International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 07 Issue: 07 | July 2020 www.irjet.net p-ISSN: 2395-0072

Text to Speech Conversion using Optical Character Recognition

Sharvari S[1], Usha A[2], Karthik P[3], Mohan Babu C[4]

[1,2,3]*Student, Dept. of Telecommunication Engineering, SJC Institute of Technology, Karnataka, India*
[4]*Professor, Dept. of Electronics and Communication Engineering, SJC Institute of Technology, Karnataka, India*

Abstract - *In present situation, communication plays a vital role in the world. Transferring on information, to the correct person, and in the proper manner is very important on a personal and professional level. The world is moving towards digitization, so are the means of communication are Phone calls, emails, text messages etc. have become a major part of message conveyance in this digital world. In order to serve the purpose of effective communication between two parties without any delay, many applications have come to existence, which acts as a mediator and help in effectively carrying messages in from text to the speech signals over miles of networks. The main purpose of this project is to overcome the problems facing by the blind people and illiterates. Because the blind people and illiterates can be easily manipulated, this leads to misuse. To overcome this problem we are proposing a device which helps in conversion of hard copy of text which is inserted into the device will be converted to speech. Most of these applications find the use of functions such as articulators, conversion from text to synthetic speech signals, language translation amongst various others. In this project, we'll be executing different techniques and algorithms that are applied to achieve the concept of Text to Speech (TST).*

Key Words: **Digitization, Conversion of hardcopy of text to speech, Synthetic speech signals, Language translation**

1. INTRODUCTION

From the past few years, Mobile Phones have become a main source of communication for this digitalized society. We can make calls and text messages from a source to a destination easily. It is known that verbal communication is the most appropriate modem of passing on and conceiving the correct information. To help the people more effectively, engage with local and/or remote services text-to-speech (TTS) were first developed to aid the visually impaired by offering a computer-generated spoken voice that would "read" text to the user. In this project, we will take a look at text to speech conversion. Using Optical Character Recognition

Optical Character Recognition can be used widely in healthcare applications to aid blind people. We have proposed a system which is used to help blind people to read books, understand different text on banners, pictures and large ad boards. OCR includes mainly three components the camera to capture the images, the programmable system to convert the captured camera into whatever format we want and finally the output system to show the output of OCR

We are using raspberry pi a compact size and low weight system for creating an OCR system. The raspberry pi is connected with the help of Ethernet cable to the computer so that we can get the display. We have proposed capturing of image through raspberry pi camera or through android camera which can be connected with Ethernet, Bluetooth or Wi-Fi. The webcam or raspberry pi camera capture the images large font size image and store it into raspberry pi system. We are using tesseract OCR library provided by python to convert the captured image to the text

The Display of Raspberry pi is shown on the computer display with the help of putty software and Ethernet cable. The camera is connected to the raspberry pi system which works as an eye for OCR system. The camera usually takes 5-7 minutes to capture and process the images. With the help of tesseract OCR library of python, the captured image is converted to the text format in the raspberry pi hard disk location. The converted text is provided to TTS system which converts the text to the voice format

1.1 OBJECTIVE

- To enhance the speech recognition and conversion of text to speech.
- To create binary image for the image recognition.
- To create binary image for the image recognition through image processing.
- Strengthening the audio output.
- Establish the common connection between the speech and text recognition.

1.2 PROBLEM STATEMENT

- The blind people and illiterates are facing difficulty in understanding the content they have.
- This leads to manipulations and scams.

1.3 METHODOLOGY

The proposed system consists of two main module the image processing and voice processing module. The image processing module captures image using camera converting the image into text. Voice processing module changes the text into sound and processes it with specific characteristics so that the sound can be understood.

At first the image processing module, where OCR converts .jpg to .txt form 2nd is voice processing module which converts .txt to speech OCR is very important element in this

module OCR or Optical Character Recognition is a technology that automatically recognizes the character through the optical mechanism, this technology intimates the ability of the human sense of sight. Before feeding the image to OCR it is converted to binary image to increase the image recognition accuracy.

