

Image Text To Speech Conversion In The Desired Language By Translating With Raspberry Pi

Rithika.H¹, B. Nithya santhoshi²

Department of Information Technology
MNM Jain Engineering College
Chennai, India

rithikachordia@yahoo.com¹, santhoshinithya@gmail.com²

Abstract—The main problem in communication is language bias between the communicators. This device basically can be used by people who do not know English and want it to be translated to their native language. The novelty component of this research work is the speech output which is available in 53 different languages translated from English. This paper is based on a prototype which helps user to hear the contents of the text images in the desired language. It involves extraction of text from the image and converting the text to translated speech in the user desired language. This is done with Raspberry Pi and a camera module by using the concepts of Tesseract OCR [optical character recognition] engine, Google Speech API [application program interface] which is the Text to speech engine and the Microsoft translator. This relieves the travelers as they can use this device to hear the English text in their own desired language. It can also be used by the visually impaired. This device helps users to hear the images being read in their desired language.

Keywords—Raspberry Pi; Tesseract OCR engine; Google Speech API; Microsoft translator; Raspberry Pi camera board.

I. INTRODUCTION

There are already many systems which read images and give voice output [2] [3] [4]. But this system gives voice output in any language desired by the user. This is done by capturing the image which is to be read using a raspberry pi camera module [7]. Raspberry pi [7] is a credit card sized single board computer. The operating system used is Raspbian. A 15 cm ribbon cable is used to attach the camera module to the raspberry pi. The coding is done using python language. The Optical character recognition engine converts the images of text into machine encoded text and saves it in a text file. Tesseract is the OCR engine which is used for extracting the English text from the image and storing it in a text file [4]. The text to speech engine converts text to speech output [4]. eSpeak is a speech synthesizer which can easily be used in raspberry pi for speech output in English. For translating it to other languages Google text to speech engine and Microsoft translator is used [5]. Google text to speech is a screen reader which speaks the text on the screen. Microsoft translator is a multilingual statistical machine translation cloud service provided by Microsoft. It supports 53 different language systems [6]. This translated speech output could be heard through speakers or head set. The platform being used for simulation of this model is putty in SSH (Secure Socket

Shell). Putty is a free and open-source terminal emulator which is used to give commands to the raspberry pi. Generally, it is done using MATLAB [2] but it is different here because translation module is an added feature which cannot be done using MATLAB.

II. SYSTEM HARDWARE DESIGN

The hardware consists of the following parts: Raspberry pi camera module, Raspberry pi 3 [model B] mounted with SD card, speakers, Internet connection via Ethernet or Wi-Fi, laptop. The Fig 1 gives the block diagram of system hardware design.

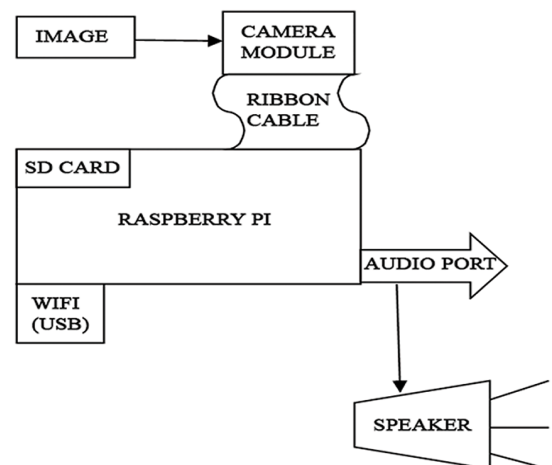


Fig. 1 System hardware design

A. Raspberry Pi 3 Model B

Raspberry Pi [7] is an ARM based credit card sized Single Board Computer created by Raspberry Pi Foundation. An 8 GB SD card is flashed with Raspbian OS and is put into the slot. An Ethernet cable is put in its slot for network connection. Power supply is enabled and raspberry pi is connected to the laptop via a USB cable. SSH client putty is used to work with raspberry pi via command line interface.

The technical specifications of raspberry pi 3 model B [3] [7] are: Broadcom BCM2837 Processor, Quad core ARM

Cortex-A53, Clock Speed of 1.2GHz, 1 GB RAM, RJ45 Port for Network Connectivity, wireless LAN (Wi-Fi) and Bluetooth 4.1, 4 USB Ports, GPIOs, Camera Interface is a 15-pin MIPI, Power Supply is of 2.5 A.

