Library Audiobook System Using Speech Recognition

[1]Nikhat Parveen, [2]Priyanka CH, [2]Ruchitha Y ,[2]Geeteeka Y,[2]VarniPriya

[1]Associate Professor, Department of Computer Science & Engineering, Koneru Lakshmaiah Education Foundation, Guntur, A.P., India,

[2] Btech Student, Department of Computer Science & Engineering, Koneru Lakshmaiah Education Foundation , Guntur, A.P., India,

Abstract:-Developments in man-made consciousness are beginning to be used successfully in human life, enabled by the use of things in different ways. Independent devices are becoming more cunning in their way of communicating in modern ways by using new technologies. The invention of perceiving modern technology to a person is one of the useful trends. For voice recognition, this project presents an exceptionally on-demand, fast, and user-friendly Python using Visual Studio Code. This software helps transform text to speech and vice versa. Voice or some other type of voice can be converted into the text instead of tapping on a keyboard. Any kind of voice message can be stated in this app as the input, and the corresponding text will appear on the screen. Modern technology can easily generate text. The recognized text can be stored in a file with GUI/Voice command. This application useful for people with physical disabilities like deaf or handicapped users who are often difficult, painful, or impossible to type, and also for people who can understand while others speaking but cannot read the words.

Keywords: - Visual Studio Code, Voice Recognition, Python, handicapped, GUI/Voice command.

Introduction: - The main aim of this project is to develop a web application for voice recognition. In this project, we convert text to speech, speech to text, pdf to speech, and image to speech. Here we take the user input in form of voice and then convert it into speech or taking input in the form of pdf and convert it into speech or taking input in the form of image and convert it into speech or raking input in the form of text and converting it into text[1]. This saves time for various activities or for people who don't know how to read or for people who can't see it. Life will be easier with this kind of development. People will have more excitement when talking about recognizing their voice. This web application will make more money in the market. Anyone can have access to it as it is web developed from wherever you need to. None of the applications would perform this kind of action based on voice recognition. This application makes it easy to convert it into various forms[2]. Instead of typing on a keyboard, we can speak so that this application converts takes input as speech and convert it into text for us. In this application, any kind of voice message can be converted[3]. A Speech Recognition. This paper is developed more for deaf and blind people. This might be the best web application in the future.

Text-to-Speech: - In our project we mainly used the text-to-Speech module is to know how to spell the words if the user doesn't know how to spell.

Pdf-to-Speech: - This module is used to read a book that is in pdf format.

Image-to-Speech: - Through this module, the user can take a picture of any paragraph in the book and it converts the words in the captured image to the speech.

Speech-to-Text: - Here the given text is converted into speech.

Literature Survey: -Before starting the project we did a survey on why people are interested in audiobooks rather than reading and also, we noted the reasons if they are not interested in audiobooks.

Most of the interested people answers are that listening to a book is convenient than reading a book and it is very useful for the people who can understand the language but cannot read, so, through audiobook system they can listen to the books they liked even though they cannot read. This Audiobook System is also useful for blind people so rather than reading they can listen to the book and understand or feel the book[4,5]. Some people those who don't like audiobook system the main reason is that they can improve their reading skills while reading a book rather than listening to a book so that they can be perfect in the language[6]. While some people say that they are happy that they don't need to carry a book every if they are suddenly interested in reading and some others say that they can do multitasking i.e., by listening to a book, they can do some other work at the same time. After reading the responses from the survey we got to know that there are more advantages than disadvantages of the audiobook system[7,8].

Methodology:-The complete project is developed in Python which is user-friendly as well as it is an object-

oriented programming language. Our project is totally web-based project. In this project we flask to connect HTML and python. There are totally four modules. They are, read a book, Listen to the captured text, Spell a word, Write a sentence. In this, we can choose where we are interested in listening by a female or male voice [9,10].

