Journal of Engineering Sciences
Vol 12, Issue 7, July/2021
ISSN NO:0377-9254

Voice Assistant in accessingReal World Applications

Dr.Ch.Rathna Jyothi[1], P.Beracah[2], M. Gowtham [3]
Department of Computer Science and Engineering, Andhra Loyola Institute of Engineering and Technology,Jawaharlal Nehru Technological University - Kakinada chrjyothi@aliet.ac.in
panam.beracah@gmail.commeesalagauti@gmail.com

***Abstract*— Voice Assistant in accessing real world applications aims to develop a personal-assistant for real world applications. Jarvis(Voice assistant) draws its inspiration from virtual assistants like Cortana ,Siri ,Alexa e.t.c. It has been designed to provide a user-friendly interface for carrying out a variety of tasks by taking certain well-defined commands. Users can interact with the assistant through voice commands. since, voice based tasks are more fast and efficient than the manual doing of tasks.**

As a personal assistant, Jarvis using the technologies like ASR(Automatic Speech Recognition),TTS(Text to Speech) assists the users with many activities that humans do in daily life like general human conversation, searching queries in web, searching for videos, downloading images, meanings and opposites, Sending Emails and Sending WhatsApp Messages, Writing documents and opens applications present in the system.

Keywords*— *Personal Assistant, Windows OS, Automation , Scrapping, Text-to-Speech, Speech Recognition.

1. INTRODUCTION

Speech Recognition is a voice based technology in which it takes voice as input and converts the speech into the text by recognition. There can be both online and offline recognition when compared to offline recognition online recognition is fast for processing and giving exact response. This technology is also called as ASR(Automatic Speech Recognition),where it automatically takes the users voice and converts to text.

Speech Recognition has different types. In which they are developed in to different worlds based on sounds and other punctuality marked sentences. The words are divided into isolated, connected and spontaneous words which make the words formation and the words pronunciation neat and clean in the regular way.

2. METHODOLOGY

I. Speech Recognition:

Speech Recognition can have different methodologies forimplementing, we can see different methodologies in the following sub sections:

i. MFCC(Mel Frequency Cepstral Coefficient):

The work of MFCC is to take the speech signal and it converts it into feature vector i.e. the input for MFCC algorithm is Speech Signal and the output for MFCC Algorithm is Feature Vector. The working of it as follows by taking speech signal and converts by Fast Fourier Transform and as spectrum as medium it scale using Mel Scale Filtering and it gives as Mel Frequency Spectrum to logarithmic Function and the obtained result from logarithmic function is sent to the Discrete Cosine Transform where the speech signals are transformed to cosine waves and it uses as cepstral coefficients as medium to convert it into Derivatives of Feature Vector and there feature vector is formed. The entire working of MFCC is shown in thefollowing diagram

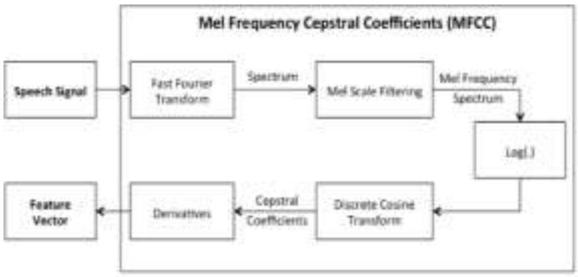

Fig 1. MFCC Architecture

The above mentioned algorithm is used in this developed project and from the vectors we take the largest equivalent similar voice recognition and convert it into speech according to the voice given to the algorithm as input to it from any respective devices as speech.

II. Selenium Web Driver:

Selenium suite was majorly introduced in software for testing purposes. But still we can make selenium web driver for scrapping work and make better applications work as fast they can by just few lines of code.

In this project we can see how selenium we driver will workand what are its uses, but now mostly selenium suite comprises of 4 categories

i. Selenium IDE
ii. Selenium RC
iii. Selenium Web Driver
iv. Selenium Grid

In this study we see about only selenium web driver and it is having most for all browsers like google chrome and Mozilla too.

The following diagram shows the selenium web driver can access through all the web browsers searching for their values in html format and takes input and downloads e.t.c.

