**A**

**REAL TIME RESEARCH PROJECT**

**ON**

**SPEECH RECOGNITION USING PYTHON**

In partial fulfilment of the requirements for the award of the degree of

**BACHELOR OF TECHNOLOGY**

**IN**

**COMPUTER SCIENCE AND ENGINEERING (DATA SCIENCE)**

Submitted By

CH. SAI KAMALA (22UP1A6713)

SK. NASREEN (22UP1A6753)

A. KEERTHI (22UP1A6701)

B. ARUNDHATHI (22UP1A6707)

Under the Guidance of

**Dr. M. VISHNU VARDHANA RAO**

Associate Professor & Head of the Department

****

**DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING (DATA SCIENCE)**

**VIGNAN’S INSTITUTE OF MANAGEMENT AND TECHNOLOGY FOR WOMEN**

**Accredited to NBA (CSE & ECE) and NAAC A+**

**(Affiliated to Jawaharlal Nehru Technological University Hyderabad) Kondapur (Village), Ghatkesar-501301**

**Academic Year: 2023-2024**



**DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING (DATA SCIENCE)**

# Vignan_logo

# **CERTIFICATE**

This is to certify that this is the bonafide record of the mini project entitled **“SPEECH RECOGNITION USING PYTHON”, submitted by Ch. Sai Kamala (22UP1A6713), SK. Nasreen (22UP1A6753), A. Keerthi (22UP1A6701), B. Arundhathi (22UP1A6707)** of B. Tech in the partial fulfilment of the requirements for the degree of Bachelor of Technology in Computer Science and Engineering (Data Science) during the year 2023-2024. The results embodied in this mini project report have not been submitted to any other university or institute for the award of any degree or diploma.

### INTERNAL GUIDE HEAD OF THE DEPARTMENT

**Dr. M. VISHNU VARDHANA RAO Dr. M. VISHNU VARDHANA RAO**

Associate Professor & Head of the Department **Associate professor**

**External Examiner**

|  |
| --- |
|  |

**DECLARATION**

We, hereby declare that the results embodied in this dissertation entitled **“SPEECH RECOGNITION USING PYTHON”,** is carried out by us during the year 2023-2024 in partial fulfilment of the award of Bachelor of Technology in Computer Science and Engineering from VIGNAN’S INSTITUTE OF MANAGEMENT AND TECHNOLOGY FOR WOMEN is an authentic record of our work carried under the guidance of **Dr. M. VISHNU VARDHANA RAO**, Department of Computer Science and Engineering (Data Science). We have not submitted the same to any other university or organization for the award of any other degree.

**CH. SAI KAMALA (22UP1A6713)**

**SK. NASREEN (22UP1A6753)**

**A. KEERTHI (22UP1A6701)**

**B. ARUNDHATHI (22UP1A6707)**

**ACKNOWLEDGEMENT**

We would like to express sincere gratitude to **Dr. G. Apparao Naidu**, Principal, Vignan’s Institute of Management and Technology for Women for his timely suggestions which helped us to complete the project in time.

We would also like to thank **Dr. M. VISHNU VARDHANA RAO**, Head of the Department, Computer Science and Engineering (Data Science), for providing us with constant encouragement and resources which helped us to complete the project in time.

We would like to thank the project guide **Dr. M. VISHNU VARDHANA RAO**, Associate Professor & Head of the Department, Computer Science and Engineering (Data Science), for his timely cooperation and valuable suggestions throughout the project. We indebted to his for the opportunity given to work under his guidance.

We would like to thank our Project coordinator **Mrs. P. Chamundeswari, Mrs. N. Sreeja** for their kind guidance and encouragement from time to time.

Finally, we also thank all the teaching and non-teaching staff of Department of Computer Science and Engineering for their support throughout the project work.

**CH. SAI KAMALA (22UP1A6713)**

**SK. NASREEN (22UP1A6753)**

**A. KEERTHI (22UP1A6701)**

**B. ARUNDHATHI (22UP1A6707)**

**ABSTRACT**

One of the fastest-growing engineering innovations is speech recognition. This has been planned and built with that fact in mind, and some effort has been made to accomplish this goal. It has a variety of uses and possible benefits in a variety of fields. Nearly 20% of the world's population has some kind of disability, with a huge number of them being unseeing or incapable of properly handle their arms and some persons who are blind but have difficulty examination difficulties may listen to a researched article using an accessible device. In such situations, speech recognition systems come in-handy, allowing them to exchange information with others when running a device using voice input. A speech to text system (STT) converts speech into text in a human language format and text to speech system (TTS) translates text into speech in a human language format. The proposed device is a hardware solution for synthesizing speech and allowing voice access to digital content. In the modern era, the current technological development provides greater facilities for human life. Speech recognition system (ASR) is improved a lot from the early 1900’s to present it allows computer to understand human language speech recognition is a machines ability to listen to spoken words and identify them. You can then use speech recognition in python to convert the convert the spoken words into text make a query or give reply. You can even program some devices to respond to these spoken words. This system also recognizes the error between the given inputs and corrects them accordingly for easier communication. The time taken for correcting errors it would be between 0-15 or 20seconds.