B. Raspberry Pi Camera Module

The camera module is connected with the camera serial interface of the raspberry pi using the 15 pin ribbon cable. Enable the camera support in the configurations [1] [7]. This helps in capturing a 5 MP resolution image by a single command (1). The command is:

```
sudo raspistill -o image.jpg (1)
```

The Fig. 2 shows a raspberry pi 3 model B with a raspberry pi camera module connected via a 15 pin ribbon cable.

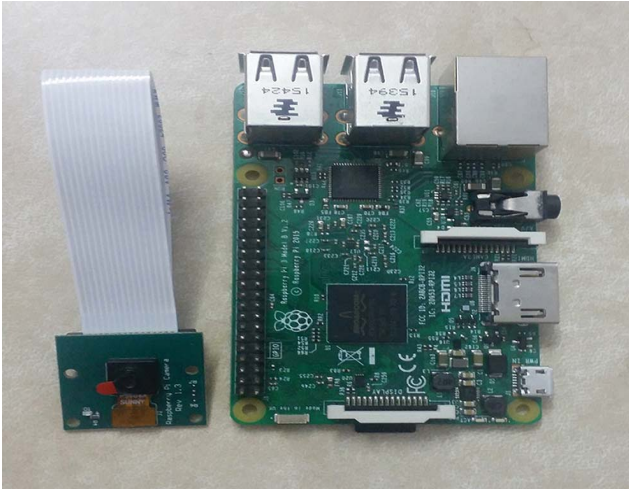


Fig. 2 Raspberry pi 3 model B with camera module

The specifications of raspberry pi camera module [3] [7] are: 5MP of image resolution, supports video and still images, has a 15 pin ribbon cable.

III. SYSTEM SOFTWARE DESIGN

The system software design consists of various phases which help in producing the end result. A machine understands the text of the image and gives a voice output with the processing of these phases. The various phases are: Image to text conversion (OCR), Text to speech and Microsoft translator. Primarily, the camera captures the image and stores it as an image file with .jpg extension. The OCR engine [3] converts it from image file to text file by extracting the numbers and characters of English alphabet only provided the text is printed. It cannot recognize handwritten texts. It is then converted to a .flac file by running a command using eSpeak [3], a text to speech engine. The flac file is given as an input to a python program which gives a translated speech and text output using Google text to speech engine and Microsoft translator [5]. The Fig. 3 gives the block diagram of system software design.

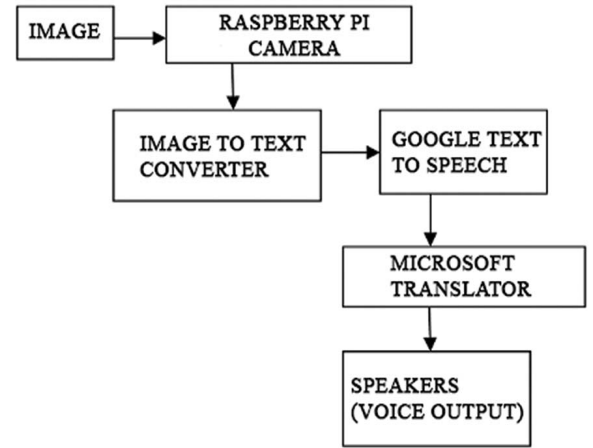


Fig. 3 System software design

A. Image To Text Converter

The raspberry pi camera module captures the image which is to be read and the image file is stored in the folder. The tesseract [3] [4] software is installed in the raspberry pi by a command (2). The command is:

```
sudo apt-get install tesseract-ocr (2)
```

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.

B. Text To Speech Synthesizer

It is a software used to synthesize speech from text [3]. A TTS Engine converts written text to a phonemic representation, the phonemic representation is converted to waveforms that can give a sound output. eSpeak [3] is a software which can be easily used in a raspberry pi by installing eSpeak engine. Here it is used for converting the text file into an audio file using .flac extension file. Flac stands for free lossless audio codec. It is an audio coding format for lossless compression of digital audio. At the end of this phase an audio file is created.

C. Translation And Speech Output

The flac file is an input to the python program which gives a translated speech output with the specified language [5]. The program has MPlayer, Google TTS engine and the credentials for translator services within it. MPlayer can play a wide variety of media formats, namely any format supported by FFmpeg libraries, and can also save all streamed content to a file locally. A companion program, called MEncoder, can take an input stream or file and transcode it into several different output formats, optionally applying various transforms along the way. Google text to speech engine is also used for converting the text to speech. Microsoft translator is a multilingual statistical machine translation cloud service provided by Microsoft [6]. During the command execution the standard language code should be given in the command for

the language required by the user. An example command for Spanish as the destination language is:

```
sudo nano pitranslate.py -o en -d es "filename" (3)
```

The output is heard by connecting the speakers to the audio jack and enabling it in the raspberry pi configuration [7]. The volumes can be adjusted by alsamixer [1] [5]. The translated speech output can be heard through the speakers.

