IMAGE TEXT TO VOICE CONVERTER WITH SENTIMENT ANALYSIS

[1]Prof. Vivek Pandey, [2]Shani Kumar Maurya, [3]Sonam Gupta, [4]ShaikhUsman Gani
[1]Assistant Professor, [2]Student, [3]Student, [4]Student
B.E. Computer Engineering, Allamuri Ratnamala Institute of Engineering and Technology
Shahapur, Thane- 421601 India

Received: February 01, 2019 Accepted: March 22, 2019

ABSTRACT: *In this paper, we proposed text to speech converter for blind people with sentiment analysis. The paper is based on Android domain along with machine learning. Reading text from text image and text board is difficult task for blind people. Visual disability is one of the biggest limitation for humanity, especially in this day and age when information is communicated a lot by text messages (electronic and paper based) rather than voice. In this work, an approach has been attempted to extract and recognize text from image (i.e. captures image that contains only text) and convert that recognized text into speech with sentiment analysis. It is implemented using an Android App and OCR (Optical Character Recognition) algorithm and uses the application of machine learning for sentiment analysis. The captured image undergoes a series of image pre-processing steps to locate only that part of the image that contains the text and removes the background. There are tools used for convert the new image (which contains only the text) to speech. There are OCR (Optical Character Recognition) software and TTS (Text-to-Speech) engines and an algorithms to perform sentiment analysis i.e. Convolution Network, SVM (Support Vector Machine). The audio output is listen through the mobile phone's audio jack using speakers or earphones.*

Key Words*: Android, OCR, Text Translator, TTS, Sentiment Analysis, visually impaired person.*

I. INTRODUCTION

Every year, the number of visually challenged persons are increasing due to eye diseases, age related causes, traffic accidents and other causes. As reading is one of major importance in the daily routine (text being present everywhere from newspapers, commercial products, signboards, digital screens etc.) of humankind, visually impaired people face many difficulties.Mobile applications that give support to the visually challenged persons for reading out the text.The focus of our research is that the visually challenged person can get information about printed text, text boards, scene text, hoardings, and instructions on traffic signboards in audio form.The application design for a camera based reading system that extract text from image and identify the text characters and strings from the captured image and finally text will be converted into audio.

The captured image undergoes a series of image pre-processing steps to locate only that part of the image that contains the text and removes the background. The image is processed by the OCR and TTS. OCR has become most successful applications of technology in the field of text recognition and artificial intelligence. 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[1].Speech synthesis is the artificial synthesis of human speech. A Text-To- Speech (TTS) synthesizer isa android-based application that should be able to read any text aloud, whether it was directly introduced in the android application by an operator or scanned and submitted to an OCR system[2].

Operational stages of the system consist of image capture, image preprocessing, image filtering, character recognition, text to speech conversion and sentiment analysis. Sentiment analysis is done by using an algorithms i.e. Convolution Network, SVM (Support Vector Machine). The software platforms used are OCR Engine, TTS Software, android platform and application of machine learning.

II. LITRETURE REVIEW

In this section, we present some previous research works for assisting visually challenged people with text to speech technology. Previous paper proposes a system, which is used for converting the input string of text into the corresponding speech using Raspberry-pi [3]. Limited and proper combination of words with grammar rules gives a clear picture of the ideas or thoughts that speaker wants to convey [4]. The system includes Python coding which is done on Raspberry-pi for the generation of speech signal based on the user defined input text [4]. A simple File Accessing Protocol (FAP) is adapted to achieve the task of retrieving the

audio files on the database. This paper also throws light on the mechanism and strategy used to convert the text input to speech output [5].

III. EXISTING SYSTEM

The devices are exist for blind people, which convert the text to voice. Take the input in an image form and output gives in voice. That device consist of raspberry pi board, which controls the peripherals like Camera, Speaker, and Display. That works based on OCR algorithm. In this project setup, the camera is mounted on a stand in such a position that if a paper is placed in between the area marked by angular braces it captures a full view of paper. The content on the paper should written in English preferably Times New Roman and good font size (24 or more as per MS Word).This device is hardware-based device.

