Mini Project Report On

### **SPEECH TO TEXT AND TEXT TO SPEECH CONVERSION**

### **(MINI ASSISTANT)**

Submitted to **Jawaharlal Nehru Technological University** in partial fulfillment of the requirements for the award of degree of

### **BACHELOR OF TECHNOLOGY**

In

### **Computer Science and Engineering**

Submitted By

###### **DOULAGHAR BHAVANA (17R01A05D0)**

**JOGI NIHARIKA (17R01A05D6)**

**NISHA THAKUR (17R01A05F2)**

**YAMSANI SRUJANA (17R01A05H8)**

Under the Guidance of

**Mr.Ch. Mahender Reddy M.Tech.,(PhD)**

(Asst.professor, CSE)

**CMR INSTITUTE OF TECHNOLOGY**

**(UGC Autonomous)**

(Approved by AICTE, Affiliated to JNTU Hyderabad)

Kandlakoya,Medchal Road,Hyderabad

**CMR INSTITUTE OF TECHNOLOGY**

**(UGC Autonomous)**

(Approved by AICTE, Affiliated to JNTU Hyderabad) Kandlakoya, Medchal Road , Hyderabad.

|  |  |
| --- | --- |
|  |  |
| **CERTIFICATE** |  |

This is to certify that the Mini Project "**SPEECH TO TEXT AND TEXT TO SPEECH CONVERSION**" is being

Submitted By

###### **DOULAGHAR BHAVANA (17R01A05D0)**

**JOGI NIHARIKA (17R01A05D6)**

**NISHA THAKUR (17R01A05F2)**

**YAMSANI SRUJANA (17R01A05H8)**

The results in this project have been verified and are found to be satisfactory. The results embodied in this work have not been submitted to any other University for the award of any other degree or diploma.

**SIGNATURE OF GUIDE SIGNATURE OF COORDINATOR SIGNATURE OF HOD**

Mr.Suresh Nomula Dr.A.M.Krishnan Dr.B.Satyanarayana

(Asst.Professor, CSE) (Assoc.Professor, CSE) (HOD, CSE)

|  |
| --- |
|  |

**ACKNOWLEDGEMENT**

We are extremely grateful to **Dr.M.Janga Reddy**, Director, **Dr.B. Satyanarayana,**Principal and **Dr. B.Satyanarayana**, HOD, Department of CSE, CMR Institute of Technology for their inspiration and valuable guidance during the entire course duration.

We are extremely grateful to **Mr. A.M.Krishnan**, Mini Project Coordinator and Internal Guide **Mr. Suresh Nomula**, Department of CSE, for their constant guidance, encouragement and moral support throughout the project.

We will be failing in duty if I do not acknowledge with gratitude thanks to the authors of the references and other literature referred to in this project.

We express our thanks to all staff members and friends for all the help and co-ordination extended in bringing out this project successfully in time.

Finally, we are very much thankful to our parents and relatives who guided directly or indirectly for every stage towards success.

DOULAGHAR BHAVANA (17R01A05D0)

JOGI NIHARIKA (17R01A05D6)

NISHA THAKUR (17R01A05F2)

YAMSANI SRUJANA (17R01A05H8)

**ABSTRACT:**

In present industry communication is a key element to progress.Passing on information, to the right person, and in the right manner is very important, not just on a corporate level, but also on a personal level.Speech recognition technology is one from the fast growing engineering technologies.Nearly 20% people of the world are suffering from various disabilities,many of them are blind and unable to use their hands efficiently.They share information with people by operating computer through voice input. Speech processing is widely used in many applications like security devices, household appliances, cellular phones, ATM machines and computers. The human computer interface has been developed to communicate or interact conveniently for one who is suffering from some kind of disabilities. Speech-to-Text Conversion (STT) systems have a lot of benefits for the deaf or dumb people and find their applications in our daily lives.Our proposed model is capable to recognize the speech and convert the input audio into text and further into audio.

