International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 09 Issue: 04 | Apr 2022 www.irjet.net p-ISSN: 2395-0072

Text Extraction from Image, Word, PDF and Text-to-Speech Conversion

Pooja Bendale[1], Sanika Badhe[2], Sarthak Bhagat[3], Pravin Rahate[4]

[1,2,3]Student, Computer Engineering Department, Mumbai University, Datta Meghe College of Engineering, Navi Mumbai, Maharashtra, India
[4]Assistant Professor, Computer Engineering Department, Mumbai University, Datta Meghe College of Engineering, Navi Mumbai, Maharashtra, India

---

Abstract - *Speech is one of the most seasoned and most regular method for information exchange between human. Throughout the long term, Endeavours have been made to foster vocally intuitive PCs to acknowledge voice/speech synthesis. Clearly such a point of interaction would yield extraordinary advantages. For this situation a computer can incorporate text and give out a speech. Text-To-speech is an innovation that gives a method for changing composed text from a clear structure over to a communicated in language that is without any problem reasonable by the end client (Fundamentally in English Language).*

Keywords—Speech Recognition, OCR technology, image extraction, text to speech

1. INTRODUCTION

This project is a motivation to develop an advanced software engine which could scrap textual data from clean and distorted pdf text, documented images and handwritten text and transfer the corresponding electronic data into speech signals. At the heart of this software engine lies an OCR Engine (Optical Character Recognizer) which inherits crucial morphological operations required for image conditioning & transformation, accompanied with python libraries used for character classification.

Further the processed textual data is transformed into speech signals using various Text-to-Speech synthesis techniques.

Text to speech synthesis is mainly based upon the concept of OCR. OCR stands for optical character recognition. Optical character Recognition (OCR) is a process that converts scanned or printed text images, handwritten text into editable text for further processing. This paper has presented a robust approach for text extraction and converting it to speech history trades back to telegraph during 20th century.

The first computer-based speech synthesis systems were created in the late 1950s, and the first complete text-to-speech system was completed in 1968. Earlier TTS was based on format synthesis and articulate synthesis, another method employed was diphone synthesis. BY 1990's Unit selection synthesis came to be used it was basically an update of diphone synthesis by improving pitch. After popularization of machine learning approach is neural network where the model can also b trained to predict further words and sentences. Most important properties of TEXT to speech are naturalness, accuracy and intelligibility. In the project we have a followed a simple method to devise TTS by using python

This project will be developed using gTTs, and playsound library.

In this project, we add a message which we want to convert into voice and click on play button to play the voice of that text message

2. OBJECTIVE

Multi-tasking busy lifestyle now a days demands doing work simultaneously. This text to speech is greatly beneficial as one can listen text documents, books, images even while having lunch

Boon for handicapped and special children who have lost their ability to visualise. Millions of children and adults can complete their education enjoy reading and also gain knowledge and not feel left out

For students who are more comfortable listening rather than reading or who understand it better when one is listening to it , Auralmate can be useful, user friendly and handy as one can download the recording free of cost and is portable and mobile

Further coupled with AI and machine learning it can be effectively used in guidance and navigation tools, conservational interactive voice response and smart home devices

3. METHODOLOGY

Character recognition or optical character recognition (OCR), is the process of converting scanned images of machine printed or handwritten text (numerals, letters, and symbols), into a computer format text. Speech synthesis is the artificial

synthesis of human speech. The text peruse on our site is an expert program for the transformation of text to speech. A computerized instrument just requires transferring of a text-based document from the client's end.

The remaining process is performed by our algorithm and libraries at backend. This entire strategy could appear to be drawn-out and tedious, however the clients don't need to hang tight for in excess of several seconds on this device for the text to voice change. It's a speedy apparatus for changing any sort of text over to expressed words without putting forth any attempts.

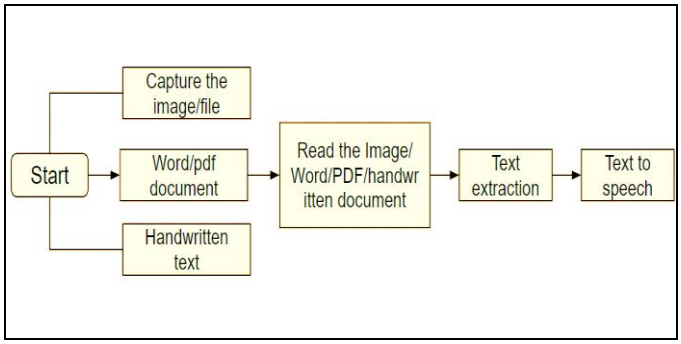

Fig -1: Sample Table format

System consists various stages named image capture, image pre-processing, image filtering, character recognition and text to speech conversion. The software used is python with set of libraries. The basic framework is a system that captures and converts that text to speech.