The existing technologies also use a similar approach as mentioned in this report, but they have certain drawbacks. Firstly, the input images taken in previous works have no complex background, i.e. the test inputs are printed on a plain white sheet. It is easy to convert such images to text without pre-processing, but such an approach will not be useful in a real-time system [6]. Also, in methods that use segmentation of characters for recognition, the characters will be read out as individual letter and not a complete word. This gives an undesirable audio output to the user.

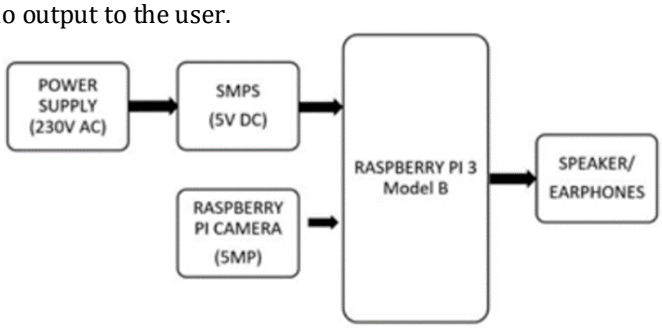

Fig: Existing System Architecture

IV. PROPOSED METHODOLOGY

There are various text to speech systems are discussed for visually challenged persons but there exits some limitations. The objective of this work is

1. Image to text convert
2. Text to voice convert

The device mainly consists of:

OCR (Optical Character Recognition)

TTS (Text To Speech)

Sentiment Analyzer

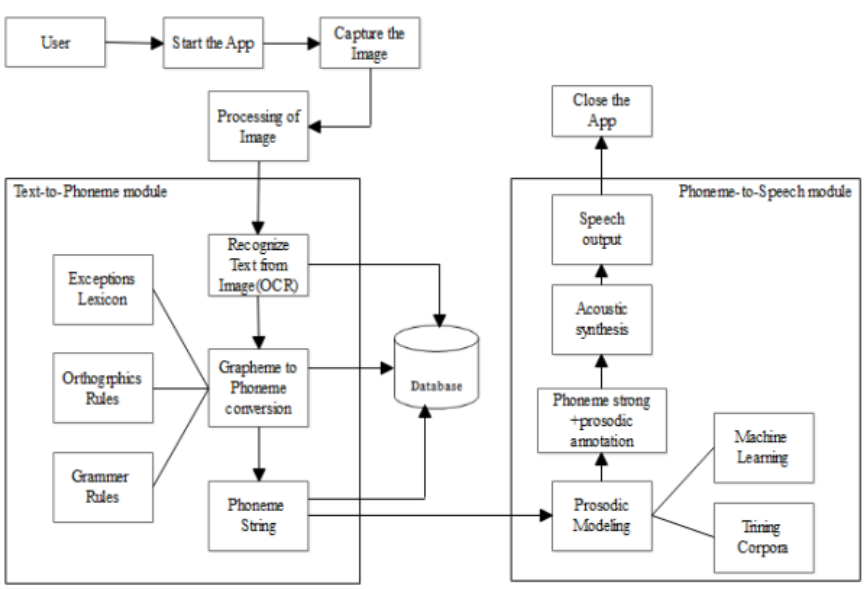

Fig: System Architecture

4.1. Processing of image

In this step, the inbuilt camera captures the images of the text. The quality of the image captured depends on the mobile phone's camera used. Normally camera required is 5MP camera with a resolution of 2592x1944.This step consists of color to gray scale conversion, edge detection, noise removal, warping and cropping and thresholding. The image is converted to gray scale, as many OpenCV functions require the input parameter as a gray scale image. Noise removal is done using bilateral filter. Canny edge detection is performed on the gray scale image for better detection of the contours. The warping and cropping of the image are performed according to the contours. This enables us to detect and extract only that region which contains text and removes the unwanted background. In the end, Thresholding is done so that the image looks like a scanned document. This is done to allow the OCR to efficiently convert the image to text.

4.2. Recognize text from image (OCR)

The extraction of the text in the image is done using optical character recognition (OCR). OCR is a field of research in pattern recognition, artificial intelligence and computer vision [7]. It is the conversion of the images of typed, handwritten or printed text into a digital text or computer format text. Earlier OCR versions had to be trained in each character of a text with its specific font. Today, advanced OCRs are available that have a high degree of accuracy, support a wide variety of image formats, languages and fonts. For our project, we have used Tesseract OCR. It is the most accurate open source OCR engine and is powered by google. It can be used on the Linux, mac and windows platform. The newest Tesseract version, 3.4 supports a hundred languages. However, images must undergo a number of pre-processing stages like noise removal, scaling etc. otherwise the output will be of low quality.