**INDEX**

|  |  |  |
| --- | --- | --- |
| S. No | Topic | Pg.no |
|  | Abstract |  |
| 1. | Introduction   * 1. What is speech recognition   2. Motivation   3. Existing System   4. Challenges in Existing System   5. Proposed System   6. Advantages of the Proposed System   7. Objectives   1.8 Methodology | 1-9  1-2  3-4  5-  5-6  7  8  8  9 |
| 2. | Literature Review | 10-11 |
| 3. | Analysis  3.1 Speech Recognition Types | 12-14  13-14 |
| 4. | System Development  4.1 Speech Synthesis  4.1.1 Evaluation of synthetic speech  4.1.2 Building speech synthesis systems  4.2 Packages used  4.3 Implementation  4.4 code  4.5 Working with audio files | 15-20  15  15  15  16  17  18-19  20 |
| 5. | Performance Analysis  5.1 System Requirements  5.1.1 Requirements  5.1.2 Best Requirements  5.2 Hardware Requirements  5.3 Accuracy | 21  21  21  21  21  21 |
| 6. | Conclusion  6.1 Advantages of Proposed System  6.2 Disadvantages  6.3 Future work  6.4 Reference | 22-23  22  22  23  23 |

**LIST OF FIGURES**

|  |  |  |
| --- | --- | --- |
| S.no | Figure Name | Pg.No |
| 1 | Fig 1: conversion of speech into text | 1 |
| 2 | Fig.2: Unified Framework | 1 |
| 3 | Fig.3: speech to text conversion process | 2 |
| 4 | Fig 4: speech recognition | 4 |
| 5 | Fig.5: Sound waves | 6 |
| 6 | Fig 6:speech recognition using python | 9 |
| 7 | Fig 7: Input through Microphone | 11 |
| 8 | Fig 8: Analysis of voice | 12 |
| 9 | Fig 9: speech recognition pipeline | 14 |
| 10 | Fig 10:system development | 15 |
| 11 | Fig 11:Output | 19 |
| 12 | Fig 12: working with audio files | 20 |

**1. INTRODUCTION**

There is huge development in speech recognition technologies from last years as it had completely brought up huge progress based on the new machine learning algorithms. The speech recognition system proves to be beneficial in many aspects as it reduces the wastage of time as well as helps the disabled individuals.



Figure-1: speech recognition using python

Speech technology with fields within the scope of the paper are to be presented in Fig. as the unified framework that encompasses covered topics, showing their complementarity, ranges and borders, interconnections, and intersections in the interdisciplinary area of Speech.

UNIFIED FRAMEWORK:

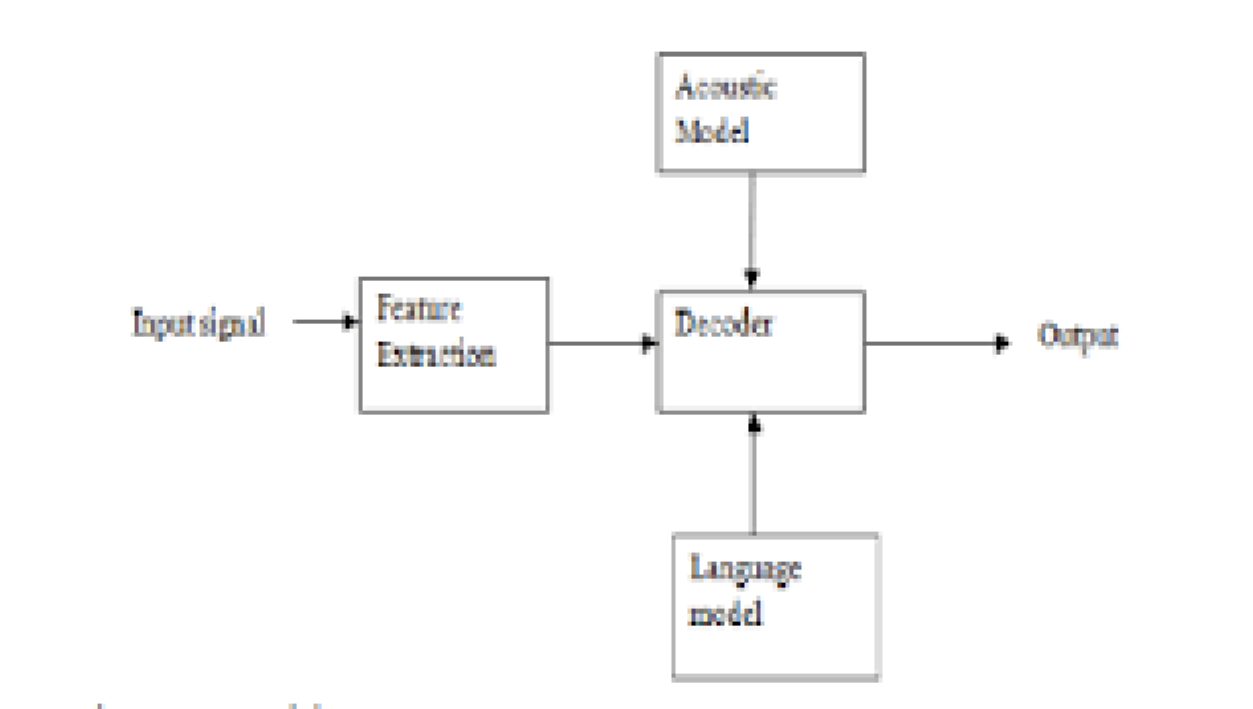


Fig.2 Unified Framework

In mostly areas of the country, there are lot of people who don’t know how to write and also how to read any word, so this project is very helpful for these type of people as you know in today’s world Everybody has its own mobile phones and they want to search a lot of things. In this project, they usually speak what they want to search, and various results of such type opens in the browser window.

