

**SUMMER INTERNSHIP PROJECT**

School of Engineering and Technology

Department of Computer science and Engineering

The Northcap University, Gurugram

06th June - 19th July 2019

Prepared by:

HRITIK KOUNSAL (17CSU083)

**Acknowledgement**

I would like to express our sincere gratitude to our supervisor **Monika Sehgal**, for giving us the opportunity to work on this topic. It would never be possible for us to take this project to this level without their innovative ideas and there relentless support and encouragement project.

HRITIK (17CSU083)

**Declaration**

I hereby declare that this project report entitled “**ALPHA CHARLIE – INTELLIGENT ASSISTANT USING PYTHON”** *by* **HRITIK (17CSU083),** being submitted in partial fulfillment of the requirements for the degree of Bachelor of Technology in **Computer Science and Engineering** NorthCap University, Gurugram during the academic year June, 2019, is a bonafide record of our original work carried out under guidance and supervision of **Mrs. Monika Sehgal (PYTHON DEVELOPER)** and has not been presented elsewhere.

HRITIK (17CSU083)

**ABSTRACT**

The report provides information on the topics I have learned during my summer internship. I purchased an offline course on **PYTHON** from APTECH. I learned how to develop working models that are useful on industrial level and how to make a voice assistant like **ALEXA**. How and which methods are available in the modules to make a voice assistant. Understanding how python concepts work is one of the most valuable and useful things you can do.

**Python** is an [interpreted](https://en.wikipedia.org/wiki/Interpreted_language), [high-level](https://en.wikipedia.org/wiki/High-level_programming_language), [general-purpose](https://en.wikipedia.org/wiki/General-purpose_programming_language) programming language. Created by [Guido van Rossum](https://en.wikipedia.org/wiki/Guido_van_Rossum) and first released in 1991, Python's design philosophy emphasizes [code readability](https://en.wikipedia.org/wiki/Code_readability) with its notable use of significant whitespace. Its language constructs and object-oriented approach aim to help programmers write clear, logical code for small and large-scale projects.

Python is [dynamically typed](https://en.wikipedia.org/wiki/Dynamic_programming_language) and [garbage-collected](https://en.wikipedia.org/wiki/Garbage_collection_(computer_science)). It supports multiple [programming paradigms](https://en.wikipedia.org/wiki/Programming_paradigm), including procedural, object-oriented, and [functional programming](https://en.wikipedia.org/wiki/Functional_programming). Python is often described as a "batteries included" language due to its comprehensive [standard library](https://en.wikipedia.org/wiki/Standard_library).

**INTRODUCTION**

A voice assistant is a digital assistant that uses voice recognition, [natural language processing](https://searchbusinessanalytics.techtarget.com/definition/natural-language-processing-NLP) and speech synthesis to provide aid to users through phones and voice recognition applications.

Voice assistants are used in help and service phone lines, smartphones and other places to assist users with tasks, including:

* Listening to an audiobook.
* Requesting information, similarly to inputting a search query.
* Making reservations.
* Adding items to a shopping list.
* Performing mathematical calculations.
* Playing music

Voice assistants are built on artificial intelligence ([AI](https://searchenterpriseai.techtarget.com/definition/AI-Artificial-Intelligence)), machine learning and voice recognition technology. As the end user interacts with the digital assistant, the AI programming uses sophisticated algorithms to learn from data input and better itself at predicting the user's needs. Some assistants are built with more advanced cognitive computing technologies which will allow a digital assistant to understand and carry out multi-step requests with numerous interactions and perform more complex tasks, such as booking seats at a movie theater.

Digital assistants can be contrasted with another application of consumer-facing AI called smart advisors. Smart advisor programs are knowledge-oriented, while digital assistants are task-oriented, although some perform both roles. Popular voice assistants currently include Apple's [Siri](https://searchmobilecomputing.techtarget.com/definition/Siri), Amazon's [Alexa](https://whatis.techtarget.com/definition/Alexa-Voice-Services-AVS), Google [Now](https://whatis.techtarget.com/definition/Google-Now), Google Assistant and Microsoft's [Crotana](https://searchenterprisedesktop.techtarget.com/definition/Cortana).

**TABLE OF CONTENTS**

|  |  |  |
| --- | --- | --- |
| S.NO | TITLE | PAGE.NO |
| 1 | ABSTRACT | 4 |
| 2 | INTRODUCTION | 5 |
| 3 | TRAINING DESCRIPTION | 8-30 |
| 4 | ANALYSIS | 31 |
| 5 | CONCLUSION | 32 |

**TABLE OF FIGURES**

|  |  |  |
| --- | --- | --- |
| FIG.NO | TITLE | PAGE NO. |
| 1 | Speech Recognition Process | 11 |
| 2 | Speech Synthesis | 17 |
| 3 | CASE DIAGRAM | 20 |
| 4 | ACITIVITY DIAGRAM | 21 |
| 5 | SEQUENCE DIAGRAM | 22 |
| 6 | COLLABORATION DIAGRAM | 23 |
| 7 | Interfaces | 25-26 |
| 8 | PROJECT OUTPUT | 30 |

**1. PROJECT OVERVIEW**

This PROJECT report considers an overview of how we can give command to the computer with the help of voice recognition system. The first section deals with the description of voice recognition process, its applications, its flaws and its future scope. Later part of the report covers voice recognition process , code to run your computer through voice recognition, finally at the end it concludes with the different use of the system and applications and further improvements and the considerations.