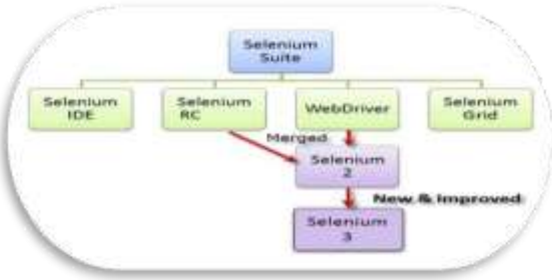

Fig 2.Selenium Suite

In this project we can see what can selenium do by just basicprogramming lines.

i. Download bulk images:

Used to train models like in machine learning

ii. Send WhatsApp Messages:

we can send WhatsApp messages also and when sendingthis WhatsApp messages we can sending 3 ways

→send message by delayed minutes
→send single contact
→send multiple contacts
iii. Search in Wikipedia and read from it

3. RESULTS & DISCUSSION

The result as shown in below graphs and each graph represents how the voice has been divided into spectrum and the voice has been divided into parallel

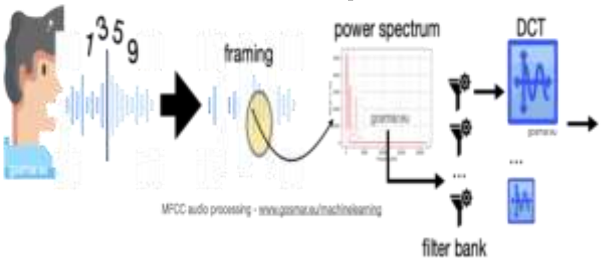

Fig 3. MFCC Working

The given speech is first framed and then based on the power used to pronounce it differentiates itself and converts into graph and then to the vector. Between this process it develops the main things like background noise reduction, filtering the data and then it sends to logarithmic functions based on the given speech by user.
MFCC does:
→Blocking of the voice given as speech

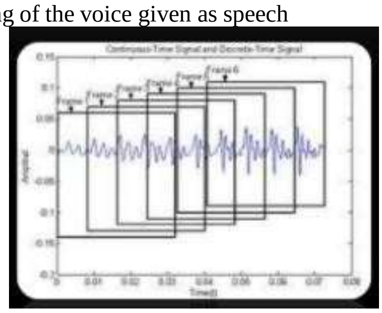

Fig 4. MFCC Speech Blocking

→Blocked voice is measured as scale called "Mel" Scale

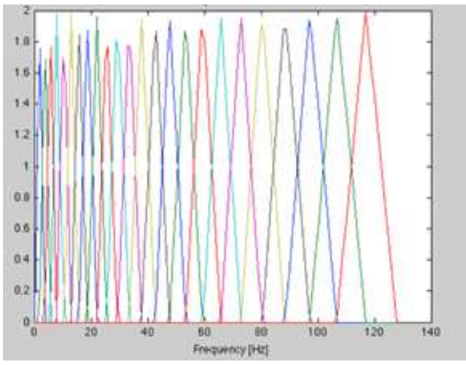

Fig 5. MFCC Wrapping

→Mel scale gets Frequency Wrapping and the wrapping gets spectrum by using the cosine and logarithmic formula

$$\tilde{c}_n = \sum_{k=1}^{K} (\log \tilde{S}_k) \cos\left[n\left(k - \frac{1}{2}\right)\frac{\pi}{K} \right], \qquad n = 0,1,\ldots,K\text{-}1$$

Fig 6. Cosine Formula

The above shown results are only partial and the actual results and way much bigger and the formulas are more complex and it can be surfaced through the references given in below sections of this page.

4. SYSTEM DESIGN AND ARCHITECTURE

Every project depends on how system is designed and how it performs on its execution to maintain this project you need to have the system with features having inbuilt or external microphone which makes to used as input i.e. voice .

The adjacent diagram shows how the system architecture is well made
→Firstly command modules i.e python script takes as input as voice
→Then based on the voice we can perform tasks if the given voicecommand is present it will execute.
→If voice command is not present then it will give text to speech as feedback saying command not present or the given matter.