Read a book: -

```python
def pdf_to_audio(input_file):

    pdfFileObj = open(input_file, "rb")
    pdfReader = PyPDF2.PdfFileReader(pdfFileObj)


    for num in range(pdfReader.numPages):
        page = pdfReader.getPage(num)
        text = page.extractText()
        player = pyttsx3.init()

        player.say(text)
        player.runAndWait()
```

Fig 1: pdf-to-speech

In read a book module first we have imported pyPDF2,pyttsx3,pdfreader.In fig 1 we can observe that pdfFileobj gets the path of the selected pdf, through pyPDF2.pdfFileReader the opened file is assigned to pdfReader.By pyttx3 the audio is assigned and started reading either in a male or female voice as per the selected voice[11].

Spell a word: -

```python
def text_to_speech(text, gender):
    """
    Function to convert text to speech
    :param text: text
    :param gender: gender
    :return: None
    """
    voice_dict = {'Male': 0, 'Female': 1}
    code = voice_dict[gender]

    engine = pyttsx3.init()

    # Setting up voice rate
    engine.setProperty('rate', 120)

    # Setting up volume level  between 0 and 1
    engine.setProperty('volume', 0.8)

    # Change voices: 0 for male and 1 for female
    voices = engine.getProperty('voices')
    engine.setProperty('voice', voices[code].id)

    engine.say(text)
    engine.runAndWait()
```

Fig 2:Text-to-Speech

This module converted the given text into speech according to the given voice of choice. Here in this module, we have imported the pyttsx3 module to start the voice. We can also set the different rates and volumes as to our wish.

Write a Sentence: -

```python
recognizer = sr.Recognizer()

''' recording the sound '''

with sr.Microphone() as source:
    print("Adjusting noise ")
    recognizer.adjust_for_ambient_noise(source, duration=1)
    print("Recording for 4 seconds")
    recorded_audio = recognizer.listen(source, timeout=4)
    print("Done recording")

''' Recorgnizing the Audio '''
try:
    print("Recognizing the text")
    text = recognizer.recognize_google(
            recorded_audio,
            language="en-US"
    )
    print("Decoded Text : {}".format(text))

except Exception as ex:
    print(ex)
```

Fig 3: Speech to text

This is where the given speech is converted into text. We have imported the recognizer module to listen to the user's voice. And the given voice is converted into text using recognizer_google. Here recognizer.listener is a user to clearly listen to the sentence given by the user.
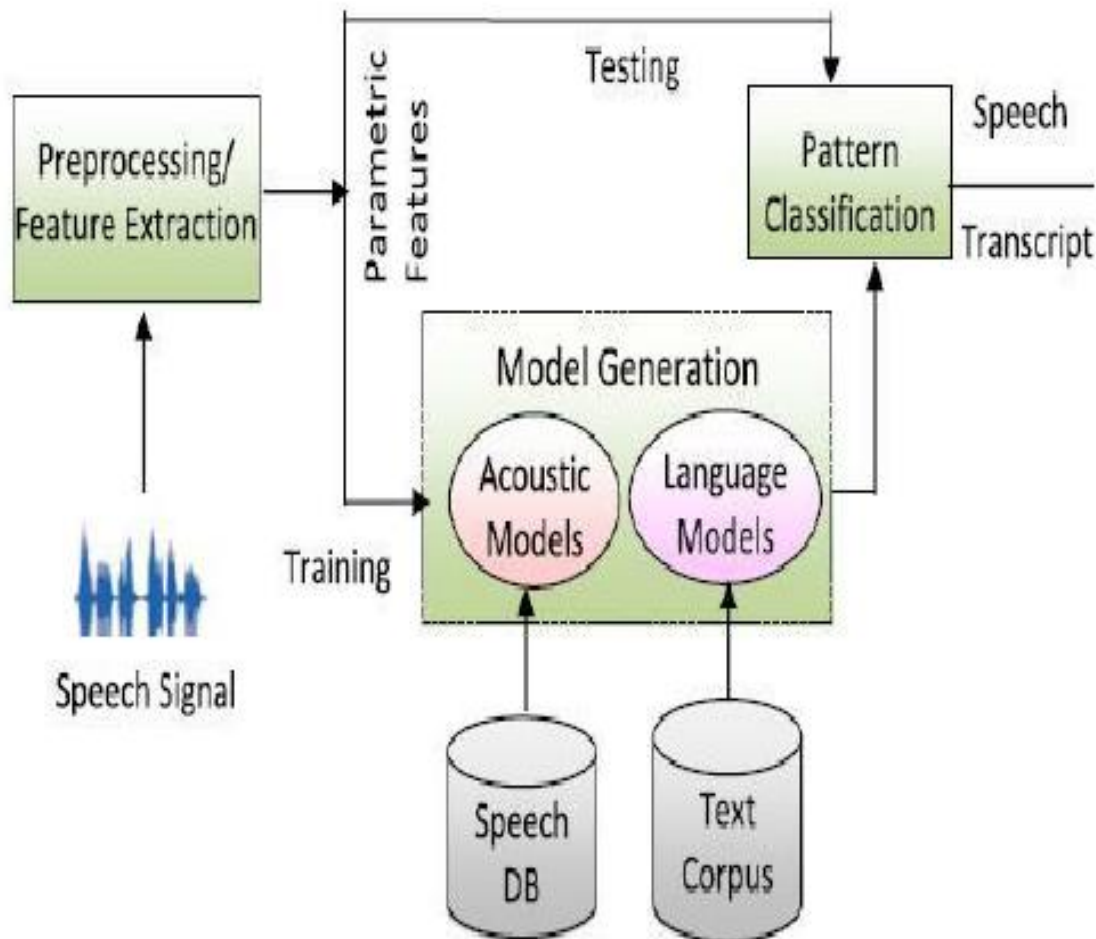