**1.1 PROJECT OBJECTIVE**

1.1.1 To understand the speech recognition and its fundamentals

.1.1.2 It’s working and applications in different areas

1.1.3 Its implementation as a desktop Application

1.1.4 Development for software that can mainly be used for:

• Speech Recognition

• Speech Generation

• Text Editing

• Tool for operating Machine through voice.

**1.2 ABSTRACT**

Voice 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 1 “open” a file by providing voice input. It also helps the user to open different system software such as play music, email. 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.

We use PyCharm software to run this project which is built in python language, in PyCharm first we have to import different modules that are needed for running various operations in the software and without importing this modules our project cannot run successfully. PyCharm provides a great platform for this functions to work properly.

1.3 PROJECT SCOPE

This project has the speech recognizing and Voice recognition technique used to create an automated computer system and we can create automated machines more and more which can be used in many fields such as medical , software industries which are fully dependent on computers to do their work on daily purpose. Many software and IT companies are trying to being fully automated and are creating robots which are fully automated and coded to do perform various works in all the companies. So it’s very important to create virtualized and automated computer system.

**2. LITERATURE REVIEW**

**2.1 An overview of voice recognition**

Voice recognition is a technology that able a computer to take the command from human with a help of microphone. These words are recognized by speech recognizer, and in the end, system works on the given command. The process of speech recognition consists of different steps that will be discussed in the following sections one by one. An ideal situation in the process of speech recognition is that, a voice recognition engine recognizes all words uttered by a human but, practically the performance of a speech recognition engine depends on number of factors. Vocabularies, multiple users and noisy environment are the major factors that are counted in as the depending factors for a voice recognition engine.

**2.2 HISTORY**

You command Siri to search for nearest pizza delivery joints and you are presented with a list of them. As can be deduced, voice recognition has come a long way; however, this technology is not something recent. In fact, it has its roots back in the 1950s. Let’s delve into the past and take a look at the brief history of how voice recognition technology has evolved over time into the [speech recognition software](http://www.totalvoicetech.com/voice-recognition-software/) we know today

.

In the history of speech recognition software technology, this was the era of ‘baby talk’; only numbers and digits could be comprehended. In 1952, ‘Audrey’ was invented by Bell Laboratories which could only understand numbers. But in 1962, the ‘shoebox’ technology was able to understand 16 words in English. Later, voice recognition was enhanced to comprehend 9 consonants and 4 vowels.

The U.S. Department of Defense contributed heavily towards the development speech recognition systems and from 1971 to 1976, it funded the DARPA SUR (Speech Understanding Research) program. As a result, ‘Harpy’ was developed by Carnegie Mellon which had the ability to comprehend 1011 words. It employed a more efficient system of searching for logical sentences.

**2.3 Types of voice recognition**

2.3.1Speaker dependent system - The voice recognition requires training before it can be used,

Which requires you to read a series of words and phrases.

2.3.2 Speaker independent system - The voice recognition software recognizes most users' voices

with no training.

2.3.3 Discrete speech recognition - The user must pause between each word so that the speech

Recognition can identify each separate word.

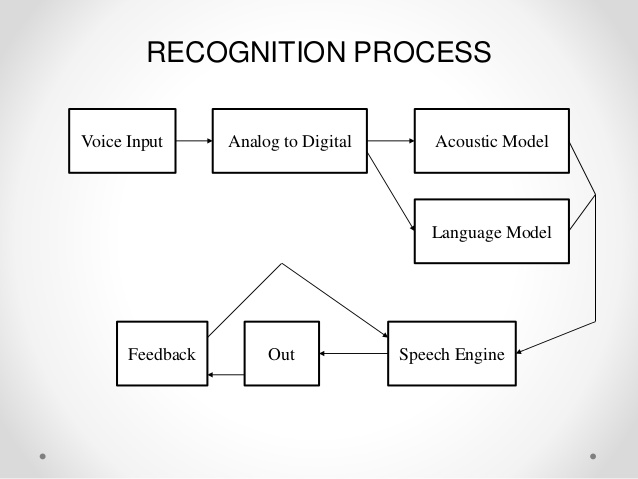
2.3.4Continuous speech recognition - The voice recognition can understand a normal rate of

Speaking.

2.3.5Natural language - The speech recognition not only can understand the voice, but can also

return answers to questions or other queries that are being asked.

**2.4 Speech recognition process**



* 1. **Types of modules:-**

1. **Pyttsx3 module**: It is a cross – platform text to speech library which is platform independent. The major advantage of using this library for text to speech conversion is that it works offline.
2. **Web browser module**: The Web browser module provides a basic interface to the system’s standard web browser. It provides an open function which takes a filename or a URL, and displays it on the browser.
3. **Smtplib module**: The Smtplib modules defines an smtp client session object that can be used to send mail to any internet machine with an SMTP and ESMTP listener daemon.
4. **Geocoder module**: The Geocoder modules is used transform a physical address description to a location on the earth’s surface. It gives the address of the client at that time in terms of longitude and latitude
5. **Random module**: Python offers random module that can generate random numbers these are pseudo random numbers as the sequence of number generated depends on the seed
6. **Speech recognition module**: Speech recognition is an important feature in several Application used as home automation, artificial intelligence, etc. this article aims to provide an introduction on how to make use of speech recognition library in python.
7. **Wikipedia module:** Wikipedia is a python library that makes it easy to access and parse data from Wikipedia. Search Wikipedia, get article summaries, and get data like links and images from a page and more.
8. **Date time module**: Date, time and date time classes provides a number of function to deal with dates, times and time intervals. Date and date time are an object in python, so when you manipulate them you are actually manipulating objects and not string or timestamps.
9. **Play music module**: It is a module used to add sound and music to your program. It contains file such play audio, pygame which are used to add sons, audio, and create games in python with the help of play music module.
10. **OS module:** This module provides a portable way of using operating system dependent functionality. If you just want to read or write a file see [open ()](https://docs.python.org/2/library/functions.html#open), if you want to manipulate paths, see the [os.path](https://docs.python.org/2/library/os.path.html#module-os.path) module.
11. **PyOWM module:** PyOWM is a client Python wrapper library for Open Weather Map web APIs. It allows quick and easy consumption of OWM data from Python applications via a simple object model and in a human-friendly fashion.