Image binary conversion is done by Image magic software, which is another open source tool for image manipulation. The output of OCR is the text, which is stored in a file (speech.txt). It needs some supporting condition in order get the minimal defect. Voice processing module changes the text into sound and processes it with specific characteristics so that the sound can be understood.

Image processing

A pixel is a single dot of a particular colour. An image is essentially a collection of pixels. More the pixels in an image, the higher will be its resolution. A computer doesn't know that an image of a signpost is really a signpost , it just knows that the first pixel is this colour, the next pixel is that colour, and displays all of its pixels for you to see.

Step 1: Pre-Processing the Image

Before text can be pulled, the image needs to be massaged in certain ways to make extraction easier and more likely to succeed. This is called pre-processing, and different software solutions use different combinations of techniques.

The more common pre-processing techniques we are using are:

Binarization

Every single pixel in the image is converted to either black or white. The goal is to make clear which pixels belong to text and which pixels belong to the background, which speeds up the actual OCR process

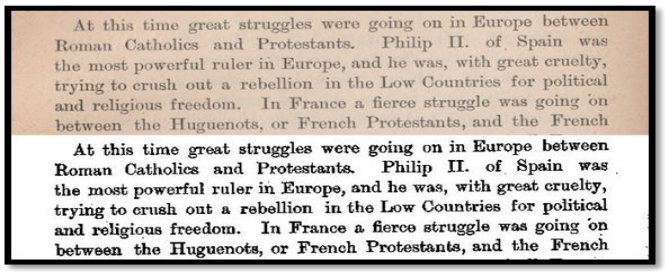

Fig -1: Binarization of Text

Deskew

Since documents are rarely scanned with perfect alignment, characters may end up slanted or even upside-down. The goal here is to identify horizontal text lines and then rotate the image so that those lines are actually horizontal.

Despeckle

Whether the image has been binarized or not, there may be noise that can interfere with the identification of characters. Despeckling gets rid of that noise and tries to smooth out the image.

Line Removal

Identifies all lines and markings that likely aren't characters, then removes them so the actual OCR process doesn't get confused. It's especially important when scanning documents with tables and boxes.

Zoning

Separates the image into distinct chunks of text, such as identifying columns in multi-column documents.

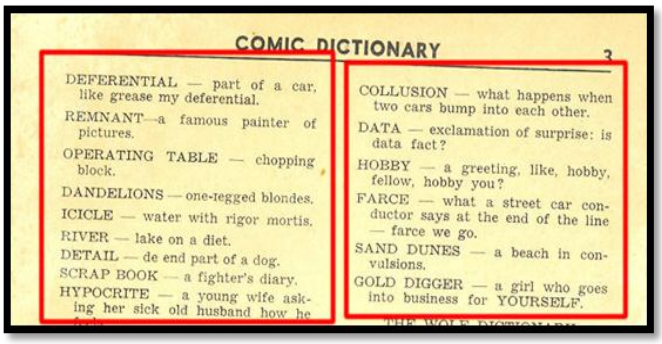

Fig -2: Zoning Process

Step 2: Processing the Image

First things first, the OCR process tries to establish the baseline for every line of text in the image (or if it was zoned in pre-processing, it will work through each zone one at a time). Each identified line of characters is handled one by one.