IV. IMPLEMENTATION

The SD card is booted and put into the slot [3] [7]. Network connectivity is given and the raspberry pi is connected to the laptop via a USB cable. Putty is used as an SSH terminal which is the simulation platform. It is a command line interface where commands need to be given for the execution. All upgrades and updates are performed. Necessary packages are installed. The configurations are set [3]. Camera and speakers are set. The software design process is implemented, codes are executed and thus a image is read and translated voice output is given. The Fig. 4 shows the setup of the project.

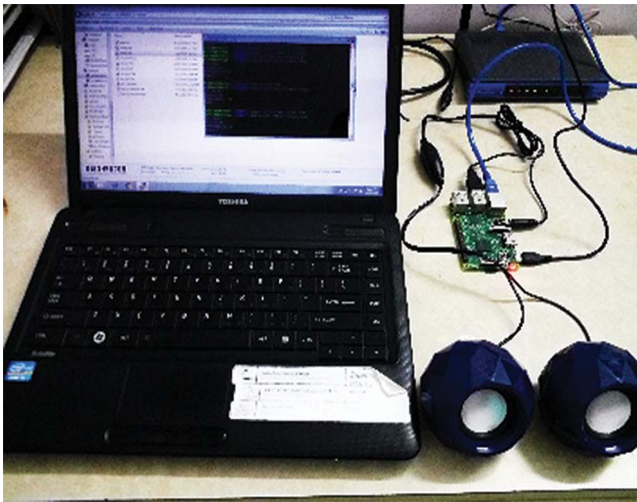


Fig. 4 Setup of the project

The input or the feed is a image captured by the raspberry pi camera module. Only printed characters or numbers in English present in the image can be converted to text file by the tesseract OCR. The text to speech engine and translator help in giving the speech output.

V. EXPERIMENT ANALYSIS

Any number or character in English alphabet can be read by this pi. Here numbers are taken as input and analyzed. A printed image of numbers is the input. The raspberry pi camera module is used to capture the image by a single command. The image is captured by the camera module and stored in a .jpg file format by a 15 pin ribbon cable. The Fig. 5 shows the image of numbers which was captured by the camera module.

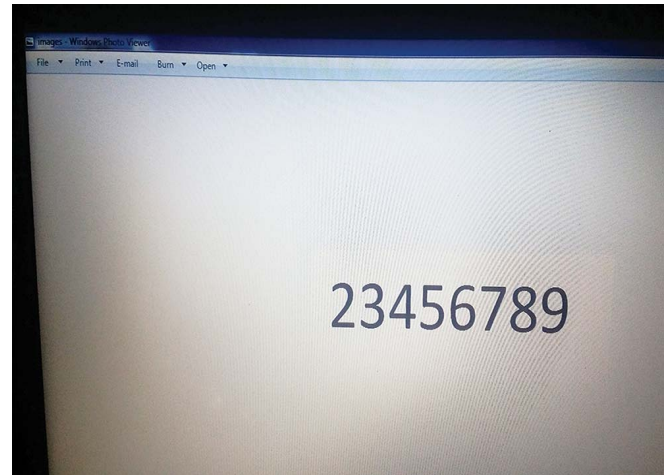


Fig. 5 Image captured

The captured image is converted to .txt file. Fig. 6 shows the text file generated by the tesseract [3] [4] software.

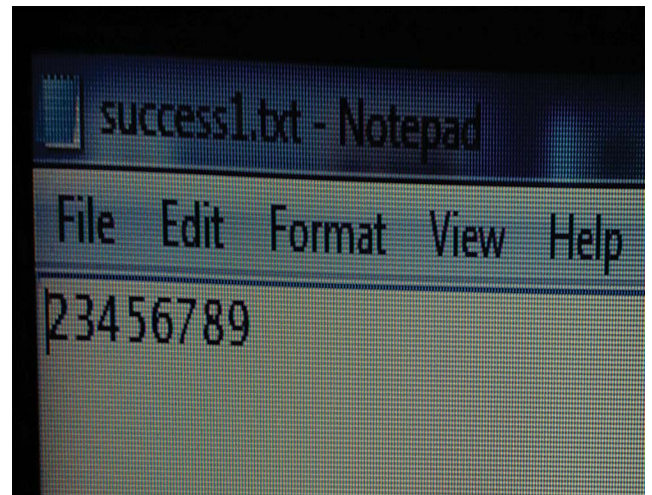


Fig. 6 Image converted to text and stored in a file

The text file is then converted to .flac file which is given as input for translation. Fig. 7 shows the conversion of text to flac file by espeak.