Fig 4: Architecture of Speech Recognition

The main methodology behind this speech recognition is that the given speech signal into small frames that is what preprocessing or feature extraction is. As we can see in fig 4 the stage is that the frames which are divided are compared with the database according to whether they are language-based or grammar-based, and it corrects according to the database. The correction of the speech and the speech comparison with the given speech is Model Generation. After the frames are converted into text according to the database they are combines to form a meaningful sentence that is patter classification do. And the pattern classification is checked with the Feature Extraction to get the maximum correct output.
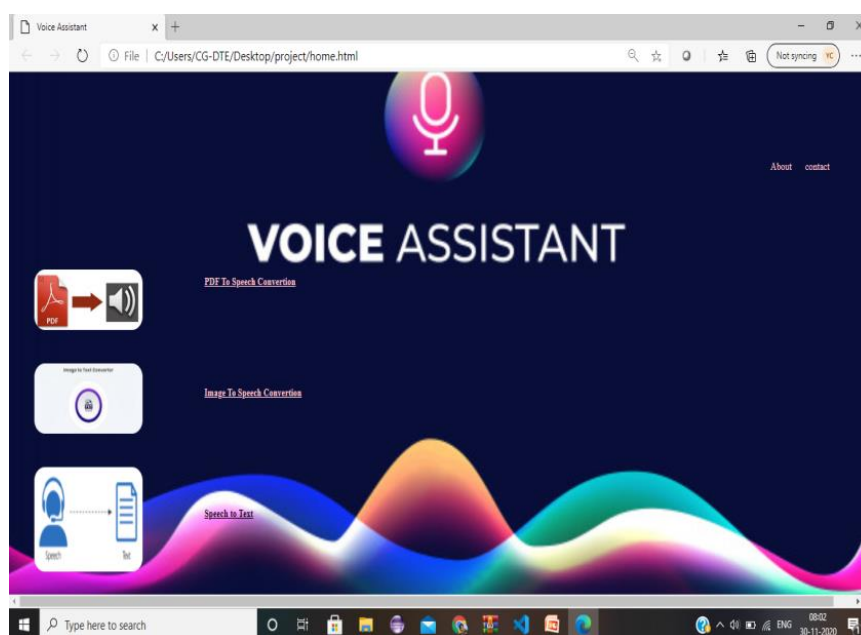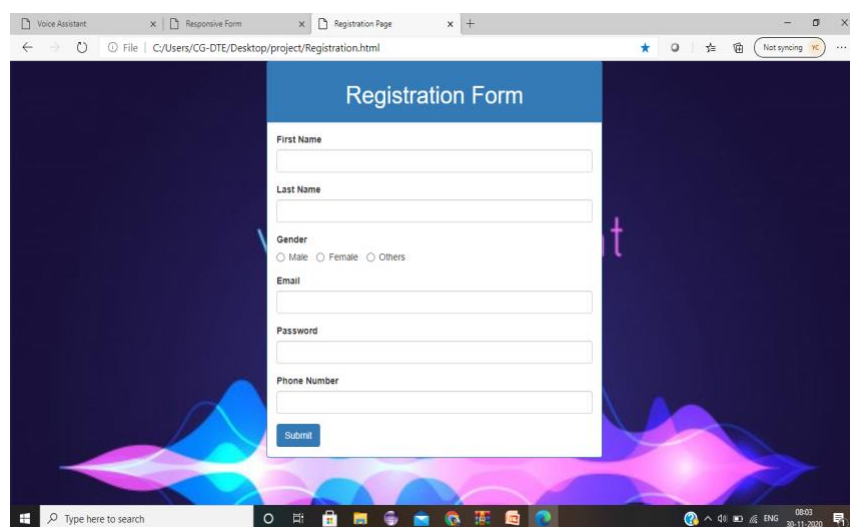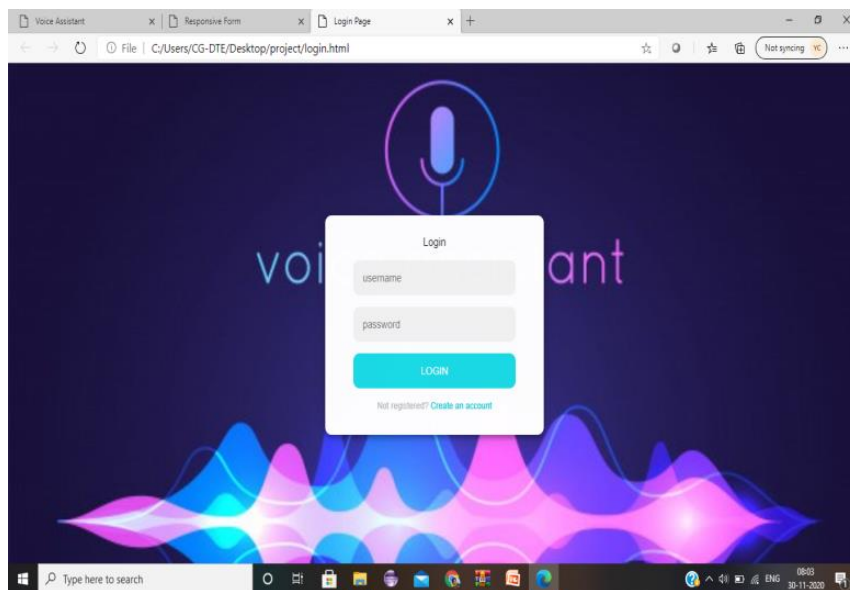
Acknowledgment: -The possibility of effectiveness needed to make the acquaintance more precise and quicker while connecting to the customer. This work can be further enhanced by updating the voice command in Google search queries. Better affirmation of discourse with the aim that the client can acquire immediate output and applications such as pc trying to lock or pc launching on the client's orders.