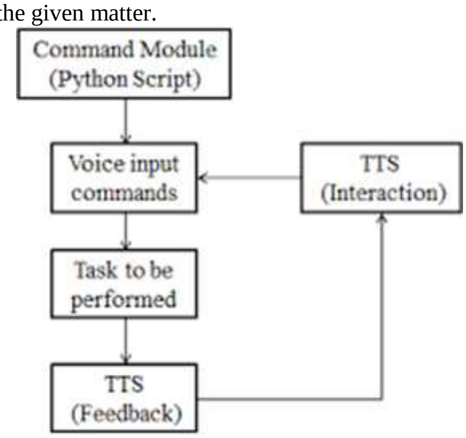

Fig 7. Voice Assistant Working Process

The following diagram shows the working of our project on voicebased input and the python modules and python backend also

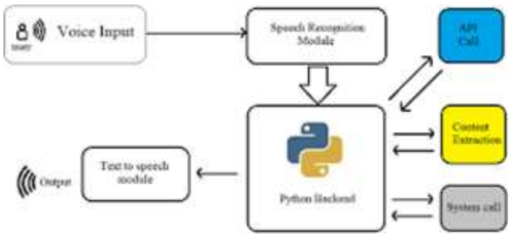

Fig 8. Python Voice Assistant Working Process

5. REQUIREMENTS

I. Software Requirements
→ Python
→ OS →Windows 10,Windows 8,7
→Required Ai and python pip packages

II. Hardware Requirements
→ Desktop/Laptop or Computer
→Internal or External Microphone

6. CODE

This project code is placed in GitHub and the GitHub

link is givenbelow.
https://codeload.github.com/BERACAH333/voice-assistant/zip/refs/heads/main
OR
visit https://github.com/BERACAH333/voice-assistant

7. CONCLUSION

This Project Voice Authorized assistant in accessing real world applications mainly aims is to develop an Assistant that will be used to identify and answer to the queries and commands given by the user.The developed system is made on python programming language to be more specific Python3. Different libraries where used such as Speech Recognition, Text to Speech convertor, Short Mail Transferring Protocols (SMTP). It provides information regarding the weather, News, it can play music, it can search for topics on Wikipedia, Display the current date and time. User can collect information through this application. It reduces both man power and time.

No need ask queries in very strict and specific way. The user should aware of general rules of English Language. The goal is to provide people a quick and easy way to have their questions answered. It can make the human daily activities go in a simpler, faster and less time consumption way. It is very Useful and efficient upon making in doing the given tasks.

8. ACKNOWLEDGMENT

This research project was partially supported by the Department of Computer Science and Engineering, Andhra Loyola Institute of Engineering and Technology, Jawaharlal Nehru Technological University-Kakinada. We are grateful to Associate Professor Dr.Ch.Rathna Jyothi forleading us to develop and contribute a paper to the conference.

9. REFERENCES

1.Hidden Markov models for speech recognition; X.D. Huang, Y.
2. Ariki, M.A. Jack. Edinburgh: Edinburgh University Press, c1990
3.Speech Recognition: The Complete Practical Reference Guide; T.Schalk, P. J. Foster: Telecom Library Inc, New York; ISBN O- 9366648-39-2; 377 pages; paperback only. Covers speech recognition in a telephony environment and wish to use call processing hardware based in PCs. It is written using Dialogic hardware as the example for the hardware.
4. Automatic speech recognition: the development of the SPHINX system; by Kai-Fu Lee; Boston; London: Kluwer Academic, c1989
5.An Introduction to the Application of the Theory of Probabilistic Functions of a Markov Process to Automatic Speech Recognition,
6. S. E. Levinson, L. R. Rabiner and M. M. Sondhi; in Bell Syst.
7. Tech. Jnl. v62(4), pp1035--1074, April 1983
8.Review of Neural Networks for Speech Recognition, R. P. Lippmann; in Neural Computation, v1(1), pp 1-38, 1989.
9.Automatic Speech and Speaker Recognition: Advanced Topics,
10. C.H. Lee, F.K. Soong and K.K. Paliwal (Eds.), Kluwer, Boston, 1996. Rabiner Lawrence, Juang Bing-Hwang. Fundamentals of Speech Recognition Prentice Hall , New Jersey, 1993, ISBN 0-13-015157-2
11. Deller John R., Jr., Hansen John J.L., Proakis John G. ,Discrete- Time Processing of Speech Signals, IEEE Press, ISBN 0-7803- 5386-2
12. Hayes H. Monson,Statistical Digital Signal Processing and Modeling, John Wiley & Sons Inc. , Toronto, 1996, ISBN 0-471-59431-8
13. Proakis John G., Manolakis Dimitris G.,Digital Signal Processing, principles, algorithms, and applications, Third Edition, Prentice Hall , New Jersey, 1996, ISBN 0-13- 394338-9
14. Hiroaki Sakoe and Seibi Chiba, Dynamic Programming algorithm Optimization for spoken word Recognition, IEEE transaction on Acoustic speech and Signal Processing, Fecruary 1978.