**TABLE OF CONTENTS**

1. INTRODUCTION 6

1.1 SYSTEM STUDY 7

2. SYSTEM ANALYSIS 8

2.1 EXISTING SYSTEM 9

2.2 PROPOSED SYSTEM 9

2.3 ALGORITHM 10

3.REQUIREMENT ANALYSIS 11

3.1 SOFTWARE REQUIREMENTS 12

3.2 HARDWARE REQUIREMENTS 12

4. SYSTEM DESIGN 13

4.1 SYSTEM ARCHITECTURE 14

5. IMPLEMENTATION 17

5.1 TECHNOLOGIES USED 18

5.2 SAMPLE CODE 20

6.OUTPUT 23

7.CONCLUSION 26

8.REFERENCES 28

**INTRODUCTION**

## **1.1 SYSTEM STUDY**

**FEASIBILITY STUDY**

The feasibility of the project is analyzed in this phase and business proposal is put forth with a very general plan for the project and some cost estimates. During system analysis the feasibility study of the proposed system is to be carried out. This is to ensure that the proposed system is not a burden to the company. For feasibility analysis, some understanding of the major requirements for the system is essential.

###### Three key considerations involved in the feasibility analysis are

* ECONOMICAL FEASIBILITY
* TECHNICAL FEASIBILITY
* SOCIAL FEASIBILITY

**ECONOMICAL FEASIBILITY**

This study is carried out to check the economic impact that the system will have on the organization. The amount of funds that the company can pour into the research and development of the system is limited. The expenditures must be justified. Thus the developed system as well within the budget and this was achieved because most of the technologies used are freely available.

**TECHNICAL FEASIBILITY**

This study is carried out to check the technical feasibility, that is, the technical requirements of the system. Any system developed must not have a high demand on the available technical resources. This will lead to high demands on the available technical resources. This will lead to high demands being placed on the client. The developed system must have a modest requirement, as only minimal or null changes are required for implementing this system.

**SOCIAL FEASIBILITY**

The aspect of study is to check the level of acceptance of the system by the user. This includes the process of training the user to use the system efficiently. The user must not feel threatened by the system, instead must accept it as a necessity. The level of acceptance by the users solely depends on the methods that are employed to educate the user about the system and to make him familiar with it. His level of confidence must be raised so that he is also able to make some constructive criticism, which is welcomed, as he is the final user of the system.

# 

# 

# 

# **SYSTEM ANALYSIS**

**2.1 EXISTING SYSTEM:**

In the current existing system we have only the conversion of speech to text but there is no further conversion into speech.Speech recognition is the ability of a computer software to identify words and phrases in spoken language and convert them to human readable text.Speech to text transcription service that converts the audio or speech files into an readable message,then send to user via mail or text.

**2.2 PROPOSED SYSTEM:**

We have improvised the outcome of the existing system by converting the produced output into speech i.e Text-to-speech(TTS).Text-to-speech (TTS) convention transforms linguistic information stored as data or text into speech. It is widely used in audio reading devices for blind people nowadays . In the last few years however, the use of text-to-speech.It usually consists of two parts.First it takes the raw text and converts letters,numbers etc into its word equivalents.This process is often called text normalization,preprocessing,or tokenization.Then it assigns phonetic transcriptions to each word,and divides and marks the text into various linguistic units like phrases,clauses,and sentences.In second it takes the symbolic linguistic representation and converts it into actual sound output.Additional functionalities are added to our proposed model such as Knowing the current date and time.One can also enable search by asking what he/she wants to search and then our model redirects to the google page by showing the result.Finding the location is also a added feature to our proposed model.People can also interact with our model by putting forward the questions like “what is your name?”.

**2.3** **ALGORITHM:**

Step1:Start

Step2:Import all modules

Step3:Initialize recognizer class

Step4:Initialize microphone as a source method

Step5:Take the audio as input

Step6:Give the text conversion of that audio as output

Step7:If user gives text as input

Step8:If the user asks “What is your name?” then it should reply “My name is mini”

Step9:If the user asks “What time is it?” then it should print time using ctime method

Step10:If the user asks “What do you want to search for?” then do the needful using the webbrowser model

Step11:If the user asks “What is the location?” then answers that using google maps.

Step12:End

# 

**REQUIREMENT**

**ANALYSIS**

###### 

###### 

###### **REQUIREMENT SPECIFICATION**

###### **3.1** **Software Requirements**

###### 

###### **Operating Systems supported**

###### Windows 10

* Windows XP
* Windows 8

###### **Technologies and Languages used to Develop**

* Python

**Integrated Development Environment**

* IDLE Python 3.8

**3.2 Hardware Requirements**

For developing the application, the following are the Hardware Requirements:

* Processor : Pentium IV or higher
* RAM : 256 MB
* Space on Hard Disk : Minimum 512MB