In this project, we made our machine recognize the speech passed as the audio file as well as the dissection of the speech based on their requirements.

WHAT IS SPEECH RECOGNITION?

Speech recognition incorporates compute science and linguistics to identify spoken words and converts them to text. It allows computer to understand human languages. Speech recognition is a machine's ability to listen to spoken words and identify them. You can then use speech recognition in Python to convert the spoken words into text, make a query or give a reply .you can even program some devices to respond to these spoken words

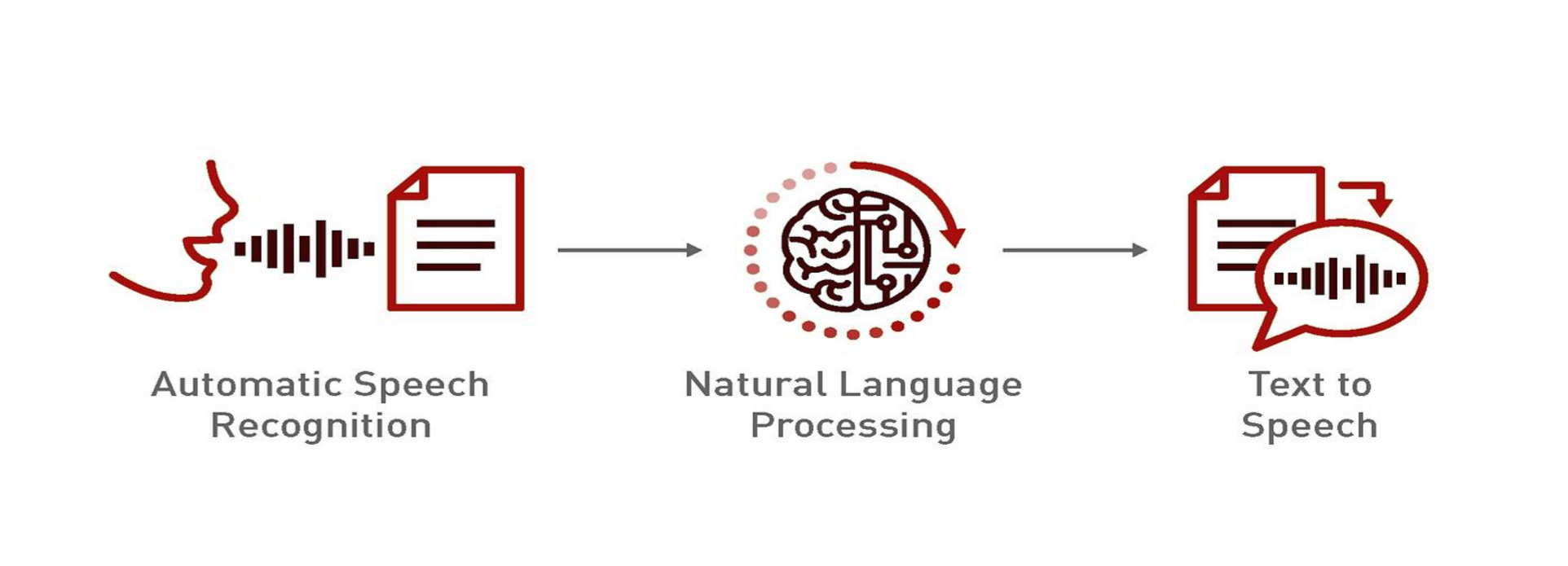


Fig.3 Speech to Text conversion

Step1 .Initially audio is taken as input with the help of microphone.

Step2 .In next step the audio input is passed through the Natural Language Processing. This helps in text processing, language understanding, and text generation. Making it easter to build robust and accurate speech recognition applications.

Step3. In the final step the voice /speech is converted into written text.

* 1. **Motivation for Speech Recognition Using Python**

***Accessibility and Inclusivity:***

* Speech recognition technology can make digital content more accessible to individuals with disabilities, such as those with visual impairments or mobility issues. By enabling voice commands and dictation, users can interact with technology more easily and efficiently.

***Improved User Experience:***

* Integrating speech recognition into applications can significantly enhance the user experience by providing a hands-free and intuitive method of interaction. This can be particularly useful in scenarios where typing is inconvenient or impossible, such as driving or cooking.

***Efficiency and Productivity:***

* Speech recognition can speed up tasks that would otherwise require manual input. For instance, professionals can use speech-to-text for quickly drafting emails, creating documents, or setting reminders, thus improving productivity and saving time.

***Technological Advancement and Innovation:***

* Developing speech recognition systems in Python contributes to the advancement of natural language processing (NLP) and artificial intelligence (AI). It pushes the boundaries of what machines can understand and respond to, fostering innovation in various fields such as customer service, healthcare, and education.

***Customizability and Flexibility:***

* Python offers a rich ecosystem of libraries and frameworks (like Speech Recognition, PyDub, and Deep Speech) that make it relatively easy to develop, customize, and extend speech recognition applications. This flexibility allows developers to tailor solutions to specific needs and integrate them with other Python-based systems.

***Research and Development:***

* Speech recognition is a vibrant area of research in AI and machine learning. Working on such projects can contribute to academic and commercial research, helping to solve complex problems related to language understanding, accent variation, noise reduction, and more.

***Market Demand and Career Opportunities:***

* There is a growing demand for voice-enabled applications in various industries, from smart home devices to virtual assistants and customer support bots. Gaining expertise in speech recognition using Python can open up numerous career opportunities in the tech industry.