For each line of characters, the OCR software identifies the spacing between characters by looking for vertical lines of non-text pixels (which should be obvious with proper binarization). Each chunk of pixels between these non-text lines is marked as a "token" that represents one character. Hence, this step is called tokenization

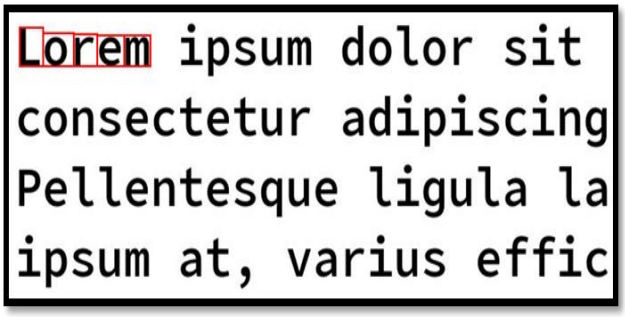

Fig -3: Tokenization

Once all of the potential characters in the image are tokenized, the OCR software can use two different techniques to identify what characters those tokens actually are: The tokens and glyphs need to be of similar size or else none of them will match.

Pattern Recognition

Each token is compared pixel-to-pixel against an entire set of known glyphs—including numbers, punctuation, and other special symbols—and the closest match is picked. This technique is also known as matrix matching.

There are several drawbacks here. First Second, the tokens need to be in a similar font as the glyphs, which rules out handwriting. But if the token's font is known, pattern recognition can be fast and accurate.

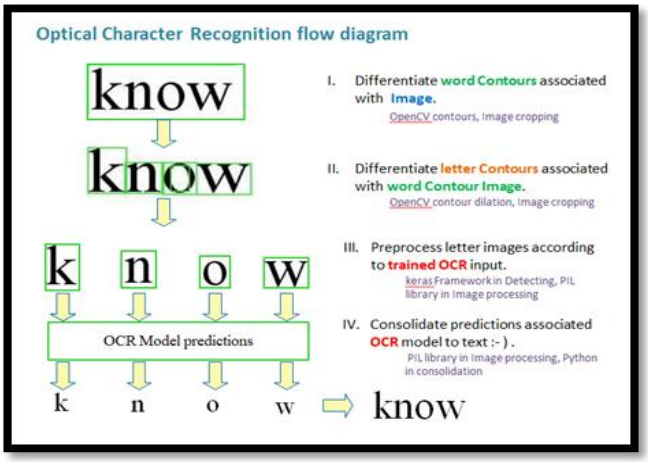

Fig -4: Pattern Recognition

Feature Extraction

Each token is compared against different rules that describe what kind of character it might be. For. example, two equal-height vertical lines connected by a single horizontal line are likely to be a capital H

Step 3: Post-Processing the Image Lexical Restriction

All words are compared against a lexicon of approved words, and any that don't match are replaced with the closest fitting word. A dictionary is one example of a lexicon. This can help correct words with erroneous characters, like "thorn" instead of "thorn".

B. Voice processing:

In this method our main aim to convert obtained text to speech with the help of coding in raspberry pi using text to speech (TTS) synthesizer. The text to speech synthesizer is installed in raspberry pi. The output can be listened through audio speakers.

Fig -5: Voice Processing

2. BLOCK DIAGRAM AND DESCRIPTION

Block Diagram

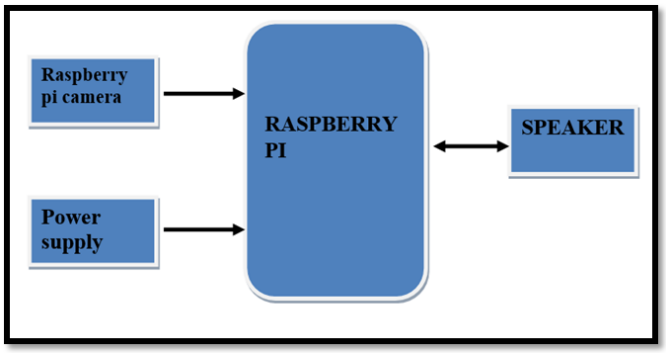

Fig -6: Block Diagram of Proposed Project

Description

Image processing: In the first step the device is moved over the printed page and the inbuilt camera captures the images of the text. The quality of the image captured will be so high so as to have fast and clear recognition due to the high resolution camera. Letters will be extracted and converted into digital form. It consists of three steps: Skew Correction, Linearization and Noise Removal. The captured image is checked for skewing. There are possibilities of image getting skewed with either left or right orientation. Here the image is first brightened and binarized.

