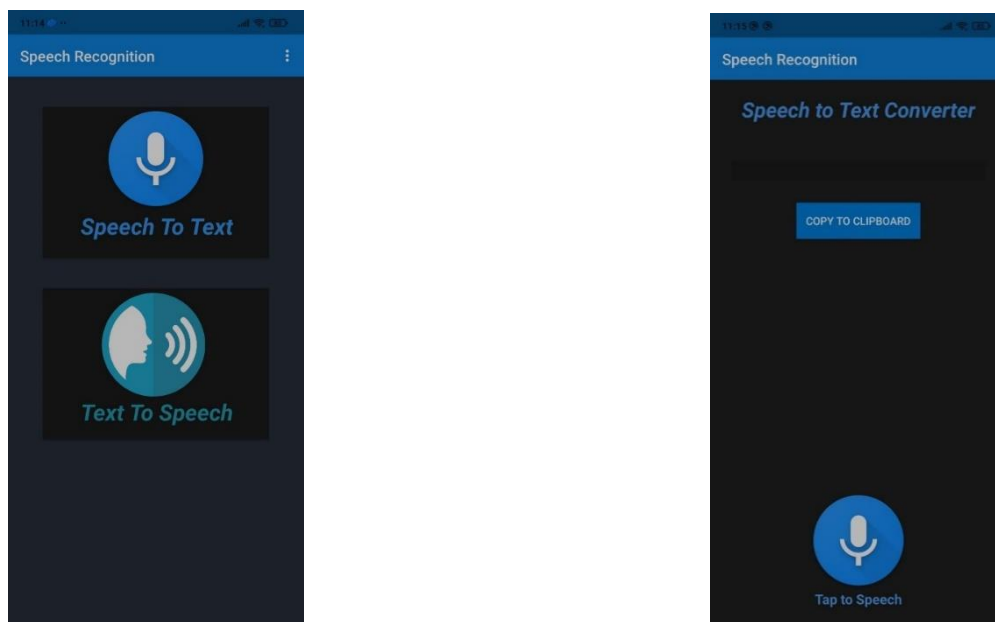


Fig.2.3

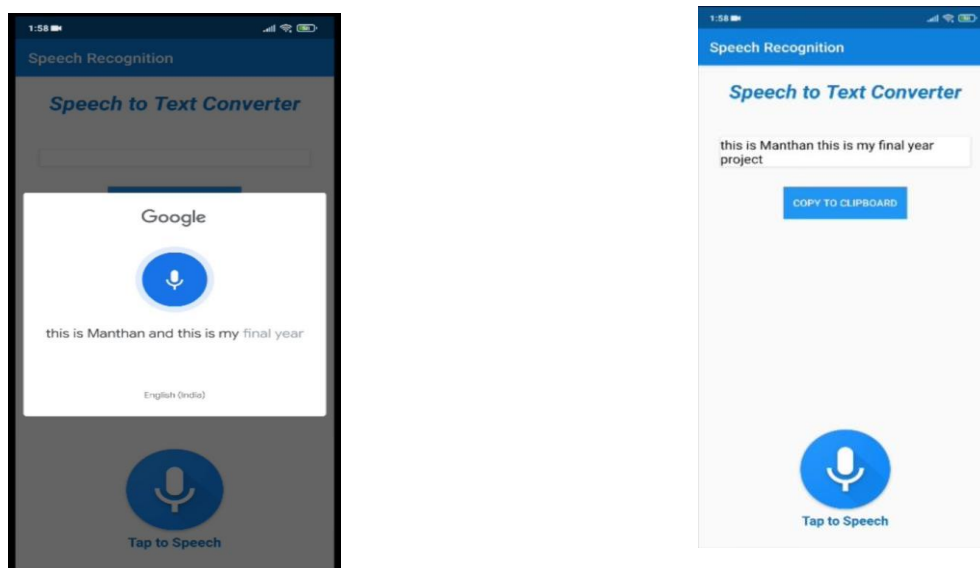


Fig.2.4

IV. CONCLUSION

This Thesis/Project work of speech recognition started with a brief introduction of the technology and its applications in different sectors. The project part of the Report was based on software development for speech recognition. At the later stage we discussed different tools for bringing that idea into practical work. After the 50 development of the software finally it was tested and results were discussed, few deficiencies factors were brought in front. After the testing work, advantages of the software were described and suggestions for further enhancement and improvement were discussed.

Acknowledgements

We shall be failing in our duty, if we will not express our sincere gratitude to all those distinguished personalities with the help of whom we have successfully completed our project our project guide Prof Meena Perla. My deep gratitude to Dr. Arun Kumar, Principal, Viva Institute Of Technology, who always been playing a great role in all round development of the student. My deep gratitude to Prof. Archana Ingle, The Head Of Electronics and Telecommunication Department for her valuable guidance, advice and constant aspiration to our work, teaching and non-teaching staff for their kind support, help and assistance, which they extended as and when required. Last but not the least we wish to thank my friends for providing technical and moral support. We hope that this project report would meet the high standards of all concerned people and for their continuous co-operation during the whole period of period of project that helped us in enhancement of this project.

REFERENCES

- [1] "Speech recognition- The next revolution" 5th edition.
- [2] Ksenia Shalnova, "Automatic Speech Recognition" 07 DEC 2007
Source: http://www.cs.bris.ac.uk/Teaching/Resources/COMS12303/lectures/Ksenia_ShalnovaSpeech_Recognition.pdf
- [3] C. Gopala Krishnan¹, Y. Harold Robinson², Naveen Chilamkurti³ Machine Learning Techniques for Speech Recognition using the Magnitude Journal of Multimedia Information System, 2020.
- [4] Dhanush Kumar S, Lavanya S, Journal of Speech to Text Conversion, IJARIT, 2018.
- [5] Nenny Anggraini, Angga Kurniawan, Luh Kesuma, Wardhani, Nashrul Hakiem, Speech Recognition Application for the Speech Impaired using the Android-based Google Cloud Speech API, TELKOMNIKA, 2018
- [6] Aditya Amberkar, Gaurav Deshmukh, Speech Recognition using Recurrent Neural Networks, IEEE, 2018.
- [7] Tanel ALUMAE¹ and Ottokar TILK, Automatic Speech Recognition System for Lithuanian Broadcast Audio, IOS PRESS, 2016.
- [8] Akhilesh Halageri, Amrita Bidappa, Arjun C, Madan Mukund Sarathy, Shabana Sultana, Speech Recognition using Deep Learning IJCSIT, 2015.
- [9] Tanel ALUMAE., Full-duplex Speech-to-text System for Estonian, IOS PRESS, 2014.
- [10] George E. Dahl, Dong Yu, Senior Member, Context Dependent Pre-Trained Deep Neural Networks for Large-Vocabulary Speech Recognition, IEEE, 2012.
- [11] Dhanush Kumar S, Lavanya S, Journal of Speech to Text Conversion, IJARIT, 2018
- [12] Aditya Amberkar, Gaurav Deshmukh, Speech Recognition using Recurrent Neural Networks, IEEE, 2018.