***Personalization and User Interaction:***

* Speech recognition allows for the development of personalized applications that can learn and adapt to individual user preferences and speech patterns, providing a more engaging and responsive interaction.

***Language and Communication Enhancement:***

* Projects focused on speech recognition can also contribute to better communication tools, such as real-time translation services, language learning applications, and tools for individuals with speech disorders.

By leveraging Python for speech recognition, developers can create innovative and impactful solutions that address real-world challenges, improve accessibility, and enhance user experiences across various domains.



**Fig 4: speech recognition**

**1.2 Existing System**

There are several well-established algorithms and libraries for speech recognition in Python. Here are a few of the most popular ones:

***Google Speech Recognition API:*** Part of the Speech Recognition library, this API provides high accuracy and supports multiple languages. It requires an internet connection to work.

***CMU Sphinx (pocket sphinx):*** An offline speech recognition toolkit that is part of the CMU Sphinx project. It is not as accurate as Google's API but does not require an internet connection.

***Mozilla Deep Speech***: An open-source speech-to-text engine based on deep learning. It provides a high level of accuracy and can be used both online and offline.

***Wit.ai:*** An API that provides natural language processing capabilities, including speech recognition. It is free to use but requires an internet connection.

**1.3 Challenges in Existing System**

Challenges in existing system are:

**Drawbacks:**

Despite the popularity of voice recognition technology, there are some disadvantages to using it. While the automatic speech recognition may be the star of the show, it can still be a bit of a challenge to implement. For example, it may not be able to capture all words accurately because of pronunciation variations, or it may not be able to sort through background noise. If this is the case, it may be best to rely on a more accurate (and more expensive!) transcription service.

Using automatic speech recognition can be an advantage in some industries, such as the law industry. It can help lawyers reduce time spent on legal research and documenting their cases, and ensure the accuracy of their work. It also allows for more efficient internal processes. However, many users are still hesitant to use an ASR bot for sensitive tasks, for the reason we’ve mentioned above – lack of accuracy.

There are still many factors to consider before implementing automatic speech recognition in your office. For example, you must ensure that your office equipment is capable of recording quality audio, and that your software can accurately read the text produced.

Although the technology can make documentation easier, it can also result in errors. For instance, the system may not understand accents or slang. It may also take longer than anticipated to capture words correctly.

Some users are worried that they will not be able to trust the voice recognition system. They are hesitant to use ASR bots for sensitive tasks. The lack of trust may cause businesses to hesitate in adopting this technology.

The technology is also expensive to implement. This may include special hardware and software. Depending on the application, it may also require significant training. There may also be regulatory requirements.

Voice recognition software can be a distraction. It may not be able to differentiate between ambient noise and the actual speech. Wearing a noise-cancelling headset may help. Also, people who speak in accents need to learn to speak clearly so the system can recognize them. People must also avoid talking in a choppy manner or mumbling. This can lead to grammar and spelling errors.

Speech recognition may also have data privacy concerns. It is important to consider how a speech recognition system will handle your personal information. This information may include sensitive financial or medical information.

Then, it is often necessary to train your employees in the proper use of automatic speech recognition. This includes developing a training program based on different scenarios. A training program can also include a number of other features, such as filtering out background noise.



Fig 5: sound waves

**1.4 Proposed System:**

Creating a speech recognition project in Python typically involves several steps. Here's a basic methodology to guide you through the process

**1.Setting Up the Environment:**

• Install necessary libraries: speech\_recognition, pyaudio, and numpy.

• Ensure you have a working microphone and the necessary permissions to use it.

**2. Recording Audio:**

• Use pyaudio or similar libraries to capture audio input from the microphone.

• Save the recorded audio as a .wav file for processing.

**3.Processing Audio:**

• Use the speech\_recognition library to convert speech to text.

• This involves creating a Recognizer object and using it to process the audio.

**4.Handling Different Languages and Accents:**

• Configure the recognizer to handle different languages by specifying the language parameter.

• Use different recognizer instances or configurations to handle various accents and dialects.

**5.Error Handling:**

•Implement robust error handling to manage issues like background noise, unclear speech, and interruptions.

**1.5 Advantages of the Proposed System:**

In mostly areas of the country, there are lot of people who don’t know how to write and also how to read any word, so this project is very helpful for these type of people as you know in today’s world, everybody has its own mobile phones and they want to search a lot of things. In this project, they usually speak what they want to search and various results of such type opens in the browser window.

1. Ability to write text using speech.

2. Different windows can be opened and web searches can be made.

3. More utilization of resources and less time consumption.

4. Recognizes different audio files and convert them to text.

5. Helpful for disabled peoples.

**1.6 Objectives**

* To be familiar with the speech recognition and its fundamentals.
* Its working and application in different areas.
* To implement it as an application for relative searches.
* Software which can be used for:
* Speech recognition
* Web searches
* Word guessing

**1.7 Methodology**

The basic function of both speech synthesis and speech recognition is easy to understand as there are many powerful capabilities provided by speech recognition technology that helps many developers to understand and utilize this technology.

Despite the substantial growth and research in speech recognition technology there are still more limitations in this technology. Because of the speech recognition humans are able to utilize the time in various aspects and also it proves to be beneficial to various disabled peoples, still this system is unfamiliar with natural human to human conversations. The complete knowledge of the limitation also the strength is very important for the accurate use of speech recognition technologies as there may be differences in the output provided by the system and the output required by the user for a particular input. Due to this understanding the user or developers of these application can make a decisions about whether the technology will benefit the use of speech-to-text in a particular speech input.