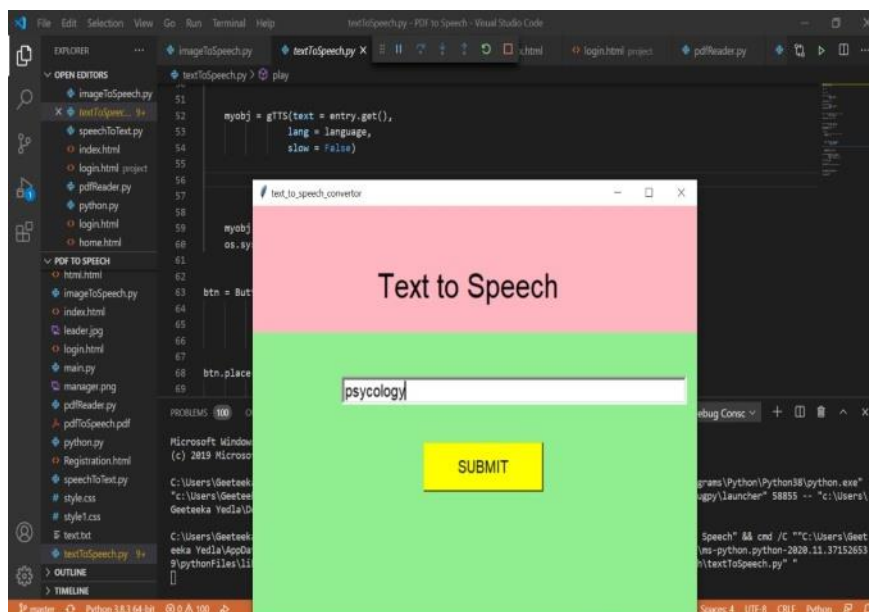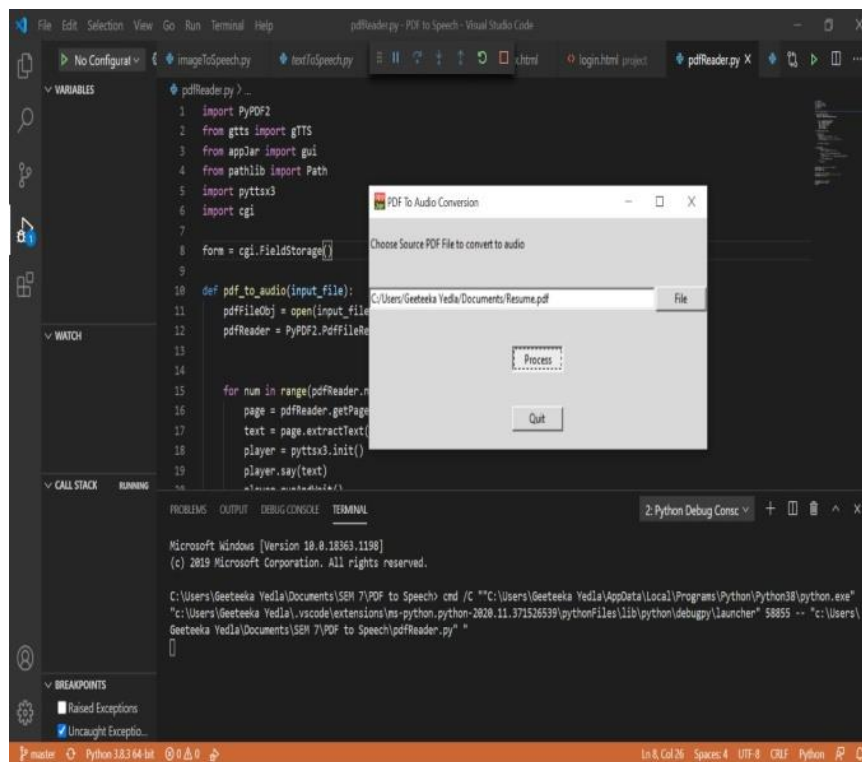
Including every time without fail so that there would be the probability of attaching a component, like saving the client's information, and when needed, the structures will be filled with the positive predictors. Our conceptual methodology can also be expressed to the IoT applications Throughout the future, our proposed approach will be able to decode the text-based portrayal in many beneficial effects. Upgrading to this approach must be feasible by attaching the features of the cash recognition.