**2.6 Uses of voice recognition system**

Basically voice recognition is used for two main purposes. First dictating that is in the context of voice recognition is translation of spoken words into text, and second controlling the computer, that is to develop such system that probably would be capable enough to take commands from a user to operate different application by voice.

Speech recognition is an alternative of keyboard. If you are unable to write or just don’t want to type then programs of speech recognition helps you to do almost any thing that you used to do with keyboard. Writing by voice let a person to write 150 words per minute or more if indeed he/she can speak that much quickly.

**2.7 Applications of voice recognition system**

2.7.1 From medical perspective: People with disabilities can benefit from speech

recognition programs. Speech recognition is especially

useful for people who have difficulty using their hands, in

such cases speech recognition programs are much

beneficial and they can use for operating computers.

Speech recognition is used in deaf telephony, such as

2.7.2 From military perspective: Speech recognition programs are important from

military perspective; in Air Force speech recognition has

definite potential for reducing pilot workload. Beside the

Air force such Programs can also be trained to be used in

helicopters, battle management and other applications.

2.7.3 From educational perspective: Individuals with learning disabilities who have

problems with thought-to-paper communication

(essentially they think of an idea but it is processed

incorrectly causing it to end up differently on paper)

can benefit from the software. Some other application

areas of speech recognition technology are described

as under.

**2.8 Voice Recognition weakness and flaws**

Besides all these advantages and benefits, yet a hundred percent perfect voice recognition system is unable to be developed. There are number of factors that can reduce the accuracy and performance of a speech recognition program.

Voice recognition process is easy for a human but it is a difficult task for a machine, comparing with a human mind speech recognition programs seems less intelligent, this is due to that fact that a human mind is God gifted thing and the capability of thinking, understanding and reacting is natural, while for a computer program it is a complicated task.

First it need to understand the spoken words with respect to their meanings, and it has to create a sufficient balance between the words, noise and spaces. A human has a built in capability of filtering the noise from a speech while a machine requires training, computer requires help for separating the speech sound from the other.

**2.9 Few factors that are considerable in this regard are :**

2.9.1 Homonyms: Are the words that are differently spelled and have the different meaning but

acquires the same meaning, for example “there” “their” “be” and “bee”. This

is a challenge for computer machine to distinguish between such types of

phrases that sound alike.

2.9.2 Overlapping speeches: A second challenge in the process, is to understand the speech

uttered by different users, current systems have a difficulty to

separate simultaneous speeches form multiple users

.

2.9.3 Noise factor: the program requires hearing the words uttered by a human distinctly and

clearly. Any extra sound can create interference, first you need to place

system away from noisy environments and then speak clearly else the

machine will confuse and will mix up the words.

**2.10 The future of speech recognition.**

2.10.1 Accuracy will become better and better.

2.10.2 Dictation speech recognition will gradually become accepted.

2.10.3 Greater use will be made of “intelligent systems” which will attempt to guess what the speaker intended to say, rather than what was actually said, as people often misspeak and make unintentional mistakes.

2.10.4 Microphone and sound systems will be designed to adapt more quickly to changing background noise levels, different environments, with better recognition of extraneous material to be discarded.

**3. METHODOLOGY AND TOOLS**

**3.1 Fundamentals to voice recognition**

3.1.1 Utterances**:** When user says some things, then this is called utterance in other words

speaking a word or a combination of words that are meaningful to the

computer is called an utterance. Utterances are then sent to speech engine to

be processed.

3.1.2 Pronunciation: A speech recognition engine uses a process word is its pronunciation, that

represents what the speech engine thinks a word should sounds like. Words

can have the multiple pronunciations associated with them.

3.1.3 Grammar : Grammar uses particular set of rules in order to define the words and phrases

that are going to be recognized by speech engine, more concisely grammar

define the domain with which the speech engine works . Grammar can be

simple as list of words or flexible enough to support the various degrees of

variations.

3.1.4 Accuracy: The performance of the speech recognition system is measurable . the ability of

recognizer can be measured by calculating its accuracy. It is useful to identify

an utterance.

3.1.5 Vocabularies: Vocabularies are the list of words that can be recognized by the speech

recognition engine . Generally the smaller vocabularies are easier to

identify by a speech recognition engine, while a large listing of words are

difficult task to be identified by engine.

**3.2 Tools**

1. Python idle 3.6

2. Visual c++ 14.0

3. PyCharm 3.6

4. Google Chrome

5. Command prompt

6. Microsoft word (documentation)

**3.3 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 python Speech API.

**3.3.1** **Speech Synthesis**

**Structure Analysis**

**Text to Phoneme Conversion**

**Prosody Analysis**

**Wave Form Production**

**3.3**.**2 A** speech synthesizer converts written text into spoken language. Speech synthesis is also referred to as text-to-speech (TTS) conversion.