# 

# 

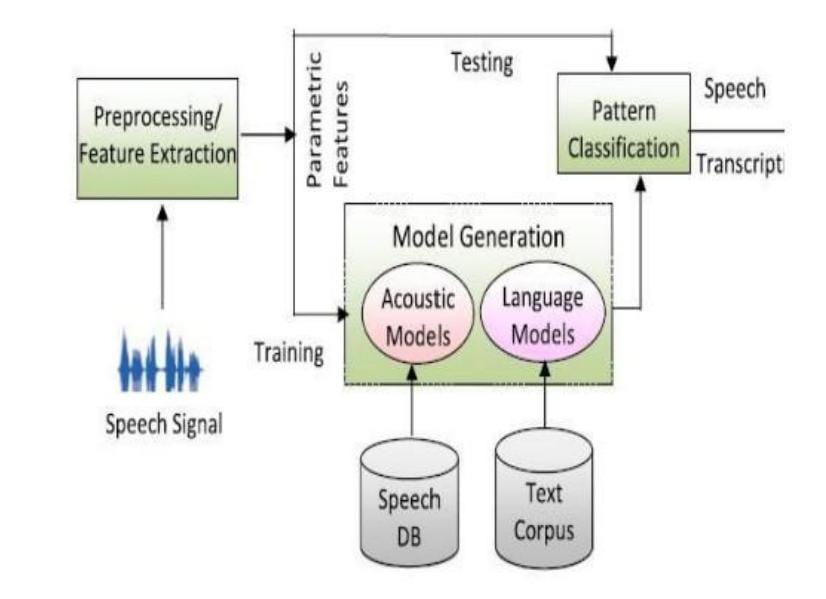
# 

# **SYSTEM DESIGN**

# 

# 

**4.1 SYSTEM ARCHITECTURE:**



i) **Pre-processing:** The analog speech signal is transformed into digital signals for later processing. This digital signal is moved to the first order filters to spectrally flatten the signals. This helps in increasing the signal’s energy at a higher frequency.

ii) **Feature Extraction:** This step finds the set of parameters of utterances that have a correlation with speech signals. These parameters, known as features, are computed through processing of the acoustic waveform. The main focus is to compute a sequence of feature vectors (relevant information) providing a compact representation of the given input signal. Commonly used feature extraction techniques are discussed below:

* **Linear Predictive Coding (LPC):** The basic idea is that the speech sample can be approximated as a linear combination of past speech samples.The digitized signal is blocked into frames of N samples. Then each sample frame is windowed to minimize signal discontinuities. Each framed window is then auto- correlated. The last step is the LPC analysis, which converts each frame of autocorrelations into LPC parameter set.
* **Mel-Frequency Cepstrum Coefficient (MFCC):** It is a very powerful technique and uses a human auditory perception system. MFCC applies certain steps to the input signal: Framing: Speech wave- form is cropped to remove interference if present; Windowing: minimizes the discontinuities in the signal; Discrete Fourier Transform: converts each frame from time domain to frequency domain; Mel Filter Bank Algorithm: the signal is plotted against the Mel spectrum to mimic human hearing.
* **Dynamic Time Warping:** This algorithm is used for measuring the similarity between two-time series which may vary in speed, based on dynamic programming. It aims at aligning two sequences of feature vectors (1 of each series) iteratively until an optimal match (according to a suitable metrics) between them is found.

iii) **Acoustic Models:** It is the fundamental part of Automated Speech Recognition (ASR) system where a connection between the acoustic information and phonetics is established. Training establishes a correlation between the basic speech units and the acoustic observations.

iv) **Language Models:** This model induces the probability of a word occurrence after a word sequence. It contains the structural constraints available in the language to generate the probabilities of occurrence. The language model distinguishes words and phrases that have a similar sound.

v) **Pattern Classification:** It is the process of comparing the unknown pattern with existing sound reference pattern and computing similarity between them. After completing the training of the system at the time of testing, patterns are classified to recognize the speech. Different approaches for pattern matching are:

* **Template Based Approach:** This approach has a collection of speech patterns which are stored as a reference representing dictionary words. Speech is recognized by matching the uttered word with the reference template.
* **Knowledge Based Approach:** This approach takes a set of features from the speech and then trains the system to generate a set of production rules automatically from the samples.
* **Neural Network Based Approach:** This approach is capable of solving more complicated recognition tasks. The basic idea is to compile and in- corporate knowledge from a variety of knowledge sources with the problem at hand.
* **Statistical Based Approach:** In this approach, variations in speech are modelled statistically (e.g. HMM) using training methods.