Fig 6: Speech Recognition using python

CHAPTER 2

**Literature Survey:  *HISTORY: (1970-2010)***

The First speech recognition system was focused on numbers, not words. In 1952 bell Laboratory designed the “Audrey System” which could recognize a single voice speaking digits aloud. Ten years later IBM introduced “shoebox” which understood 16 words in English. Across the globe other nations developed hardware that could recognize sound and sleep. And by the end of ‘60s, the technology could support words with 4 vowels and nine consonants.

***From 1970’s:***

Speech recognition made several meaningful advancements in this Decade. This was mostly due to the US Department of defence and DARPA. The Speech Understanding Program SUR program there ran was one of the largest of its kind in the history of speech recognition. Mellon ‘Harpy Speech System came from this program and was capable of understanding over 1000 kind words that is about the same a three-year Old’s vocabulary. Also significant in the 70’s was Bell Laboratories introduction odd the system that could interpret multiple voices.

***From 1980’s:***

The ‘80s saw speech Recognition vocab go from few of hundred’s words to the several thousand words. One of the Breakthroughs that came from a statistical method known as the ‘Hidden Markov Model0 ‘HMM’ ‘. Instead of just using words and looking for the sound patterns. The Hmm estimated the probability of the unknown sounds actually being words.

***From 1990’s:***

Speech recognition was propelled forward in the 90s in the large part because of the own personal computer. The faster processors made it possible for software like dragon dictate to become the more widely used bell south introduced the Voice Portal (VAL) in which was a dial in interactive voice recognition system.

This System give new birth to the myriad of the phones tree system that are still in the existence today.

***From 2000’s:***

From the year 20002 Speech recognition Technology had achieved close to the 80 percent accuracy. For almost of all the Decade there aren’t a lot of Advancements till Google has come with a start of Google search voice. As it was an application which put speech recognition into hands of lakhs of people. This was also Significant because that the processing power would be offloaded to its data Centres. Not only for that, was Google Application collecting data from many billions of the searches which could help this to predict what a human is actually saying. That time Google’s English voice search system, included 240 billion words from user searches.

***From 2010’s:***

In 2012 Apple Launched SIRI which was as same as the Google’s VOICE SEARCH. The early part of the decade saw an explosion of the other voice Recognition Applications. And with Amazon’s ALEXA, Google Home we’ve seen consumers becoming more and more comfortable talking to Machines. Today, some of the Largest Technical Companies are competing to herald the speech accuracy title. In 2015, IBM achieved a word ERROR RATE of 6.8%. IN 2016 Microsoft overpassed IBM with a 5.8 % claim. Shortly After that IBM improved their Rate to 5.4 %. However, it’s Google that claims the lowest Ratio rate at 4.8percent.



Fig 7: Speech recognition through Microphone

CHAPTER 3

Analysis:

From apple SIRI to Smart Devices of home, Speech Recognition is very drastically used in our lives. This Speech Recognition project is to Utilize Cagle Speech Recognition Challenge Dataset to Create Kera Model on above of tenser flow & to create predictions in the voice files.

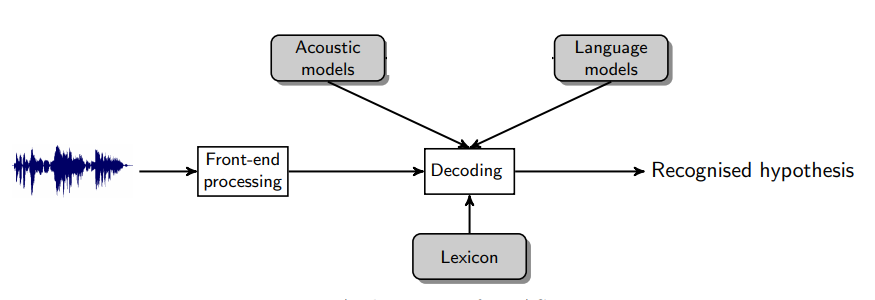


Fig 8: Analysis of Voice

Now, use the microphone to get audio input from the user in real-time, recognize it, and print it in text. As you can see, you have performed speech recognition in Python to access the microphone and used a function to convert the audio into text form.

3.1 SPEECH RECOGNITION TYPES

SPEECH RECOGNITION SYSTEM is basically divided into following depending on various types:

Speaking Mode:

Basically it means that how the words are been spoken as in connected or in isolated. In Isolated word of speech Recognition System needs that speaker take pause between the words he speak.It means single kind word In connected word of speech recognition system did not need that the speaker take pause briefly in between the words. It generally means full length sentences in which words are then artificially keep away by silence.

Speaking Style:

Generally it includes whether that the speech is in continuous form of spontaneous form. Continuous form is that spoken in natural form. Systems are too evaluated on speech read from the scripts that are prepared where as in spontaneous or extemporaneously generated, speech does not contain fluencies, and it is also difficult to figure out that speech read from the written script. It is also vastly much harder as it tends to be peppered with fluency like “UUH” and “UUM”, no full sentence, spluttering, stuttering, sneezing, cough, and also vocabulary is essentially unlimited, so there must be training to system to be able to tackle with unknown and hidden words.

Vocabulary:

IT is much simple to discriminate a smaller set of the words, but rate of error increase as the size of the vocabulary increases. For ex: 10 digits start from 0 -9 can easily be recognized rightly on the other side vocabulary whose size is 100, 4000 or 15000 have the rate of error as 3%, 6%, 40%. The vocabulary is hard to predict or recognize if it contains confused kind of words.