The major steps in producing speech from text are as follows:

**1. Structure analysis:** Process the input text to determine where paragraphs, sentences and other

structures start and end. For most languages, punctuation and formatting

data are used in this stage.

**2.** **Text pre-processing:** Analyze the input text for special constructs of the language. In English,

special treatment is required for abbreviations, acronyms, dates, times,

numbers, currency amounts, email addresses and many other forms.

Other languages need special processing for these forms and most

languages have other specialized requirements.

**3**. Text-to-phoneme conversion: Convert each word to phonemes. A phoneme is a basic unit of

sound in a language. English has around 45 phonemes

including the consonant and vowel sounds. For example, "times"

is spoken as four phonemes "t ay m s". Different languages have

different sets of sounds (different phonemes). For example,

Japanese has fewer phonemes including sounds not found in

English, such as "ts" in "tsunami".

**4. Prosody analysis**: process the sentence structure, words and phonemes to determine

appsropriate prosody for the sentence. Prosody includes many of the features

of speech other than the sounds of the words being spoken. This includes

the pitch (or methodology),the timing (or rhythm), the pausing, the

speaking rate, the emphasis on words and many other features. Correct

prosody is important for making speech sound right and for correctly

conveying the meaning of a sentence.

**5. Waveform production**: finally, the phonemes and prosody information are used to produce

the audio waveform for each sentence. There are many ways in which

the speech can be produced from the phoneme and prosody

information. Most current systems do it in one of two ways:

concatenation of chunks of recorded human speech, or formant

synthesis using signal processing techniques based on knowledge of

how phonemes sound and how prosody affects those phonemes. The

details of waveform generation are not typically important to

application developers.

**3.3.3 Selecting Voices**

Most speech synthesizers are able to produce a number of voices. In most cases voices attempt to sound natural and human, but some voices may be deliberately mechanical or robotic.

The Voice class is used to encapsulate the four features that describe each voice: voice name, gender, age and speaking style.

The voice name and speaking style are both String objects and the contents of those strings are determined by the synthesizer. Typical voice names might be "Victor", "Monica", "Ahmed", "Jose", "My Robot" or something completely different. Speaking styles might include "casual", "business", "robotic" or "happy" (or similar words in other languages) but the API does not impose any restrictions upon the speaking styles.

For both voice name and speaking style, synthesizers are encouraged to use strings that are meaningful to users so that they can make sensible judgments when selecting voices. By contrast the gender and age are both defined by the API so that programmatic selection is possible

**3.4 DIAGRAMS**

3.4.1 Use case diagram

**USER**

**3.4.2 ACTIVITY DIAGRAM**

Open PyCharm software

Enter the code and run

**E**

Take voice input

Take text input

Save document

**3. SEQUENCE DIAGRAM**

Location, weather

Open Gmail , YouTube

Play music

Open Google

Time, date

**1.** web browser

2. Play audio

**3.** Time date

4. Geocoder

**4. COLLABORATION DIAGRAM**

Hello sir, I am your personal assistant Alpha Charlie, how I can help you.

I am sorry I didn’t get that command please try writing the command.

Add recipient, what should I say. Email sent, I am searching internet for you.

**4. IMPLEMENTATION AND TESTING**

**4.1 System requirements**

**4.1.1 Minimum requirements**

Pentium 200 MHz processor

64 MB of RAM

Microphone

Sound card

**4.1.2 Best requirements**

1.6 GHz Processor

128 MB or more of RAM

Sound cards with very clear signals

High quality microphones

**4.2 Hardware Requirements**

**1. Sound cards**

Speech requires relatively low bandwidth, high quality 16 bit sound card will be better enough to work. Sound must be enabled, and proper driver should be installed. Sound cards with the 'cleanest' A/D (analog to digital) conversions are recommended, but most often the clarity of the digital sample is more dependent on the microphone quality and even more dependent on the environmental noise. Some speech recognition systems might require specific sound cards.

**2. Microphones**

A quality microphone is key when utilizing the speech recognition system. Desktop microphones are not suitable to continue with speech recognition system, because they have tendency to pick up more ambient noise. The best choice, and most common is the headset style. It allows the ambient noise to be minimized, while allowing you to have the microphone at the tip of your tongue all the time. Headsets are available without earphones and with earphones (mono or stereo).