The function for skew detection checks for an angle of orientation and if detected then a simple image rotation is carried out till the lines match with the horizontal axis, which produced a skew corrected image. The noise introduced during capturing or of sequence of characters into sub image of individual symbol (due to poor-quality of the page has to be cleared for further processing. Segmentation: This operation seeks to decompose an image characters). The binarized image is checked for interline spaces. If inter line spaces are detected then the image is segmented into sets of paragraphs across the interline gap.

The lines in the paragraphs are scanned for horizontal space intersection with respect to the background. Histogram of the image issued to detect the width of the horizontal lines. Then the lines are scanned vertically for vertical space intersection. Here histograms are used to detect the width of the words. Then the words are decomposed into character width computation.

Feature Extraction: In this stage we gather the essential features of the image called feature maps. One such method is to detect the edge in the image, as they will contain the required text. Tesseract: This software is used to convert the image file to text file by extracting the texts from the image and storing it in the file with.txt extension.

Flowchart

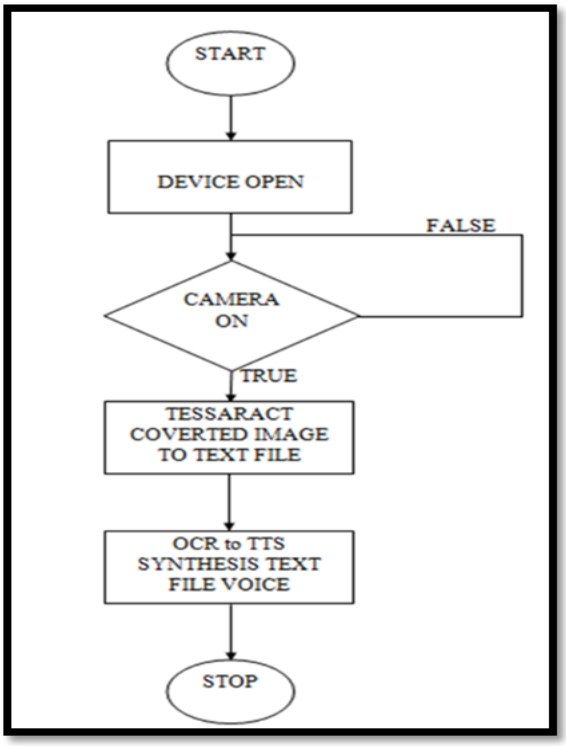

Fig -7: Flowchart

Flowchart Description

- Device boot up by power on the device, it takes command form user, raspberry pi cam with Sony Imax image sensor it takes image, before taking image user as to focus the image so that image get good.
- Optical character recognition is method where image to text by using Tesseract. Once camera condition gets true .camera takes noise less image with high resolution image than OCR engine convert image into text.
- This text. File is converted into voice by TTS engine, again processing continues.

3. HARDWARE AND SOFTWARE REQUIREMENTS

Hardware Requirements:

- Raspberry Pi 3 Model B+
- Speaker
- Raspberry Pi Camera Module V2

Software Requirements:

