

e- Narrator (Image/PDF to Speech Converter)

Dr. Asif Ali¹ Ankit Thakur² Harsh Malviya³ Harshal Jain⁴

¹Associate Professor ^{2,3,3}B.E. Scholar

^{1,2,3,4}Department of Information Technology

^{1,2,3,4}A.I.T.R, Indore, Madhya Pradesh, India

Abstract— e-Narrator is a free website that automatically speaks any portable document format (PDF) and Image which you upload. Its main aim is to help patients suffering from visual impairment. The main idea of this project is to recognize the text character and convert it into speech signal. The text contained in the page is first pre-processed. The pre-processing module prepares the text for recognition. Then the text is segmented to separate the character from each other. Segmentation is followed by extraction of letters and resizing them and stores them in the text file. These processes are done with the help of Optical Character Recognition (OCR). This text is then converted into speech.

Keywords: Text Processing, Image Processing, Text-to-Speech, Image-to-Text, Speech-to-Text

I. INTRODUCTION

The most significant type of correspondence is speech communication rather the communication is characterized by speech itself. In our regular day to day existence, we have to speak with other individuals so as to do our tasks. This has alluded as man-to-man communication. Yet, there exists each other part of communication known as man-to-machine communication. In this time of innovation an individual needs to collaborate with an enormous number of customer emotionally supportive networks to complete least complex of tasks. This is the reason why humans want to have speech as communication/interaction medium with computers as well. e- Narrator is a free website that automatically speaks any portable document format (PDF) and Image which you upload. Its main aim is to help patients suffering from visual impairment. The main idea of this project is to recognize the text character and convert it into speech signal. The text contained in the page is first pre-processed. The pre-processing module prepares the text for recognition. Then the text is segmented to separate the character from each other. Segmentation is followed by extraction of letters and resizing them and stores them in the text file. These processes are done with the help of Optical Character Recognition (OCR). This text is then converted into speech.

II. IMAGE TO TEXT:

Text recognition is the process of detecting and recognizing text in images and video streams. The recognizer decides the actual text in each block once it has been identified and fragments it into lines and phrases. The Text API senses text in real-time on the computer in English language.

The Text Recognizer segments text into blocks, lines, and words.

A. Roughly speaking:

- 1) A Block is a contiguous set of text lines, such as a paragraph or column.

- 2) A Line is a contiguous set of words on the same vertical axis.
- 3) A Word is a contiguous set of alphanumeric characters on the same vertical axis.

In order to improve recognition performance, IIT includes preprocessing of objects. Preprocessing involves many methods such as De-skew, Despeckle, binarization, row deletion, study of the format, etc. Binarization involves transforming an object into black and white from a blue or gray scale. It is done to isolate the text from the background Fixed-pitch fonts are segmented by aligning the image to a fixed grid depending on where vertical grid lines cross black areas least frequently.

A graded list of candidate characters is generated by the key process, character recognition. It consists of two methods that are known as matching matrix and extraction of features. Matrix matching also known as pattern recognition or image correlation involves pixel-by pixel-based comparison of an image to a stored glyph. While the abstraction of the function breaks down glyphs into attributes such as columns, closed loops, line orientation and line intersections. These are similar to a character's abstract vector image, which could be simplified to one or more glyph models.

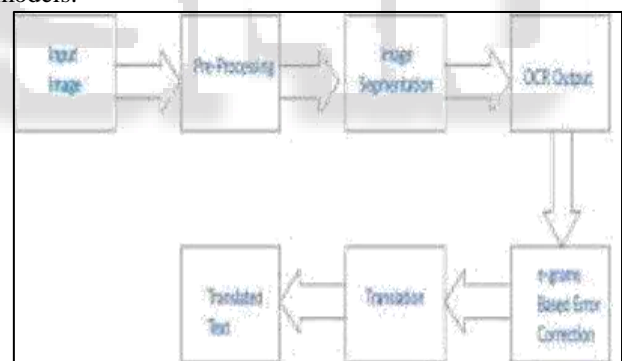


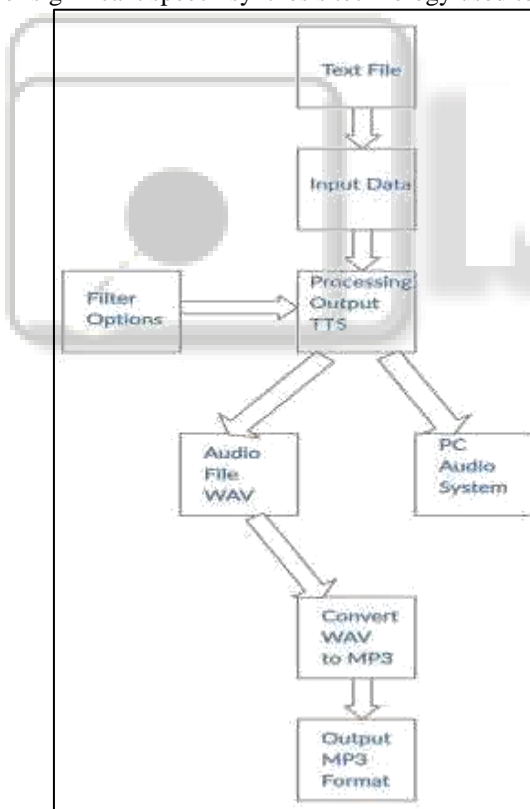
Fig. 1: Image to Text Conversion

III. TEXT TO SPEECH

TTS is a type of speech synthesis for Text-to-Speech (also written as Text-to-Speech), which transforms text into voice production. Text-to-speech program basically takes the text you compose and converts it into speech files you can use. There are many ways to create text-based sound. There is a Unit Selection Synthesis (USS) method. The process starts at both ends, creating voice server word text storage which meets to generate speech in the center. Unified Messaging Technology, primarily email interpretation technology, virtual voice reply systems, plays an important role in full or partial customer support programs, examples of TTS device banking implementations. Sadly, this ability is robbed of some of the educationally deficient, visually impaired, an illiterate people. A system is designed to address this issue by saying the document in the language in which such people

can understand it. This is done through a voice-related text framework.