Result: -
The code and procedure for the implementation is clearly explained and presented in the following link.
GitLab Link:<u>https://projects.kluniversity.in/170031438/batch65</u>

Future Scope: -As per our project present, we declare that a person who is having the interest to learn English if he comes from his native language background our project helps then more present we are having 4 modules in that like speech-text conversion, text-speech conversion,pdf-speech conversion,image-pdf conversion this concept in a single platform ad's beauty to our project. In further we want to update this project by adding some more modules which can help future generation people access a variety of functions related to the library in this platform that makes the library system of student into easy way .as per the present technology this type of conversions have more impact on the generation as we can take an example of Alexa and so on which makes the life in a smooth way. In the same manner, the conversions also play a major role. In the future, we will implement our project in life by having a library in this virtual manner where we can upload any book into it and get the content of the book as a voice where just by listing, we can just gain knowledge in the best way as the every one say the thing which we hear makes a major impact.

Conclusion: -In this Paperwork on speech recognition, we started with a brief introduction to android technology. Any person can use this application for the recognition of text or speech. In the days ahead, our proposed platform will be introduced in a multilingual application, enabling users to use the software in their own language without difficulty.

Our proposed structure will be able to decipher the text-based representation in a significantly improved manner. This application is useful likewise in daily life playing with their every single movement.

References: -

[1] Ayushi Trivedi,Navya Pant, Pinal Shah, Simran Sonik and Supriya Agrawal DepartmentofComputerScience, NMIMS University,Mumbai,India. Corresponding Author: Navya Pant.Speech to text and text to speech recognition systems-AreviewArtificialintelligence(AI),sometimescalledmachineintelligence.https://www.iosrjournals.org/iosr-jce/papers/Vol20-issue2/Version-1/E2002013643.pdf

[2] Speech Recognition using Android. Bhushan Mokal, Sahil Patil, Aniket Kale, Prof. Archana Arudkar in 2020

https://www.irjet.net/archives/V7/i2/IRJET-V7I2628.pdf

[3] B.Marr,The Amazing Ways Google Uses Deep Learning AI. Cortana Intelligence, Google Assistant, AppleSiri.

[1] Priyanka, J.H., Parveen, N, Clustering algorithms and storage of clustered data in green cloud environment using hadoop, Journal of Green Engineering " ISSN: 2245-4586, 2020, 10(9), pp. 4744-4751,

[2] Hill, J., Ford, W.R. and Farreras, I.G., 2015. Real conversations with artificial intelligence: A comparison between human– human online conversations and human–chatbot conversations. Computers in Human Behavior, 49,pp.245-250.

[3] "CMUSphnix Basic concepts of speech - Speech Recognition process". http://cmusphinx.sourceforge.netlwiki/tutorialconcepts

[4] Huang,J.,Zhou,M.andYang,D.,2007,January.ExtractingChatbotKnowledgefromOnlineDiscussionForums.In IJCAI(Vol. 7, pp.423-428).

[5] Varish, N., Parveen, N, et.al, Image Retrieval Scheme Using Quantized Bins of Color Image Components and Adaptive Tetrolet Transform, IEEE Access 2020, 8, pp. 117639-117665, 9121956

[6] Parveen, N., Roy, A., Sai Sandesh, D., Sai Srinivasulu, J.Y.P.R., Srikanth, N., Human computer interaction through hand gesture recognition technology, International Journal of Scientific and Technology Research, 2020, 9(4), pp. 505-513

[7] Gayathri .S , Porkodi Venkatesh , PushpapriyaPremkumar on Voice Assistant for Visually Impaired in 2019, https://ijesc.org/upload/40664e91149af2618afd09aaf1fca8f8.Voice%20Assistant%20for%20Visually%20Impaired%20(2).pdf

[8] Mohasi, L. and Mashao, D., 2006. Text-to-Speech Technology in Human-Computer Interaction. In 5th Conference on Human Computer Interaction in Southern Africa, South Africa (CHISA 2006, ACM SIGHI) (pp.79-84).

[9] Fryer, L.K. and Carpenter, R., 2006. Bots as language learning tools. Language Learning &Technology.

[10] J Kiran, N Parveen, Holistic Review of Software Testing and Challenges, International Journal Of Innovative Technology and Exploring Engineering (IJITEE) , Volume.8, Issue 7, Page No pp.1506-1521, June 2019,

[11] Sreeram, G., Pradeep, S., Rao, K.S., Raju, B.D., Nikhat, P. , Moving ridge neuronal espionage network simulation for reticulum invasion sensing, International Journal of Pervasive Computing and Communications, 2020