**IMPLEMENTATION**

## 

## **5.1TECHNOLOGIES USED**

### **PYTHON**

Python is a general-purpose interpreted, interactive, object-oriented, and high-level programming language. An [interpreted language](https://en.wikipedia.org/wiki/Interpreted_language), Python has a design philosophy that emphasizes code [readability](https://en.wikipedia.org/wiki/Readability) (notably using [whitespace](https://en.wikipedia.org/wiki/Whitespace_character) indentation to delimit [code blocks](https://en.wikipedia.org/wiki/Code_block) rather than curly brackets or keywords), and a syntax that allows programmers to express concepts in fewer [lines of code](https://en.wikipedia.org/wiki/Source_lines_of_code) than might be used in languages such as [C++](https://en.wikipedia.org/wiki/C%2B%2B) or [Java](https://en.wikipedia.org/wiki/Java_(programming_language)). It provides constructs that enable clear programming on both small and large scales. Python interpreters are available for many [operating systems.](https://en.wikipedia.org/wiki/Operating_system)

**MODULES:**

* Python Speech Recognition module
* Python webbrowser module
* Python time module
* Python playsound module
* Python os module
* Python gTTS module

**Python Speech Recognition module:**

* Library for performing speech recognition, with support for several engines and APIs, online and offline.
* The first component of speech recognition is, of course, speech. Speech must be converted from physical sound to an electrical signal with a microphone, and then to digital data with an analog-to-digital converter.

**Python webbrowser module:**

* The webbrowser module provides a high-level interface to allow displaying Web-based documents to users.
* Under most circumstances, simply calling the open() function from this module will do the right thing.You have to import the module and use the open() function.

**Python time module:**

* This module provides various time-related functions. For related functionality, see also the [datetime](https://docs.python.org/3/library/datetime.html#module-datetime) and [calendar](https://docs.python.org/3/library/calendar.html#module-calendar) modules.
* Although this module is always available, not all functions are available on all platforms.

**Python playsound module:**

* The playsound module is a cross platform module that can play audio files. This doesn't have any dependencies, simply install with pip in your virtual environment and run.
* Implementation is different on platforms.

**Python os module:**

* The OS module in python provides functions for interacting with the operating system.
* OS, comes under Python's standard utility modules.
* This module provides a portable way of using operating system dependent functionality.
* These modules include many functions to interact with the file system.

**Python gTTS module:**

* gTTS (*Google Text-to-Speech*), a Python library and CLI tool to interact with Google Translate text-to-speech API.
* Writes spoken mp3 data to a file, a file-like object (bytestring) for further audio manipulation, or stdout.
* It features flexible pre-processing and tokenizing.

**5.2 SAMPLE CODE**

import speech\_recognition as sr

import webbrowser

import time

import playsound

import os

import random

from gtts import gTTS

from time import ctime

r=sr.Recognizer()

def record\_audio(ask=False):

with sr.Microphone() as source:

if ask:

alexis\_speak(ask)

audio=r.listen(source)

voice\_data=''

try:

voice\_data=r.recognize\_google(audio)

except sr.UnknownValueError:

alexis\_speak('sorry,Idid not get that')

except sr.RequestError:

alexis\_speak('sorry,my speech service is down')

return voice\_data

def alexis\_speak(audio\_string):

tts=gTTS(text=audio\_string,lang='en')

r=random.randint(1,1000000)

audio\_file='audio-'+str(r)+'.mp3'

tts.save(audio\_file)

playsound.playsound(audio\_file)

print(audio\_string)

os.remove(audio\_file)

def respond(voice\_data):

if 'what is your name' in voice\_data:

alexis\_speak('My name is alexix')

if 'what time is it' in voice\_data:

alexis\_speak(ctime())

if 'search' in voice\_data:

search=record\_audio('what do you want to search for?')

url='https://google.com/search?q='+search

webbrowser.get().open(url)

alexis\_speak('here is what i found for'+search)

if 'find location' in voice\_data:

location=record\_audio('what is the location?')

url='https://google.nl/maps/place/' +location+'/&amp;'

webbrowser.get().open(url)

alexis\_speak('here is the location of'+location)

if 'exit' in voice\_data:

exit()

time.sleep(1)

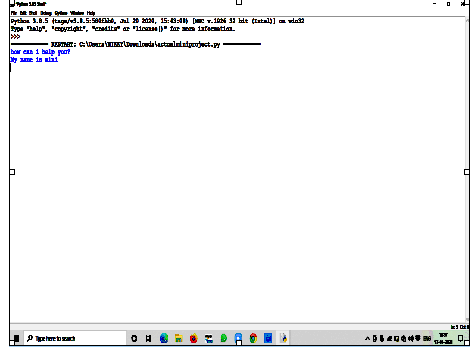
alexis\_speak('how can i help you?')