Prerequisites include python 3.0 is an interpreted, object-oriented, high-level programming language with dynamic semantics. Python library used while executing the project are:

1. Flask: powerful tool used to create a user-friendly object-oriented interface. Fast and easy way to create GUI applications.
2. Pytesseract: It's an Optical character recognition tool.It mainly performs the task of recognising the text embedded in the document, file, image . It can read all types of images and file format like png, jpeg, jpg giff
3. gTTs: google text to speech. It's used to create a mp3 file of extracted text. Unlimited lengths are allowed
4. PIL: python image library basically to add extra capability to python interpreter

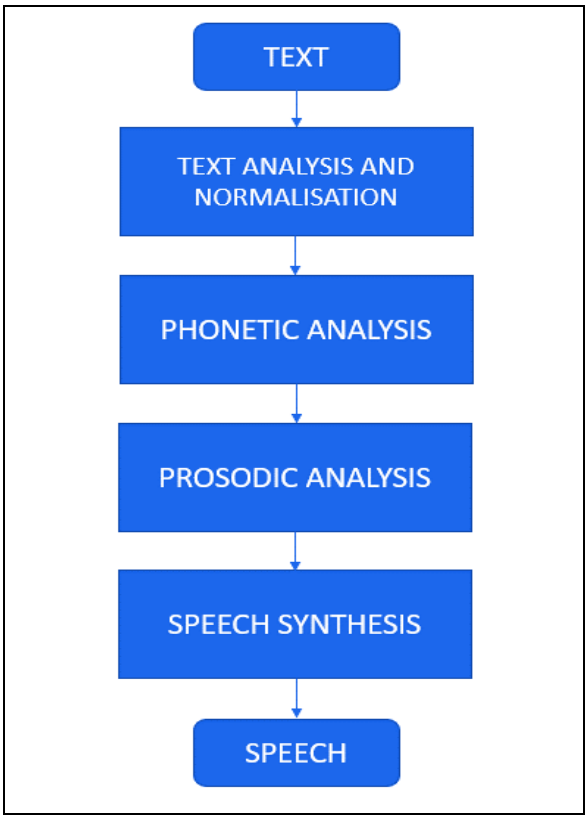

Fig -2: Proposed System

Our proposed system works as follows:

Step 1: User uploads document

Step 2: It is saved temporarily to our server

Step 3: Program checks what file type

Step 4: We select the appropriate method for the file type

Step 5: Text is extracted

Step 6: User is allowed to look at the extracted text

Step 7: That text is sent to the TTS API

Step 8: User downloads the audio file

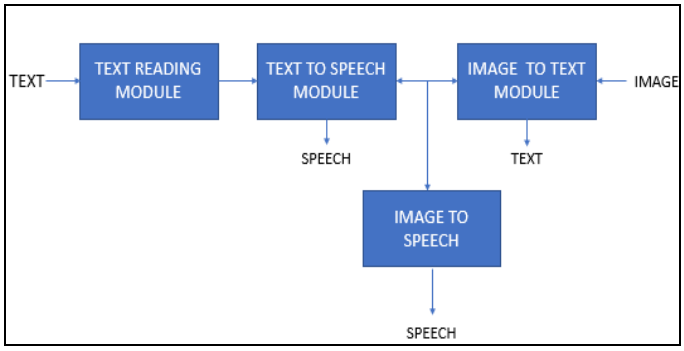

Fig -3: Work flow

4. DEPLOYMENT OF MODEL

The algorithm is served through a website that is hosted on heroku.

The text in the image is recognized and saved in the Text document.Now click on the convert button that identifies the text in the given image.

Fig -4: Front-end to upload file

The output of the text document is converted to voice.

The file is then downloaded on the user's device.

Fig -5: Screenshot of output file being downloaded

5. FUTURE SCOPE

This technology has been researched for a long time now. Collaboration with other algorithms and approaches and by increasing its capabilities it can be useful in many real life applications.

The text reader on our website is a professional program for the conversion of text to speech. It's an automated tool that only requires uploading of a textual file from the user's end. The rest of the process for the text speech conversion is done by the advanced algorithms of our tool in the backend. This whole procedure might seem tedious and time-consuming, but the users don't have to wait for more than a couple of seconds on this tool for the text to voice conversion. It's an expeditious tool for converting any type of text to spoken words without making any efforts. You can make use of our converter on the go; the users aren't restricted to get themselves registered for using our service.

Accessibility: Aural text-to-speech solutions provide improved digital accessibility to populations with learning and speech disabilities, visual impairments, and low literacy across devices and platforms. Audio enabled website and Augmented and Alternative Communication (AAC) devices and other communication devices used by those with a speech impairment.

Automotive: Can be effectively used in navigation systems and GPS, Outbound correspondences among showrooms and clients for things like arrangement affirmations, planned assistance updates, and advancement and deals updates can without much of a stretch be computerized utilizing one of Auralmate's applications Government websites can be made read aloud hence they can reach to each and every person. It can be used for emergency alerts and speech enabled tax visa fillings Health: Can be effectively used in health monitoring, medical devices, dial in pharmacy and appointment reminders

6. CONCLUSION

Today and in coming future there will be huge demand of TTS and audio assistance. In this paper we have attempted to extract text from text documents, images as well as handwritten text Also the model works with satisfactory accuracy. By this approach text and images from a word document, Web page or e-Book can be read and can generate synthesised speech through a computer's speakers.

Further prospects in the project will be to streamline the same technology for handwritten text.

REFERENCES

[1] Nafiz Arica, Student Member, IEEE, and Fatos T. Yarman-Vural, Senior Member, IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE, VOL. 24, NO. 6, JUNE 2002

[2] N.P. Narendra ·K. Sreenivasa Rao ·Krishnendu Ghosh ·Ramu Reddy Vempada ·Sudhamay Maity, Int J Speech Technol (2011) 14:167–181DOI 10.1007/s10772-011-9094-4

[3] Sagar Patil, Mayuri Phonde, Siddharth Prajapati, Saranga Rane and Anita L ahane, International Journal of Engineering Research & Technology (IJERT), ISSN: 2278-0181, Vol. 5 Issue 04, April-2016

[4] Prof. Teena Varma, Stephen S Madari, Lenita L Montheiro and Rachna S Pooojary, International Journal of Engineering Research & Technology (IJERT), ISSN: 2278-0181, NTASU - 2020 Conference Proceedings.