Speech Recognition And Various Actions Using Speech Commands

Sqn Ldr Nishan Tiwari

Roll No 204102316

ABSTRACT

Speech recognition, also known as automatic speech recognition (ASR), computer speech recognition, or speech-to-text, is a capability which enables a program to process human speech into a written format. While it's commonly confused with voice recognition, speech recognition focuses on the translation of speech from a verbal format to a text one whereas voice recognition just seeks to identify an individual user's voice.

Here in this project, a program has been written to perform very basic commands on the host as spoken by the user. This program can be further elaborated for making a computerized Function Generator where just by saying the function needed it can be plotted. The proof of concept is shown here.

2 Introduction

In the 1980s the computers were operated using command line interface or CLI. In the mid-1990s, with Microsoft Windows operating systems, the world got familiar with Graphical User Interface or GUI. In today's generation it is required that the major routine works be performed using voice user interface.

Voice user interfaces (VUIs) allow the user to interact with a system through voice or speech commands. Virtual assistants, such as Siri, Google Assistant, and Alexa, are examples of VUIs. The primary advantage of a VUI is that it allows for a hands-free, eyes-free way in which users can interact with a product while focusing their attention elsewhere.

However, in Digital Signal Processing lab, a lot of signals need to plotted in a routine manner for which same code is again and again written. This can be solved just by using a VUI which asks the user which signal he/she wants to plot and then does the same. Here only the proof of concept is given. A full fledged program for the same can be made in which complex functions such as DFT, Weiner Filtering or Linear Prediction Filtering can be done just by speaking. This can be used by TAs in IIT to demonstrate various filtering effects to students.

Speech recognition is an interdisciplinary subfield of computer science and computational linguistics that develops methodologies and technologies that enable the recognition and translation of spoken language into text by computers. It is also known as automatic

speech recognition (ASR), computer speech recognition or speech to text (STT). It incorporates knowledge and research in the computer science, linguistics and computer engineering fields.

Some speech recognition systems require "training" (also called "enrollment") where an individual speaker reads text or isolated vocabulary into the system. The system analyzes the person's specific voice and uses it to fine-tune the recognition of that person's speech, resulting in increased accuracy. Systems that do not use training are called "speaker independent" systems. Systems that use training are called "speaker dependent".

Speech recognition applications include voice user interfaces such as voice dialing (e.g. "call home"), call routing (e.g. "I would like to make a collect call"), demotic appliance control, search key words (e.g. find a podcast where particular words were spoken), simple data entry (e.g., entering a credit card number), preparation of structured documents (e.g. a radiology report), determining speaker characteristics, speech-to-text processing (e.g., word processors or emails), and aircraft (usually termed direct voice input).

3 Methodology

Here we are using speech_recognition module of python to identify the spoken words and then the text is converted to respective commands. The flow of program is as shown

    (a)   First, the program asks whether user wants to convert a live speech to text or a recorded speech.

```
Please press
1 for recognition of speech through microphone.
2 for recognition of speech through file.
┌──────────────────────────────────────────────────────┐
│                                                        │
└──────────────────────────────────────────────────────┘
```

    (b)   Then if user selects 1 then program asks the user
"Hello, please tell what you want to do? Open a website or know time or plot a sinusoidal signal. Say close the session to close. Please speak only when i say."

```
Please press
1 for recognition of speech through microphone.
2 for recognition of speech through file.
1
Hello, please tell what you want to do? Open a website or know time or plot a sinusoidal signal. Say close the session to clos
e. Please speak only when i say.
Please Speak
Did you say open a website
Which website you want me to open? Like google or news or Facebook or Twitter or Youtube or Corona cases?
```

(c)   Now we have 3 options i.e.

   (i)   Open a website for that user has to say "*open a website*".

      (ii)   Know time for that user has to say "*know time*".

      (iii)   Plot a sinusoidal signal. For this the user has to say "*plot a sinusoidal signal*".

      (iv)   To close the session, user has to say "*close the session*".

(d)   If the user says *"open a website",* then the program asks "*Which website you want me to open? Like google or news or Facebook or Twitter or Youtube or Corona Cases".* Now user can say any one of these and that particular website will be opened.