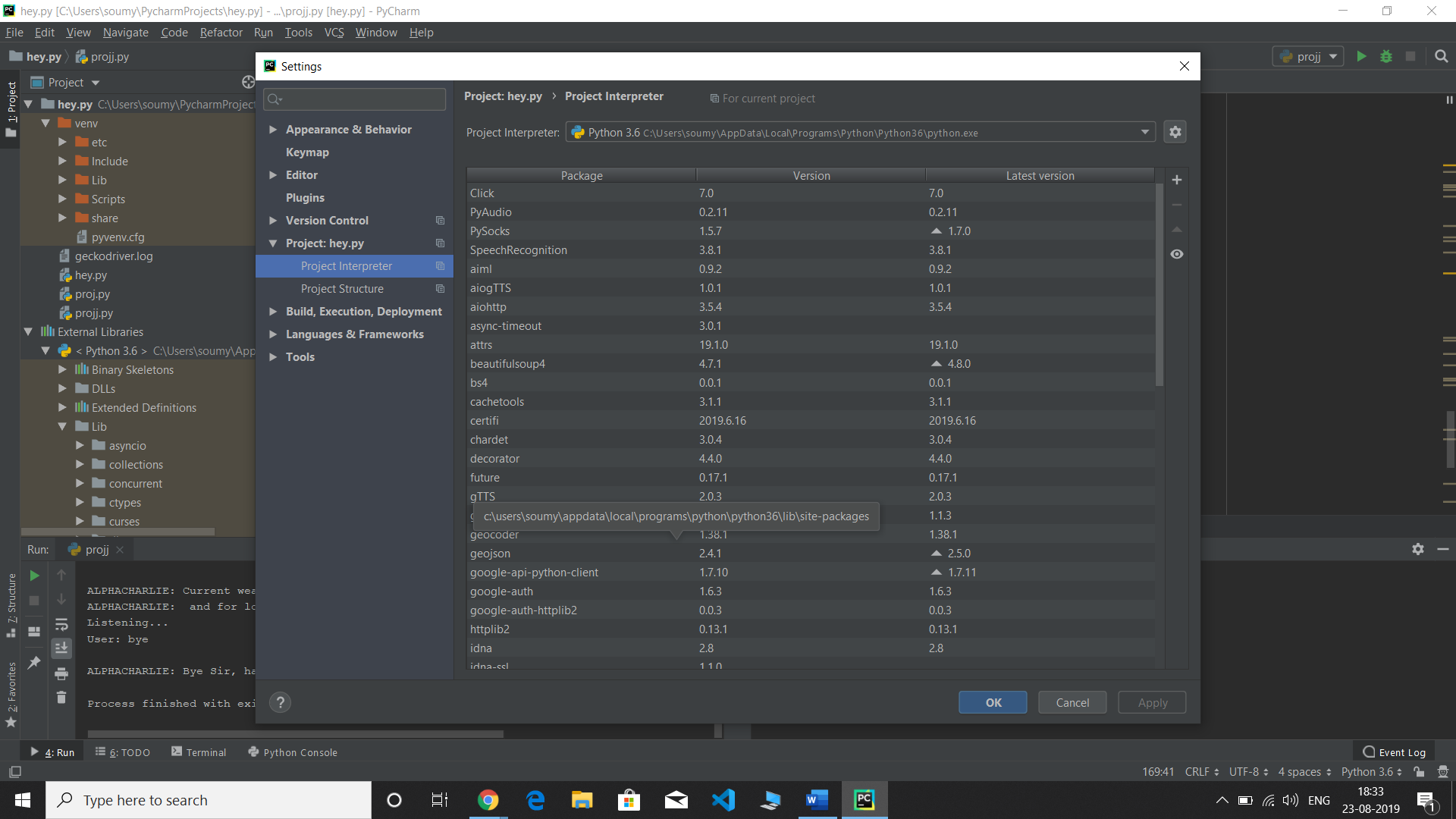
**3. Computer/ Processors**

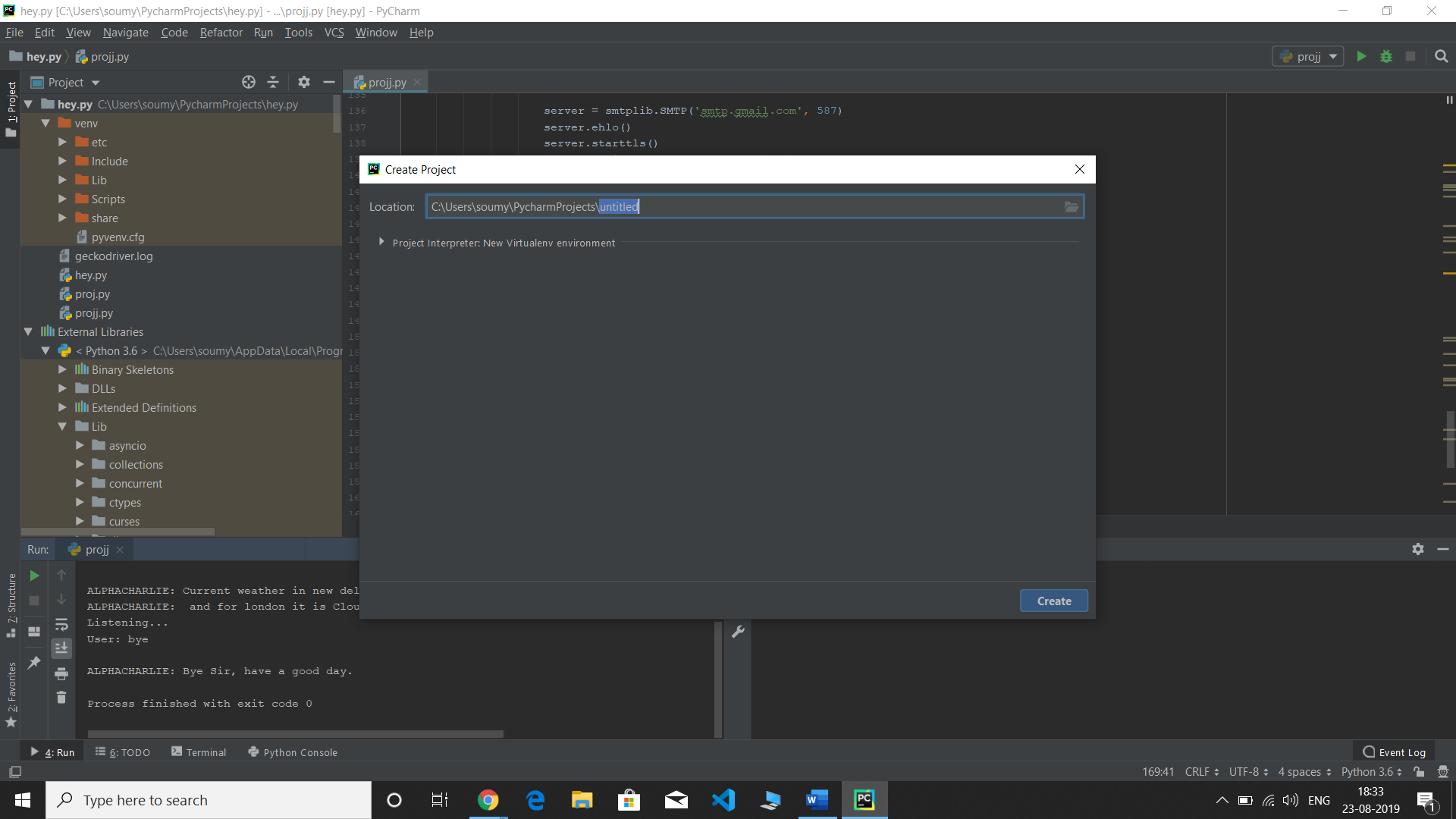
Speech recognition applications can be heavily dependent on processing speed. This is because a large amount of digital filtering and signal processing can take place in ASR.