- Python 3.7.1
- Raspbian OS

4. HARDWARE AND SOFTWARE IMPLEMENTATION

SYSTEM DESIGN AND CONTROL FLOW

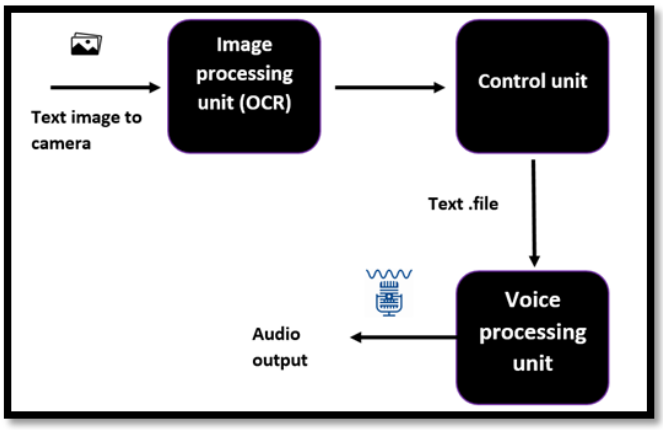

Fig -8: Control Flow

Feature Extraction

In this stage we gather the essential features of the image called feature maps. One such method is to detect the edges in the image, as they will contain the required text. For this we can use various axes detecting techniques like: Sobel, Kirsch, Canny, Prewitt etc. The most accurate in finding the four directional axes: horizontal, vertical, right diagonal and left diagonal is the Kirsch detector. This technique uses the eight point neighborhood of each pixel.

Optical Character Recognition

Optical character recognition, usually abbreviated to OCR, is the mechanical or electronic conversion of scanned images of handwritten, typewritten or printed text into machine encoded text. It is widely used as a form of data entry from some sort of original paper data source, whether documents, sales receipts, mail, or any number of printed records. It is crucial to the computerization of printed texts so that they can be electronically searched, stored more compactly, displayed on-line and used in machine processes such as machine translation, text-to- speech and text mining. OCR is a field of research in pattern recognition, artificial intelligence and computer vision.

Tesseract

Tesseract is a free software optical character recognition engine for various operating systems. Tesseract is considered as one of the most accurate free software OCR engines currently available. It is available for Linux, Windows and Mac OS. An image with the text is given as input to the Tesseract engine that is command based tool. Then it is processed by Tesseract command. Tesseract command takes two arguments: First argument is image file name that contains text and second argument is output text file in which, extracted text is stored. The output file extension is given as .txt by Tesseract, so no need to specify the file extension while specifying the output file name as a second argument in Tesseract command.

After processing is completed, the content of the output is present in .txt file. In simple images with or without colour (grey scale), Tesseract provides results with 100% accuracy. But in the case of some complex images Tesseract provides better accuracy results if the images are in the gray scale mode as compared to colour images. Although Tesseract is command-based tool but as it is open source and it is available in the form of Dynamic Link Library, it can be easily made available in graphics mode.

Text To Speech

A text to speech (TTS) synthesizer is a computer based system that can read text aloud automatically, regardless of whether the text is introduced by a computer input stream or a scanned input submitted to an Optical character recognition (OCR) engine. A speech synthesizer can be implemented by both hardware and software. Speech is often based on concatenation of natural speech i.e. units that are taken from natural speech put together to form a word or sentence.

The output observed for image text to speech conversion is

- Image is captured with The Camera – The captured image is converted to the text and saved at same location of image.
- It takes approximately 7-8 sec to convert the text.
- The converted text is processed in TTS.
- The speech is obtained as an output in Headphones or speaker.
- It recognizes numbers as well as letter in English.
- Range of reading Distance is 30cm.

CIRCUIT DESGIN AND INTERFACE

We are using raspberry pi a compact size and low weight system for creating an OCR system. The raspberry pi is connected with the help of Ethernet cable to the computer so that we can get the display. We have proposed capturing of image through raspberry pi camera or through android camera which can be connected USB.`

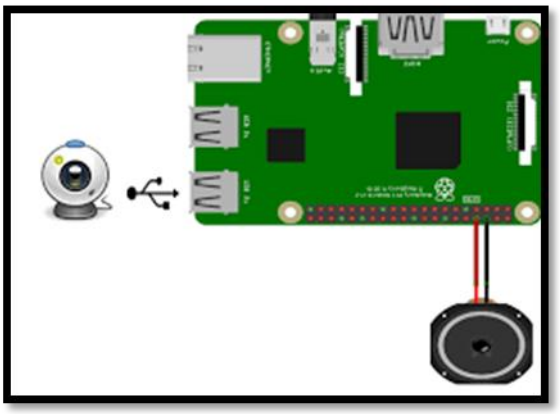

Fig -9: Interfacing Camera with Pi

Webcam or raspberry pi camera capture the images large font size image and store it into raspberry pi system. We are using tesseract OCR library provided by python to convert the captured image to the text. The image text will be saved in .txt file or as per the user requirement. The camera and tesseract takes 5-7 minutes to capture and convert the image to text.