Enrolment:

This is kind of 2 ways

1) Speaker Dependent

2) Speaker independent

In speaker dependent the user must be providing various samples of her or his speech before they’re used, a speaker dependent system is meant for use of only single kind speaker, where as speaker independent system is allowed or intended to use any type or kind of speaker.

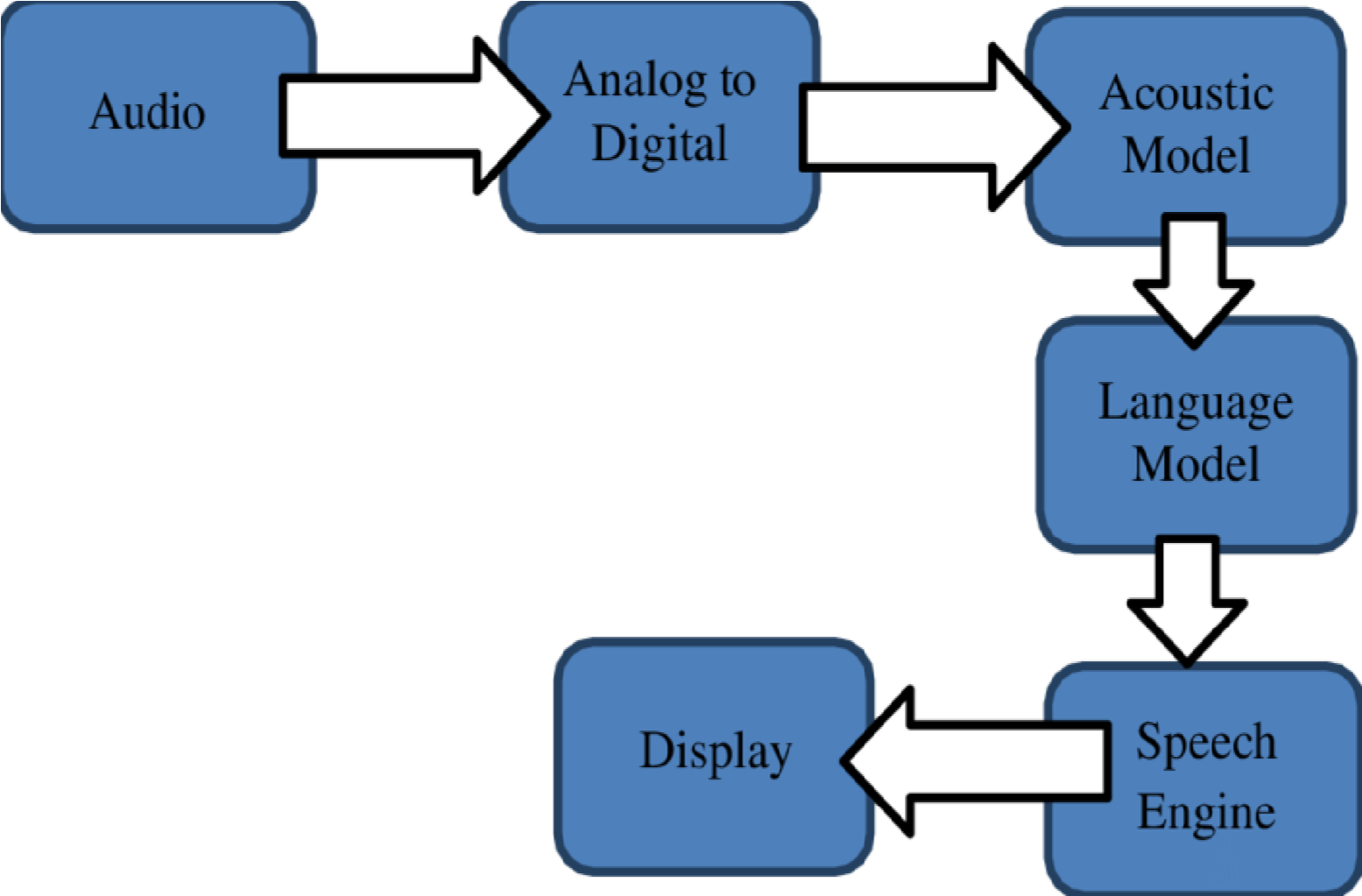


Fig:9 Speech Recognition Process

The above diagram explains about the speech recognition process:

Step1: Initially audio is taken as input by using microphone. Then this audio is sent to next step.

Step2: Here the analog signals are converted into digital signals.

Step3: Acoustic model converts raw audio signals into a sequence of phonetic units, which are then used by subsequent components of a speech recognition system to decode spoken language into text.

Step4: A Language model in speech recognition helps in understanding the meaning and structure of spoken language, thereby enhancing the accuracy and reliability of converting spoken words into written text.

Step5: It plays a crucial role in query processing, Information retrieval, Filtering and Natural Language Understanding.

Step6: After converting spoken words into text (speech-to-text), the “Display” function may involve showing the recognized text on a screen or Interface. This allows users to verify the accuracy of the transcription.

CHAPTER 4

SYSTEM DEVELOPMENT

4.1 Speech Synthesis

4.1.1 Evaluation of Synthetic Speech:

Speech Synthesis Systems can be calculate I terms of different requirements such as speech intelligibility, Speech Naturalness, System Complexity, and so on. For Ambient Intelligent Application it is Reasonable to imagine that new Evaluation Criteria will be Require for example , emotional Influence on the User, Ability to get the User to Act, mastery over Language generation, and Whether the system takes the Environmental Variables into Account and adjusts its behaviour Accordingly.