4.3 Interfaces

Opening Software

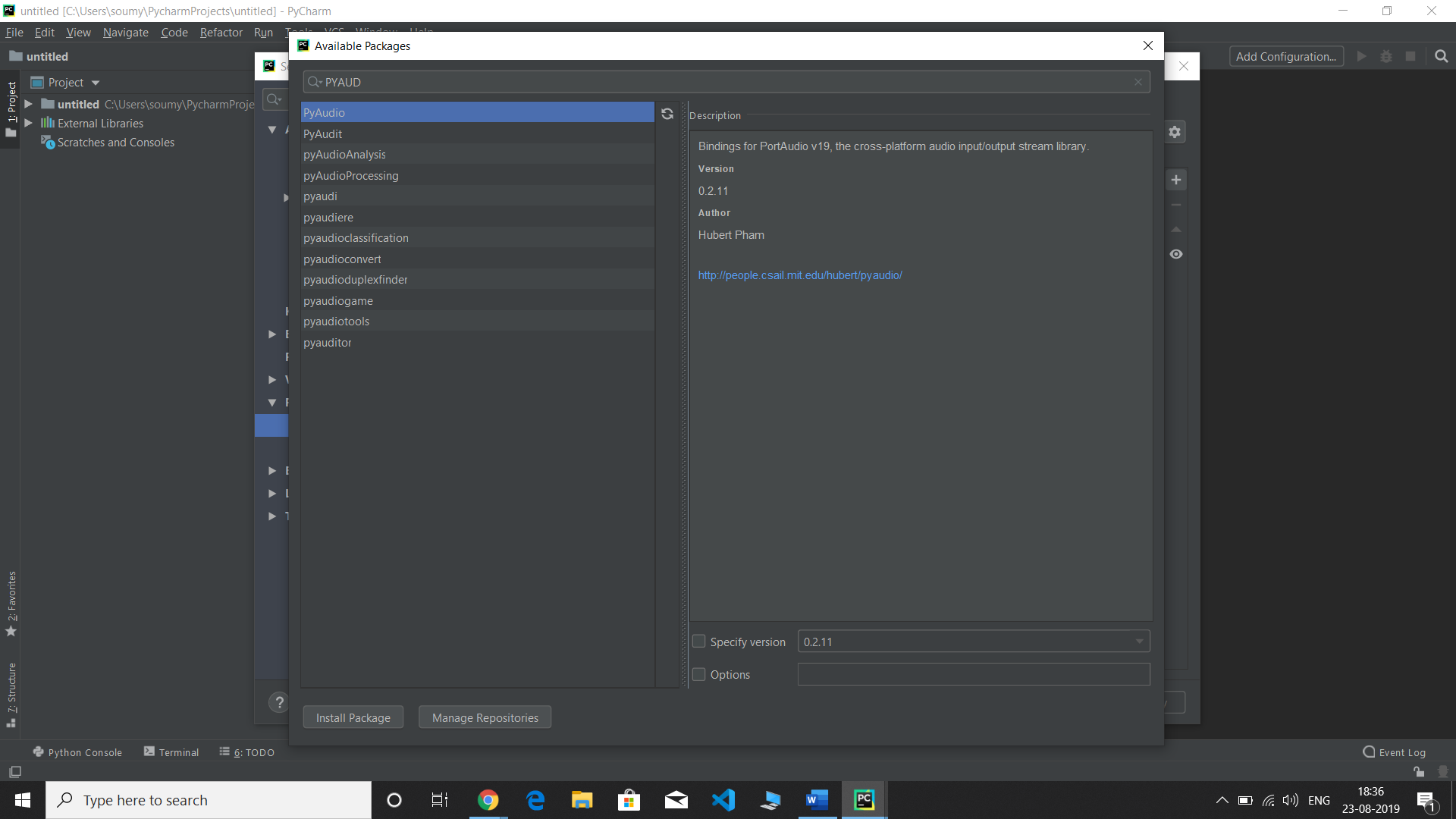
1. Software is opened by using PyCharm IDE, after selecting python 3.6 interpreter.