Raspberry Pi OCR consists of two main methods, the Image Processing method and voice processing method. The image processing method includes the image capturing and image to text conversion. Voice processing method includes captured image to speech conversion. The speech conversion can be optimized i.e. the voice can be changed to male or female voice so that it can be understood by the specific user. We can also control the echo for the voice generated from voice processing method.

Open source OCR software called Tesseract used as a basis for the implementation of text reading system for visually disabled in Android platform. Google is currently developing the project and sponsors the open development project. Today, Tesseract is considered the most accurate free OCR engine in existence.

When the OCR process is complete it produces a returns a string of text which is displayed on the user interface screen, where the user is also allowed to edit the text then using the TTS API enables our Android device to speak text of different languages. The TTS engine that ships with the Android platform supports a number of languages: English, French, German, Italian and Spanish. Also American and British accents for English are both supported. The TTS engine needs to know which language to speak. So the voice and dictionary are language specific resources that need to be loaded before the engine can start to speak

SOFTWARE ANALYSIS

The display of raspberry pi is shown on the computer display with the help of putty software and Ethernet cable.

VNC (virtual network computing)

VNC is platform-independent – there are clients and servers for many GUI-based operating systems and for Java. Multiple clients may connect to a VNC server at the same time. Popular uses for this technology include remote technical support and accessing files on one's work computer from one's home computer, or vice versa.

VNC was originally developed at the Olivetti & Oracle Research Lab in Cambridge, United Kingdom. The original VNC source code and many modern derivatives are open source under the GNU General Public License.

VNC may be tunneled over an SSH or VPN connection which would add an extra security layer with stronger encryption. SSH clients are available for most platforms; SSH tunnels can be created from UNIX clients, Microsoft Windows clients, Macintosh clients (including Mac OS X and System 7 and up) – and many others. There are also freeware applications that create instant VPN tunnels between computers.

- Direct connections are quick and simple providing you're joined to the same private local network as your Raspberry Pi. For example, this might be a wired or wireless network at home, at school, or in the office).
- On your Raspberry Pi (using a terminal window or via SSH) use these instructions or run IP Config to discover your private IP address.
- On the device you'll use to take control, download VNC Viewer. For best results, use the compatible app from Real VNC.
- Enter your Raspberry Pi's private IP address into VNC Viewer:
- Cloud connections are convenient and encrypted end-to-end. They are highly recommended for connecting to your Raspberry Pi over the internet.

The Raspberry Pi is mostly a Linux system. The most popular Raspberry Pi operating system is Raspbian, which the Foundation officially recommends.

Raspbian built on top of Debian Linux. It comes with standard Linux software including a web browser, mail client, and various system tools, as well as programming tools for Java and Python, and a program called Scratch.

The camera is connected to the raspberry pi system which works as an eye for OCR system. The camera usually takes 5-7 minutes to capture and process the images. With the help of tesseract OCR library of python, the captured image is converted to the text format in the raspberry pi hard disk location. The converted text is provided to TTS system which converts the text to the voice format.

5. APPLICATIONS AND ADVANTAGES

ADVANTAGES

- Improved word recognition skills, fluency and accuracy.
- Global market penetration.
- It is very reliable and user friendly.
- Enhanced employee performance.
- Extend the reach of your content.
- People with different learning styles.
- It is very useful for illiterates.