Some Of the Just Mentioned evaluation Criteria are for the Complete System. Having Evaluation Criteria for the Whole System is reasonable because a single, miss performing component would negatively impact how the system is perceived by humans.

4.1.2 Building Speech Synthesis Systems:

Building Speech Synthesis Systems require a speech Units Corpus. Natural Speech must have been recorded for all Units- For Example, all Phonemes – in all possible Contexts.

Next the Units in the Spoken Speech Data are segmented and labelled. Finally, the most Appropriate Speech Units are Chosen (Black and Campbell, 1995).

Generally, concatenative Synthesis yields high quality Speech. With the Large Speech Units Corpus, high quality speech waveforms can be generated. Such synthesized speech preserves waveforms can be generated. Such synthesized speech preserves naturalness and intelligibility. Separate prosody modelling is not necessary for speech unit selection due to the availability of many units corresponding to varied contexts. Picture

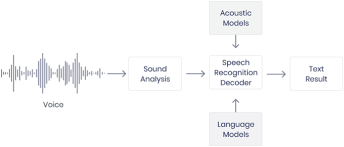


Fig 10: System Development

4.2 Packages Used :

The following is the install packages in this project:

1.Import speech recognition:

Speech recognition helps to take the input with ease and helps in running model in just a few minutes.

The speech recognition library has several popular speech APIs and is thus extremely flexible. It consists of seven APIs which can be used to speech recognition but all six APIs comes with authentication key and password except Google speech API which makes it extremely flexible and with its ability of free usage and ease of use it makes it excellent choice for speech recognition.

2. Import audio:

The pip install Audio command installs py audio to the python interpreter and thus make it easier to work with microphones which helps in real time speech recognition. With Py Audio, we can easily use Python to record and to play audio on a kind of variety of platforms.

3. Import web browser :

with this package we can make use of our default browser used to locate, retrieve and display data .The URL and the query is passed to the instance of the web browser package and basis on the” URL” provided and the query the particular webpage opens.

Speech Recognition is an important feature in several applications used such as home automation, artificial intelligence, etc. This article aims to provide an introduction on how to make use of the Speech Recognition and pyttsx3 library of Python.

**Implementation**

**SAMPLE CODE:**

* **Installation required:**
* **Python Speech Recognition module:**
  + pip install speech recognition

PyAudio: Use the following command for Linux users

* + sudo apt-get install python3-pyaudio
  + Windows users can install pyaudio by executing the following command in a terminal
  + pip install pyaudio
  + Python pyttsx3 module:  pip install pyttsx3

 Speech Input Using a Microphone and Translation of Speech to Text

* **Allow Adjusting for Ambient Noise:**

Since the surrounding noise varies, we must allow the program a second or too to adjust the energy threshold of recording so it is adjusted according to the external noise level.

* **Speech to text translation:**

This is done with the help of Google Speech Recognition. This requires an active internet connection to work. However, there are certain offline Recognition systems such as Pocket Sphinx, but have a very rigorous installation process that requires several dependencies. Google Speech Recognition.

* **Translation of Speech to Text:**

First, we need to import the library and then initialize it using init() function. This function may take 2 arguments.

* Init (driver Name string, debug bool)
* driver name: [Name of available driver] sapi5 on Windows | nsss on MacOS
* debug: to enable or disable debug output

After initialization, we will make the program speak the text using say() function.   
 This method may also take 2 arguments.

Text: anything you want to hear

**python**:

import speech\_recognition as sr

# Initialize recognizer class (for recognizing the speech)

recognizer = sr.Recognizer()

# Function to capture and recognize speech

def recognize\_speech\_from\_mic(recognizer, microphone):

# Adjust for ambient noise

with microphone as source:

print("Adjusting for ambient noise, please wait...")

recognizer.adjust\_for\_ambient\_noise(source)

while True:

# Capture audio

with microphone as source:

audio\_data = recognizer.listen(source)

try:

# Recognize (convert from speech to text)

text = recognizer.recognize\_google(audio\_data)

print(text)

except sr.UnknownValueError:

print("Sorry, I could not understand the audio.")

except sr.RequestError:

print("Sorry, my speech service is down.")

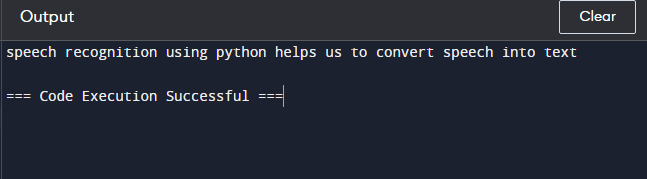
# Create a microphone instance

mic = sr.Microphone()

# Start recognizing speech

recognize\_speech\_from\_mic(recognizer, mic)

OUTPUT:



**Working with Audio Files:**

After the installation of Speech Recognition in the command line it becomes easy to use the audio files because of its an Audio file class. The path of the audio file can be passed as the argument to the Audio File class and it also provides with the context manager as it helps in reading and working with the file material. The context manager then is responsible for opening of the audio file and finally stores the data of file in the instance of the Audio File. Then the record () method is used to store the data from the entire audio file and initialize it into the instance of Audio Data. The recognize google () is used to recognize any kind of speech in the audio. The results depend on the internet’s connection speed and are displayed and the speech to text conversion depends immensely on the accent and the speed of the speaker. As we have used the audio file our speech recognition system caught some words differently because of the vocabulary of the speaker.