(e)   If user says covid then website like https://www.covid19india.org is opened as shown
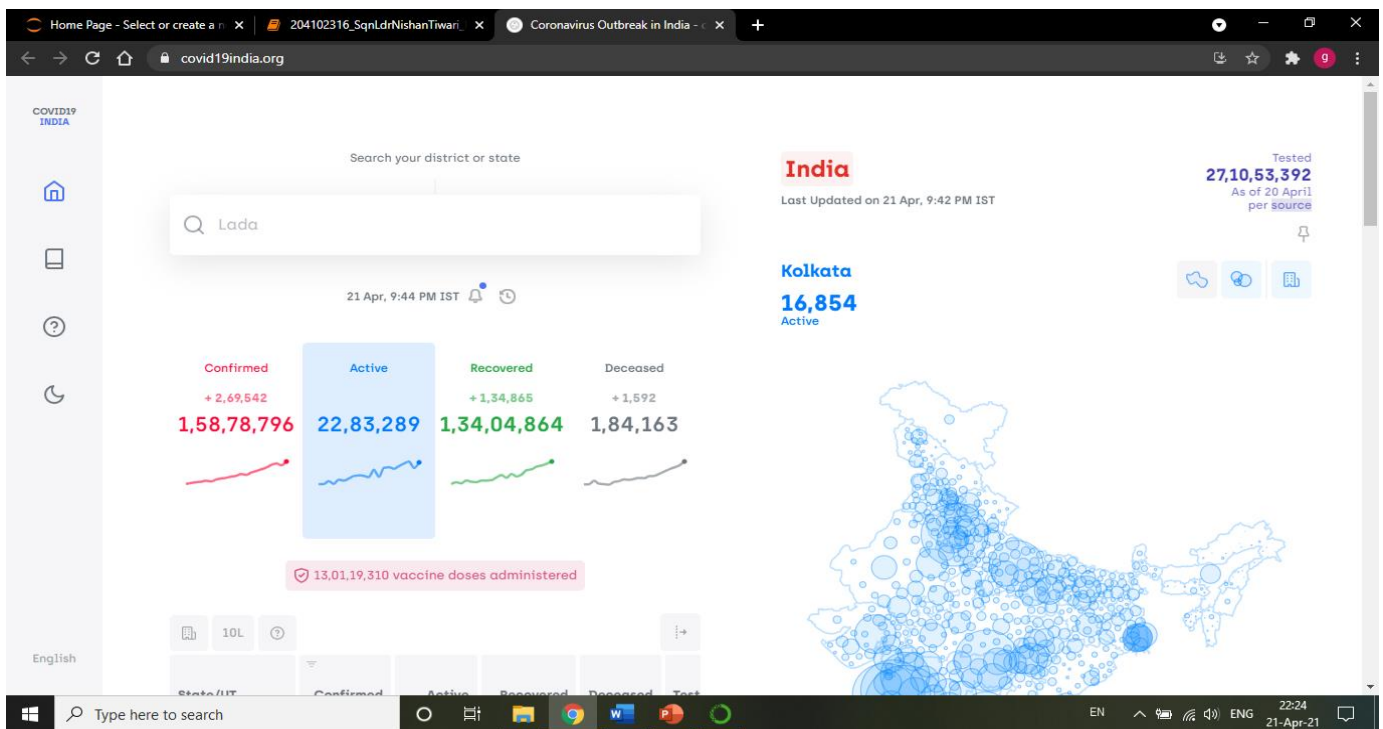
(f)   Similarly, if the user says "*plot a sinusoidal signal*", then a prompt appears along with the voiceover that "*please enter the frequency, amplitude and phase in the box*". On entering 100 as frequency, 2 as Amplitude and 180 as phase, the graph is plotted as shown.

```
Please press
1 for recognition of speech through microphone.
2 for recognition of speech through file.
1
Hello, please tell what you want to do? Open a website or know time or plot a sinusoidal signal. Say close the session to clos
e. Please speak only when i say.
Please Speak
Did you say plot a sinusoidal signal
Please enter the frequency, amplitude and phase in the box
Enter the frequency
100
Enter Amplitude
2
Enter Phase in Degree180
Thankyou for using the software created by Squadron Leader Nishan Tiwari. Good Bye
```


Sine wave x(t)=2.0sin(2πft+phi) where f=100Hz and phi=180
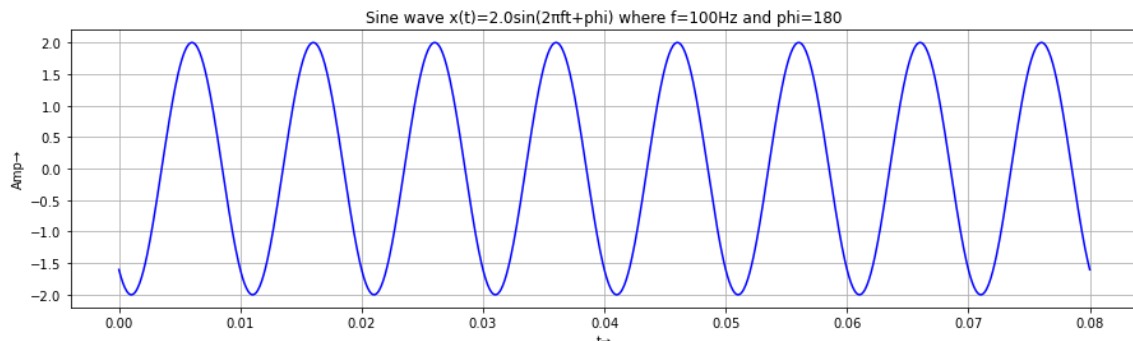
(g)   Finally if the "*know date*" option is chosen then the date and time are shown.