APPLICATIONS

- Converts text into spoken voice output.
- It is mainly designed for blind people.
- It helps small kinds to improve word recognition, remember information while reading.
- We can give different language access for those who are not able to read and write.
- Impact of Text to Speech on Tourism.
- It can be used as announcing device in schools/colleges and in other places.

6. OUTPUT AND EXPERIMENTAL RESULTS

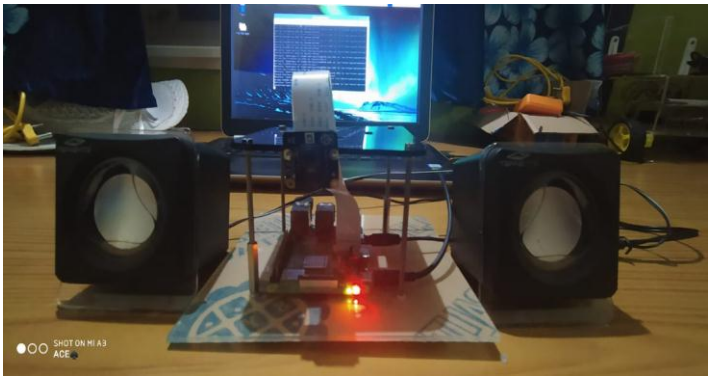

Fig -10: Raspberry pi interfaced with Camera Module

Here we are using Tesseract for converting image into text file and TTS for text file into voice processing

1. The first image says that we are capture image by focusing the image.
2. Second image says the device converted image file into text file.
3. Third image says about the processing the image file to text. File.
4. Forth image say voice processing of the text file.

CONCLUSIONS

This research work has provided numerical data, which can serve as an essential proof to the theory that the integrity of the position of syllables should be maintained during concatenation in speech synthesis. Naturalness of TTS system is based on how properly segmentation units are formed and after synthesizing different words how carefully the concatenation joint distortion is reduced. In this research work implementation of this system, visually impaired persons can easily listen to the text of the document. Through translation tools, one convert the text to the desired language and then again by using the TTS speech recognition tool can convert that changed text into audio. This idea will tend to a new technology that involves neuron based visual system which can give information directly to damaged neurons in mind in disabled people.

REFERENCES

[1] AyushiTrivedi, NavyaPant, Pinal Shah, SimranSonik and Supriya Agrawal,"Speech to text and text to speech recognition systems" , IOSR Journal of Computer Engineering (IOSR-JCE) e-ISSN: 2278-0661, p-ISSN: 2278-8727, Vol-20, Issue-2, Ver-I, March.April-2018.

[2] Hussain Rangoonwala, Vishal Kaushik, P Mohith and DhanalakshmiSamiappan "Text to speech module International Journal of Pure and Applied Mathematics" Vol-115 No. 6 2017.

[3] S. Venkateswarlu1, D. B. K. Kamesh1, J. K. R. Sastry and Radhika Rani , "Text to speech conversion", Indian Journal of Science and Technology", Vol-9(38), October-2016

[4] Gordana Laštovička-Medin, Itana Bubanja, "Hardware approach of text-to-speech in embedded applications" Kaladharan N, An english text to speech conversion system,International Journal of Advanced Research in Computer Science and Software Engineering,, Vol-5, Issue-10, October-2015

[5] Poonam.S.Shetake, S.A.Patil and P.M Jadhav "Review of text to speech conversion methods", International Journal of Industrial Electronics and Electrical Engineering, Vol-2, Issue-8, August-2014

[6] Kaveri Kamble and Ramesh Kagalkar, "Translation of text to speech conversion for hindi language", "International Journal of Science and Research", Impact Factor (2012): 3.358 Vol- 3 Issue-11, November 2014.

[7] Kumar Patra, Biplab Patra, Puspanjali Mohapatra , "Text to speech conversion with phonematic concatenation," International Journal of Electronics Communication and Computer Technology " (IJECCT) Vol- 2 Issue-5, September-2012, ISSN:2249-7838