2. Now open new project than select new python file

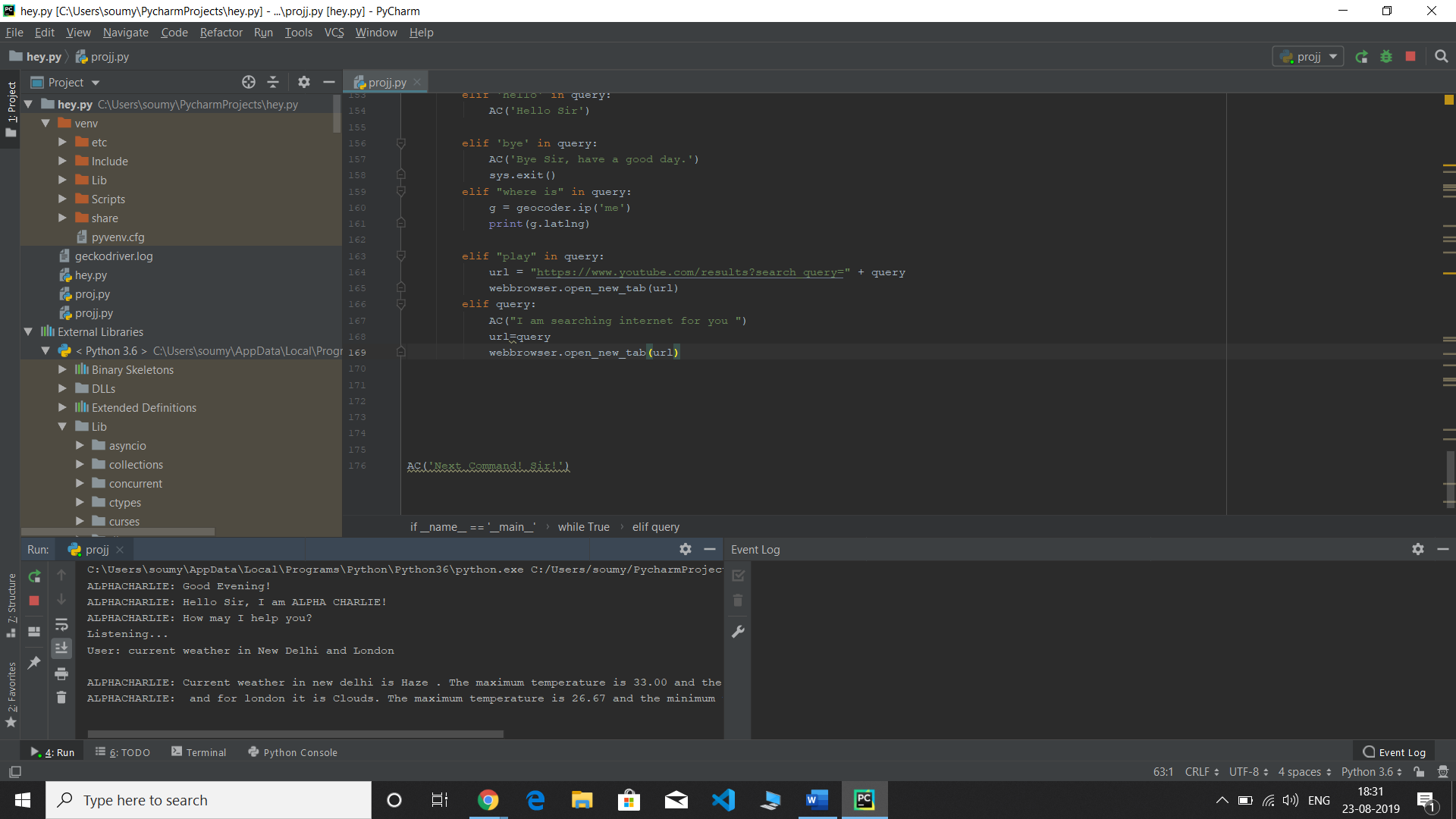
3. Now import all the required modules in the code