(h)   Now, finally if "*close the session*" is told then the program terminates as shown.

```
Please press
1 for recognition of speech through microphone.
2 for recognition of speech through file.
1
Hello, please tell what you want to do? Open a website or know time or plot a sinusoidal signal. Say close the session to clos
e. Please speak only when i say.
Please Speak
Did you say no time
Hello, please tell what you want to do? Open a website or know time or plot a sinusoidal signal. Say close the session to clos
e. Please speak only when i say.
Please Speak
Did you say close the session
Thankyou for using the software created by Squadron Leader Nishan Tiwari. Good Bye
```

(i)   However, if in the beginning the option 2 was chosen then the program would ask to enter the file name which contains the speech. Finally, it outputs the speech converted to text and reads out the same as shown.

```
Please press
1 for recognition of speech through microphone.
2 for recognition of speech through file.
2
Enter file name
male.wav
The Text from audio file are:
summary the sides to break a teacher for the you keep adequate coverage the works of places to save money baby is taking longer
to getting squared away then the bank was expected during the life events company in AVN heartattack se retirement income the B
ritish were inadequate news of the saving lives are heard it has done that you naked Bond what a discussion can insert when the
title of this type of song is in question or waxing or gasing needed I prevent my be personalized number work lace leather and
lace work on a flat surface and smooths out this post and a separate system uses a single sirf contained Unity op shop at store
holds a good mechanical isliye bad bus figures with Gauhar in late summer curable chairs cabinets chest down house is a set
```

The block diagram of the methodology used to convert speech to text is as shown below in the general model of automatic speech recognition.

General model of ASR
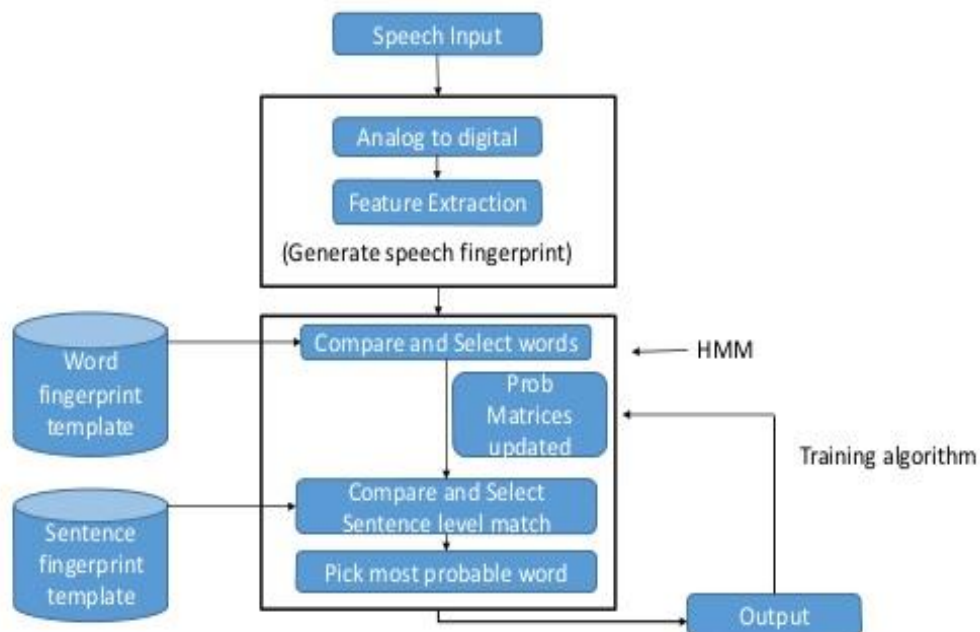
5 Conclusion

In this project, a speech recognition technique is utilized to demonstrate how useful it can be in the form of Voice User Interface. We have demonstrated how a computer can be made to operate based on the set of instructions given in the form of speech commands. The same technology is used in many applications like Alexa, Google Assistant, Siri etc. These openly available APIs for converting speech to text can be invariably used in many applications like:

    (a)   In the workplace for

        (i)   Search for reports or documents on your computer

        (ii)   Create a graph or tables using data

        (iii)   Dictate the information you want to be incorporated into a document

        (iv)   Print documents on request

        (v)   Record minutes

        (vi)   Make travel arrangements

    (b)   In banking for
        (i)   Request information regarding your balance, transactions, and spending habits without having to open your cell phone

        (ii)   Make payments

        (iii)   Receive information about your transaction history

    (c)   In healthcare for

        (i)   Quickly finding information from medical records

        (ii)   Nurses can be reminded of processes or given specific instructions

        (iii)   Nurses can ask for administrative information, such as the number of patients on a floor and the number of available units

        (iv)   At home, parents can ask for common symptoms of diseases, when they should go to the doctor, and how to look after a sick child

6 References

1. https://en.wikipedia.org/wiki/Speech_recognition

2. https://searchcustomerexperience.techtarget.com/definition/speech-recognition

3. https://realpython.com/python-speech-recognition/

4. https://www.slideshare.net/sonukumar142/speech-recognition-techniques