4.3. Grapheme to Phoneme Conversion

Grapheme to phoneme is the main part of Text To Speech (TTS). TTS takes the text as an input and the output will be the voice which is synthesis voice. Grapheme is way of writing down phoneme, Phoneme is the smallest unit of sound phoneme can be put together to make word. Grapheme does the arrangement of the text that how it is pronounced. And these arrangements know as phoneme, which is in synthesis voice. Example of the G2P is if there is a text "cat" for correct, pronounce it will be write as "k ae t". Theanother example is the text "Through" become "th r oo" where the sound /oo/ is represented by the letters 'o u g h'.

4.4. Phoneme String:

The problem of translating text to speech is usually approached at the phoneme level using a rule-based system [8]. A rule based system specifies how each letter, or group of letters, will be translated into a basic unit of sound, a phoneme [8]. These rules are very context sensitive, depending on possibly lengthy left and right contexts[8].

4.5. Prosodic modelling:

In this, one key module in a text-to-Speech system is prosody modeling. Prosody refers to duration, intonation and intensity patterns of speech associated to the sequence of syllables, words and phrases. A good prosody model should capture the duration, intonation and intensity patterns of natural speech. In complete prosody generation model, the quantities like phrasing, stress and the like are determined to generate naturalness bearing synthetic voice.Pragmatics examines the distinction between the literal meaning of a sentence and the meaning intended by the speaker [9]. Prosody can have the effect of changing the meaning of a sentence by indicating a speaker's attitude to what is being said (e.g. it can indicate irony, sarcasm, etc.) particularly when prosody works in conjunction with the social/situational context of an utterance [10]. Prosodic modeling makes the speech more understandable.

4.6. Acoustic Sound:

An acoustic model is used in automatic speech recognition to represent the relationship between an audio signal and the phonemes or other linguistic units that make up speech [11]. The application is learned from a set of audio recordings and their corresponding transcripts[11]**.**An acoustic model is created by taking a large database of speech and using special training algorithms to create statistical representations for each phoneme in a language [12].Acoustic model make the sound more understandable.

V. CONCLUSION

In this paper, we discussed about the Text-to-Speech system (TTS).The text to speech conversion may seem effective and efficient to its users if it produces natural speech and by making several modifications to it. This system will be helpful for bind persons to access information in written form and in the surrounding. It is useful to understand the written text messages, warnings, and traffic direction in voice form by converting it from Text to Speech. This paper made a clear and simple overview of working of text to speech system (TTS) in step by step process. The user capture the image and the system reads it from the database or data

store where the words, phones, diaphones, triphone are stored then text is convert into speech with semantic analysis.The speech output originating from the system can be manipulated as per the user's objectives. This portable device, does not require internet connection, and can be used independently by people.

VI. ACKNOWLEDGMENT

This research was supported by Alamuri Ratnamala Institute of Engineering and technology. We thank our guide Prof. Vivek Pandey who provided insight and expertise that greatly assisted the research.

VII. REFERENCES:

1. https://www.scribd.com/document/325331905/kh
2. https://www.researchgate.net/publication/232641635_High_quality_text-tospeech_synthesis_A_comparison_ of_four_candidate_algorithms
3. https://www.semanticscholar.org/paper/Text-to-Speech-Conversion-Using-Raspberry-Pi-for-Reddy/dc581b fbd201cc3f0527b417d5711d402ea50130
4. http://www.ijirset.com/upload/november/19_2012_ece_1_SET_ISS%201.pdf
5. http://www.ijirset.com/upload/november/19.html
6. http://ijlret.com/Papers/Vol-3-issue-6/2-B2017160.pdf
7. https://en.wikipedia.org/wiki/Optical_character_recognition
8. https://link.springer.com/chapter/10.1007/BFb0038474
9. http://clas.mq.edu.au/speech/phonetics/phonology/intonation/prosody.html
10. https://www.academia.edu/9578800/Pragmatics_Methodology_Course
11. https://www.youtube.com/watch?v=5ktDTa8glaA
12. http://www.voxforge.org/home/docs/faq/faq/what-is-an-acoustic-model