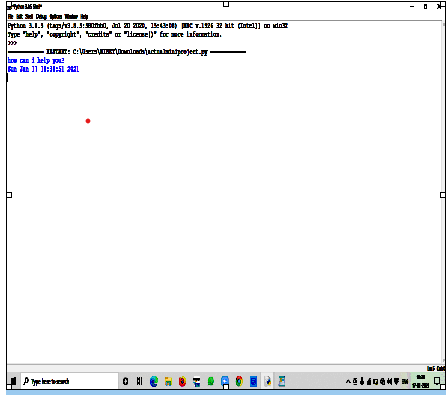
while 1:

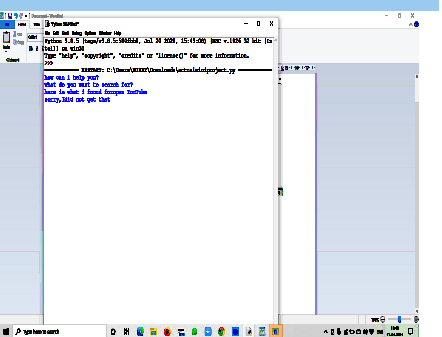
voice\_data=record\_audio()

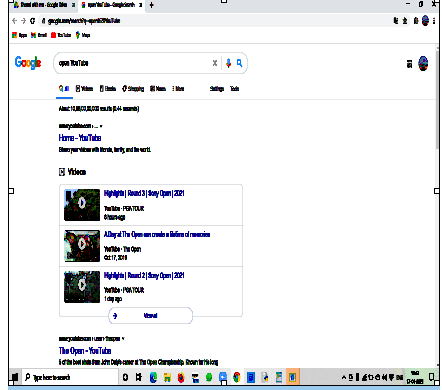
respond(voice\_data)

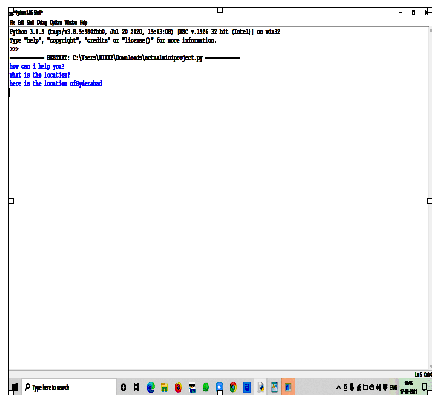
**OUTPUT**

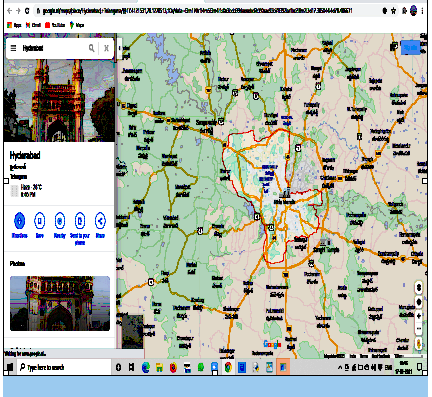
****

****

****

****

****

****

**CONCLUSION**

**CONCLUSION:**

The speech to text conversion may seem effective and efficient to its users if it produces natural speech and by making several modifications to it. This system is useful for deaf and dumb people to Interact with the other peoples from society. Speech to Text synthesis is a critical research and application area in the field of multimedia interfaces.Our proposed model is capable to recognize the speech and convert the input audio into text and further into audio.The system gives the input data from mice in the form of voice, then preprocessed that data & converted into text format displayed on PC.

**REFERENCES:**

* <https://hackernoon.com/how-to-convert-speech-to-text-in-python-q0263tzp>
* <http://www.igntu.ac.in/eContent/IGNTU-eContent-815947141046-MA-Linguistics-4-HarjitSingh-ComputationalLinguistics-5.pdf>
* <https://pypi.org/project/SpeechRecognition/>
* https://www.programiz.com/python-programming/time
* [https://pypi.org/project/gTTS](https://pypi.org/project/gTTS/)/
* <https://docs.python.org/3/library/os.html>
* <https://docs.python.org/3/library/webbrowser.html>
* <https://pypi.org/project/playsound/>