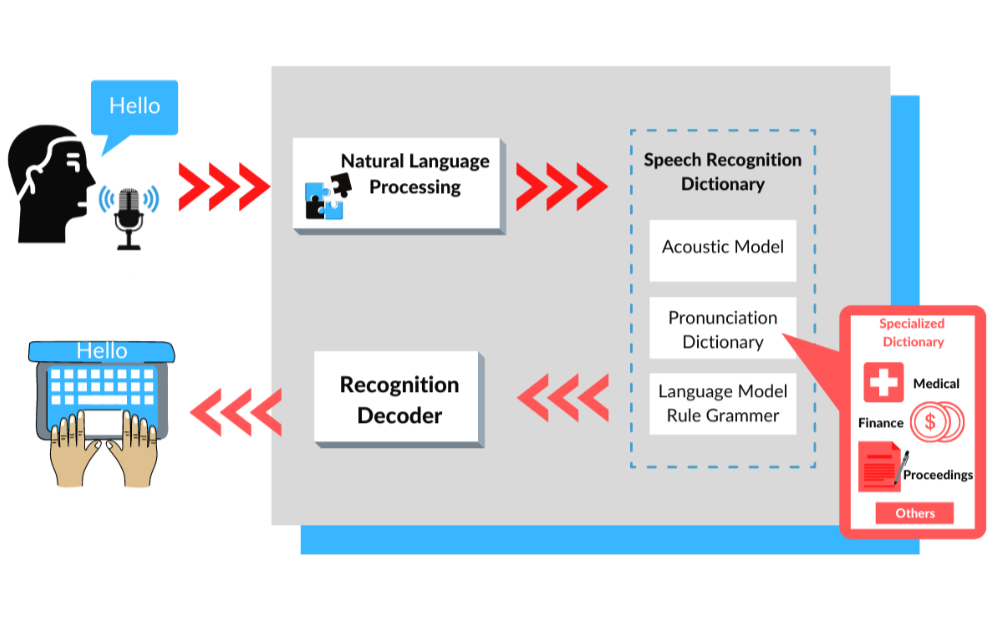


Fig 11: Working of Speech Recognition

**CHAPTER 5**

**Performance Analysis:**

5.1 System Requirements:

5.1.1 Requirements:

a. 1.6 MHz Processor

b. 128 MB RAM

c. Microphones for good audio.

5.1.2 Best Requirements:

a. 2.4 GHz processor

b. Greater than128 MB RAM

c.10% consumption of memory

d. Best quality microphones

5.2 Hardware Requirements:

Microphones:

Microphones are the most important tools for the real time speech to text conversion in . Therefore the pre-installed ones cannot be used as they are more prone to the background noise and also of poor quality terms of speech.

Computer Processor:

Speech recognition application depends majorly on processing speed. The input from the user can take some

time if the processing speed is low and thus user wasted more time on waiting compared to performing the

task which makes the application less feasible for use

## **5.3 Accuracy:**

Various factors can affect the accuracy of speech recognition. For example, a speaker’s age and accent can thwart the system’s ability to recognize speech. Another challenge is the number of variations found in human speech. These variations include speech sounds, inflections, and accents. Fortunately, recent advances in artificial intelligence have made it possible to develop speaker-independent speech recognition. The technology uses [self-supervised learning models](https://medium.com/geekculture/self-supervised-learning-finds-speech-recognition-d3460ecfc06), allowing the system to learn more about the speakers’ voices and adjust its accuracy. One of the most important advances in speaker-independent speech recognition is deep learning. This technology utilizes neural networks as feature transformations. By utilizing faster processors and cheaper memory, the system can increase its vocabulary size and recognition accuracy. In addition, deep learning has decreased word error rates by 30%.

**6. CONCLUSION**

The project of speech recognition gives us the introduction of this technology and its various application in different sectors. The project is divided into three parts, the first which helps in converting audio to text, the second which recognises the spoken word and the third which performs the operations provided as the command by the user. After the development of these parts these models were tested and the results were produced which tells about the accuracy of each model. Various advantages and disadvantages of this software is discussed.

**6.1 Advantages of Proposed Method:**

In mostly areas of the country, there are lot of people who don’t know how to write and also how to read any word, so this project is very helpful for these type of people as you know in today’s world, everybody has its own mobile phones and they want to search a lot of things. In this project, they usually speak what they want to search and various results of such type opens in the browser window.

1. Ability to write text using speech.

2. Different windows can be opened and web searches can be made.

3. More utilization of resources and less time consumption.

4. Recognizes different audio files and convert them to text.

5. Helpful for disabled peoples.

**6.2 Disadvantages:**

1. Low accuracy because of its limited ability.

2. Fails in noisy environment.

3.Depends majorly on Google API thus not an original software.

4. Limited operations can be performed

**6.3 Future work:**

The tech to support speech Applications is today both Relatively Inexpensive and Powerful. With the betterment or the advance tech in Artificial Intelligence and to the increase amounts of Speech Data that can be easily mined, it is now possible to that voice becomes the next Dominant Interface. At So, We can also applause the many companies before us that propelled speech Recognition to where it is today. We Automate Transcription workflow and make it fast, easy and more affordable. We wouldn’t do this without the proper Work that has to been done before we.

**6.4 Reference:**

[1]-------KOGILA RAGHU Speech emotion recognition system performance analysis with optimized features using different classification algorithms