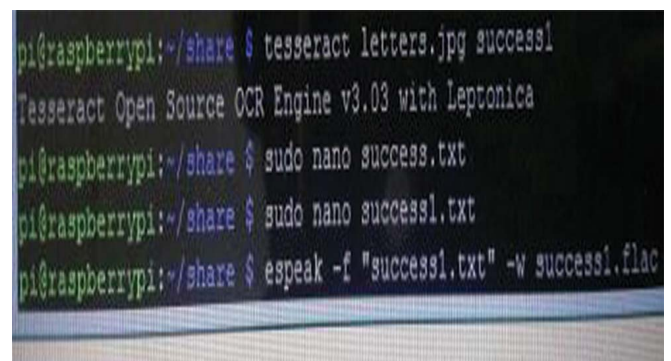
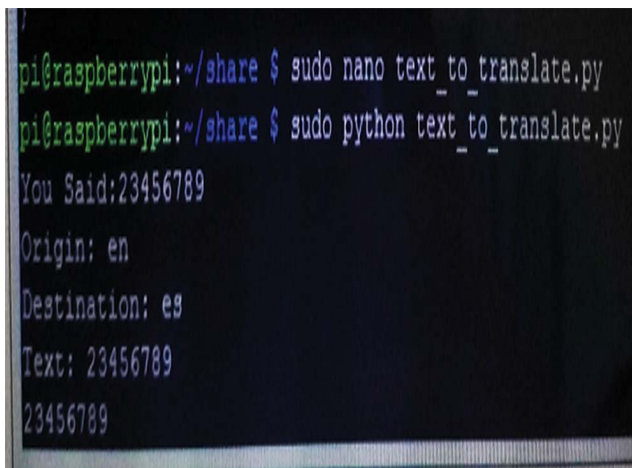


Fig. 7 Text file converted to audio file

The python program is then executed for translated speech output. Fig. 8 shows the command line execution for the translated speech output from English to Spanish language.



```
pi@raspberrypi:~/share $ sudo nano text_to_translate.py
pi@raspberrypi:~/share $ sudo python text_to_translate.py
You Said:23456789
Origin: en
Destination: es
Text: 23456789
23456789
```

Fig. 8 Translated speech output in Spanish language

The pictures illustrate an example of the speech output from an image consisting of numbers in English which is translated to Spanish. Similarly, we can also read characters with the help of this prototype.

VI. CONCLUSION AND FUTURE WORK

This project helps the travelers to hear their native language when they are in foreign country. It is also useful for people who do not know the English language. It can be used by any person who wants to read an image and hear the translated voice output. Future enhancements can be made in

the image to text conversion for more accuracy and in the speech output to eliminate noise by using various algorithms.

ACKNOWLEDGMENT

We would like to thank our project guide Dr. A. Srinivasan, who has been an inspiration. He always has motivated us and helped us in understanding various concepts. We are also grateful to the professor of SSN college of engineering, Mr. Vinob chandder, who taught us the basics of working with raspberry pi.

REFERENCES

- [1] V. Ajantha Devi, Dr. S Santhosh Baboo (Jul-Aug 2014), "Optical Character Recognition on Tamil Text Image Using Raspberry Pi" International Journal of Computer Science Trends and Technology (IJCST) – Vol. 2 Issue 4.
- [2] Raja Venkatesan.T, M.Karthigaa, P.Ranjith, C.Arunkumar, M.Gowtham, "Intelligent Transalate System for Visually Challenged People" International Journal for Scientific Research & Development (IJSRD), ISSN (online): 2321-0613, Vol. 3, Issue 12, 2016.
- [3] Anusha Bhargava, Karthik V. Nath, Pritish Sachdeva, Monil Samel (April 2015), "Reading Assistant for the Visually Impaired" International Journal of Current Engineering and Technology (IJCET), E-ISSN 2277 – 4106, P-ISSN 2347 – 5161, Vol.5, No.2.
- [4] K Nirmala Kumari, Meghana Reddy J (May 2016), "Image Text to Speech Conversion Using OCR Technique in Raspberry Pi" International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering(IJAREEIE), ISSN (Print): 2320 – 3765, ISSN (Online): 2278 – 8875, Vol. 5, Issue 5
- [5] <http://www.daveconroy.com/turn-raspberry-pi-translator-speech-recognition-playback-60-languages/>
- [6] <https://msdn.microsoft.com/en-us/library/hh456380.aspx>
- [7] www.raspberrypi.org