Next, they need to understand the natural sound of humans in order to create a full text-to-speech system. For this purpose, a voice actor (with a great sounding voice) who is fluent in certain language is chosen to speak the entire sentences or paragraphs. This is recorded and a database of voices is created. It builds a server of thousands of recorded sound files. Such records need to be categorized and organized; the units of expression are marked and sections are classified by phonemes, syllables, morphemes, terms, phrases and sentences. There is one device devoted to the analysis of natural language. Text is condensed and broken into phonetic tone. Then a set of analyzes were made to understand the meaning of sentences and to assess the context of the word for pronunciation. Such two systems are now combining to generate audio, namely the voice registry and the natural language processing. This is achieved as follows: the voice inventory is scanned once the NLP is full and the speech units are chosen which best match together to produce the sounds associated with the provided document. This whole process is called the Synthesis of System Selection (USS). We use USS because the most natural sounding speech is known to be made. HTS (HTS-based speech synthesis system) is the other significant speech synthesis technology used today.



IV. RELATED WORKS

Texttospeech.io can only convert image not PDF its user interface is complex and cannot easy for user to use.[5]

I2S OCR's major drawback is it does recognize text very well in English and it does read it out loud. However, there is no control over the speech. There is no stop or pause.[11]

Naturalreaders.com's drawback is its several features and several languages are paid. After 20 days of free trail we have to buy premium membership.[10]

V. METHODOLOGY

The main idea of this project is to recognize the text character and convert it into speech signal. Method which is implemented in e-Narrator is mentioned below step by step The text contained in the page is first pre-processed then The pre-processing module prepares the text for recognition. After text recognition the text is segmented to separate the character from each other. Segmentation is followed by extraction of letters and resizing them and stores them in the text file. These processes are done with the help of Optical Character Recognition (OCR). This text is then converted into speech.

VI. RESULT

e-Narrator is a free platform which converts portable t (PDF) and Images is converted into Speech and Narration. e-Narrator provides several languages as well as several accents in which narration or converted speech is converted to help user for better understanding. When Image or PDF is uploaded by user first text is segmented into words and it then it is converted into text with the help of image processing. After this conversion text is converted into speech with the help of text processing. Conversion of image to text and then text to speech is done with the help of Optical Character Recognition Engine (OCR Engine).

VII. PROPOSED SYSTEM

Our development concept takes the form of Portable Document Format and Photo as input. For further storage purposes, this data is processed. Convert the voice or picture output to text format utilizing speech-to-text and photo-to-text conversion is the main processing stage. When we obtain the full text format data, it will be evaluated and interpreted as follows. Second, the text is structured in the style of the destination to adapt it to the use. In some cases, we also need to separate keywords from the normal text. Such pre-processed text is supplied to the device processing unit required to perform its actual use as output. The output which will be generated from these various system components will again be converted into speech signal as the final output using text-to-speech conversion.

VIII. CONCLUSION

In this paper we are proposed a system which has demonstrated a website which helps users to convert Image or portable document format into speech or narration. The main idea behind this project is to recognize the text character and convert it into speech signal. The text contained in the page is first pre-processed. The pre-processing module prepares the text for recognition. Then the text is segmented to separate the character from each other. Segmentation is followed by extraction of letters and resizing them and stores them in the text file. These processes are done with the help of Optical Character Recognition (OCR). This text is then converted into speech for visually impaired people.

REFERENCES

- [1] Yogita H. Ghadage, Sushama D. Shelke, "Speech to text conversion for multilingual languages", Communication and Signal Processing (ICCSP), 2016 International Conference on 6-8 April 2016
- [2] Akhil S. Deshpande¹, Shreyas S. Vaidya, Pravin B. Swami³, Pavan R. Jaiswal⁴ "Integration of speech, Image and text processing technologies", International Research Journal of Engineering and Technology (IRJET), Volume: 04 Issue: 04 Apr -2017
- [3] Christos Liambas, Miltiadis Saratzidis, "Autonomous OCR dictating system for blind people", Global Humanitarian Technology Conference (GHTC), 2016
- [4] Asha G. Hagargund, Sharsha Vanria Thota, Mitadru Bera, Eram Fatima Shaik, "Image to Speech Conversion for Visually Impaired", International Journal of Latest Research in Engineering and Technology (IJLRET), 2017
- [5] <https://texttospeech.io/online-image-to-text-reader>
- [6] https://en.wikipedia.org/wiki/Optical_character_recognition
- [7] <http://www.codeguru.com/cpp/gm/multimedia/audio/article.php/c12363/HowSpeech-Recognition-Works.htm>
- [8] https://en.wikipedia.org/wiki/Image_translation
- [9] <https://smallseotools.com/image-to-text-converter/>
- [10] <https://www.naturalreaders.com/online/>
- [11] <https://i2s.symisc.net/>