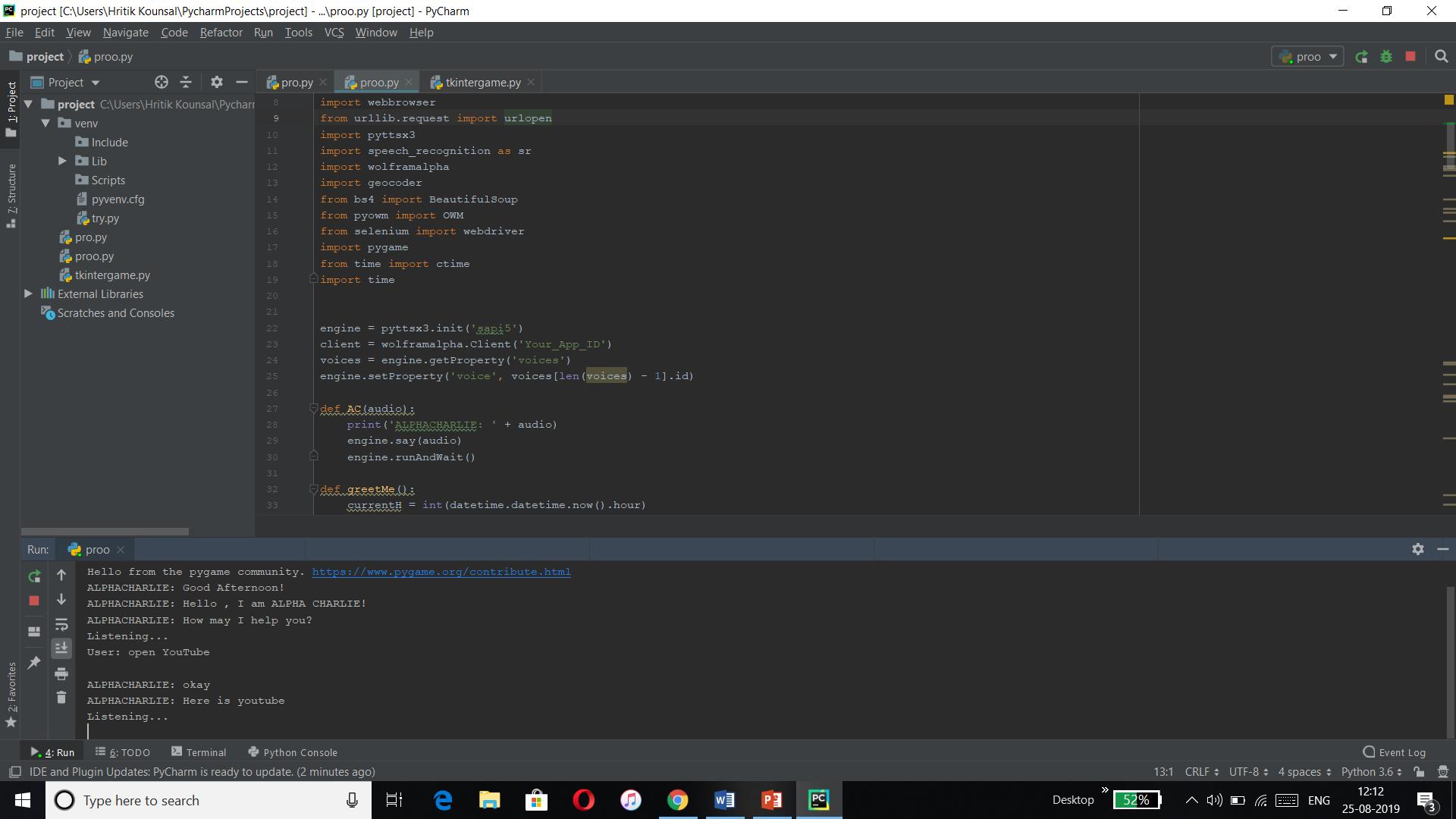
E.g. Import PyAudio



**Project output**

**This project will give this output after running is**



****

**ANALYSIS**

**Advantages of software**

1. Able to write the text through both keyboard and voice input.

2. Voice recognition of different notepad commands such as open save and clear.

3. Open different windows soft wares, based on voice input.

4. Requires less consumption of time in writing text.

5. Provide significant help for the people with disabilities.

6. Lower operational costs.

7. Artificially intelligent

8. It is secure

**Disadvantages**

1. Low accuracy

2. Not good in the noisy environment

3. Expensive equipment.

4. It is very costly.

**Future Enhancements**

This work can be taken into more detail and more work can be done on the project in order to bring

Modifications and additional features. The current software doesn’t support a large vocabulary, the

Work will be done in order to accumulate more number of samples and increase the efficiency of the software. The current version of the software supports only few areas of the notepad but more areas can be covered and effort will be made in this regard.

**Conclusion**

This 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 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.

**REFERENCES**

1. Www. GeeksforGeeks.Com

2. www.Tutorial Point.com

3. http://en.wikipedia.org/wiki/Speech\_recognition Visited: 12/July/2019

4. http://www.jisc.ac.uk/media/documents/techwatch/ruchi.pdf

|  |  |  |
| --- | --- | --- |
| Project Daily Task | | |
| Date | **Day** | **Work** |
| 3rd June,2019 | Monday | Learned the basic python concepts. |
| 4th June,2019 | Tuesday | Learned the basic python concepts |
| 5th June,2019 | Wednesday | Learned the basic python concepts |
| 6th June,2019 | Thursday | Learned the basic python concepts |
| 7th June,2019 | Friday | free |
| 10th June,2019 | Monday | Learned the basic python concepts like list and tuples |
| 11th june,2019 | Tuesday | Learned the basic python concepts.  Sets and Dictionary |
| 12th June - 14th June,2019 | Wednesday-Friday | Free |
| 17th June,2019 | Monday | Learned the basic python concepts  If else conditions. |
| 18rd june,2019 | Tuesday | Learned the basic python concepts of loops and functions. |
| 19th june,2019 | Wednesday | Learned the basic python concepts of classes and modules. |
| 20th June – 21st June,2019 | Thursday-Friday | Learned the new module of GUI i.e. Tkinter. |
| 24th June – 27th June,2019 | Monday-Thursday | Design Form and game using Tkinter. |
| 28th June,2019 | Friday | Free |
| 1ST July – 3rd July.2019 | Monday-Wednesday | Learned SQL connection with Python. |
| 4th July,2019 | Thursday | Added it to the final project of Tkinter. (SQL connected with Tkinter) |
| 5th July,2019 | Friday | Allotment of voice assistant Project |
| 8th July,2019 | Monday | Web scrapping for the project. |
| 9th July- 10th July,2019 | Tuesday-Wednesday | Learned about different modules needed for project like Speech recognition etc. |
| 11th July,2019 | Thursday | Started Programming small functions which we want to add. |
| 12th July.2019 | Friday | Free |
| 15th July – 17th July,2019 | Monday-Wednesday | Combining all functions together and spotting errors in code. |
| 18th July,2019 | Thursday | Fixing all the errors and running the code. Finally all functions working, Few are not tried fixing them. |
| 19th July,2019 | Friday | Received the letter of internship and finally project is ready. |
|  |  |  |