Advanced Optical Character Recognition for Blind Peoples Using Rasberry PI

Chaithanya Jyothi Y S[1], Dr K M Sadyojatha[2]
[1]*M. Tech in VLSI Design & Embedded Systems, Department of ECE, Ballari Institute of Technology and Management (BITM)*
[2]*Professor, Department of ECE, Ballari Institute of Technology and Management (BITM)*

Abstract **- The main objective of this paper is to provide blinds voice data and an inventive system that can help the blind to read altered. Many techniques have been used for the processing of finding obstacles and alert them by voice and convert images into text and then to the sound that can be heard by the blind. All systems are implemented using a Raspberry Pi Micro controller-based system. This article presents a novel approach, identifying the area of interest in the object that can be identified, using text scan and optical recognition to transform the text of binary characters tools. The second-order characters are identified by the scanned image being emitted before it is detected. A third method has also been introduced that can be used to convert e-speech exposure characters by matching that model. Embedded systems have been developed based on ARM technology which helps blind people to detect obstacles. All methods presented in this paper have been implemented in embedded applications. The attached board has been tested with different words and speech in English, more can be done to generate speech in other differences both National and International Languages.**

Index Terms **- blind, Raspberry Pi, text, ARM, embedded**

1.INTRODUCTION

A camera-based reader encourages blind people to peruse marks on the items and other handheld gadgets in their step by step lives. To separate the article from hefty backgrounds and different environmental factors, a compelling movement-based strategy is utilized to characterize Region of Interest (ROI) in the camera see. In the acquired ROI text, confinement and recognition are done2. The printed setting or filtered picture is changed over into PC recognition design by utilizing Optical Character Recognition (OCR) so it can speed up operation3. In model distinguishing proof, all the characters are limited and disengaged and afterward the end character picture is coordinated to a pre-processor for eliminating the noise. All the characters are contrasted and an information base of recognized characters which are amassed together to frame introductory content example. The yield is then given to the e-talk motor to change over content to discourse and this yield is given to blind clients through headphones. The obstruction in the process is distinguished by utilizing ultrasonic sensor.

285 million individuals are assessed to be visually impaired around the world, 39 million are blind and 246 million have low vision. This paper essentially offers an ease framework to support the blind people. The most recent advancement in computerized cameras, versatile PCs helps in planning the camera-based items that consolidates the PC vision innovation and the optical character recognition framework. Programming's, for example, video magnifiers, screen readers and optical guides are accessible to support the blind individuals and those with vision misfortune to utilize a PC. There are just barely any gadgets that can offer great admittance to blind clients to peruse printed text in outside world. In this day and age, number of difficulties have been looked by the blind individuals on the grounds that printed text is wherever as receipts bank guidelines over the drugs and so forth. At the point when blind individuals are helped to peruse printed text and item marks, it will build their free living just as cultivate social and financial self-proficiency.

2. PROPOSED SYSTEM

Existing systems such as a movable bar code reader built in to facilitate blind users with information about these products on speech to name various products in an extended database. On the other hand, it is extremely tough for blind users to locate the barcode place and point the barcode reader at the barcode

precisely for a gigantic drawback. Another reading aid system of such a pen scanner has been developed under similar circumstances. Other major problems for blind people are found to be an obstacle. Sample system, a camera used for reading text and obstacle. Text varies in different languages such as English. The complete application was published by the Raspberry Pi Board.

A. Software Specifications

1. Operating system (Linux), Raspbian Pixel
2. Language (python)
3. Library (Espeak (Linux-library))
4. Library (Tessaract (Linux-library))

B. Hardware Specifications

1. Raspberry Pi 3
2. SD Card
3. Ultrasonic Sensor
4. HDMI to VGA Converter
5. 5v,2A Power Supply

2.1 Raspberry Pi

The Raspberry Pi board utilized for improvement and experimenting the techniques introduced in this paper is appeared in the Figure 1. It is a smaller than usual PC which is in MasterCard size that connects to a TV or screen and uses a mouse and console. It permits individuals of all age gatherings to utilize a computer and help them in getting the hang of programming dialects like python and scratch. The Raspberry Pi equipment has developed through a few forms that feature varieties in memory limit and peripheral-gadget support.

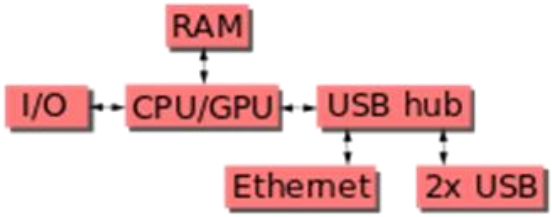

Figure 1 Block Diagram of Raspberry pi

This block diagram portrays Models A, B, A+, and B+. Model An, A+, and the Pi Zero do not have the Ethernet and USB center parts. The Ethernet connector is inside associated with an extra USB port. In Model An, A+, and the Pi Zero, the USB port is associated legitimately to the framework on a chip (SoC). On the Pi 1 Model B+ and later models the USB/Ethernet chip contains a five-point USB center, of which four ports are accessible, while the Pi 1 Model B just gives two. On the Pi Zero, the USB port is additionally associated legitimately to the SoC, yet it utilizes a miniature USB (OTG) port.

2.1 Proposed System Configuration

The text to speech structure comprises of three essential instruments Image catching, Data overseeing and speech yield. The image catching component assembles the images which contains text as video or image. In our model framework images are caught by utilizing a WEB CAM. Utilizing the camera, image of the object from the jumbled foundations or different environmental factors are extricated. Text confinement calculation is utilized for acquiring the text from images; an Ada support learning model is applied to restrict the text in the image. Text recognition is actualized to change over the image-based text to readable codes. In the model built up a PC is utilized for information preparing. The sound yield is conveyed to the blind clients through e-talk motor and the sound is introduced in English language. The block diagram that shows various parts of the framework for actualizing the proposed techniques has been appeared in the Following figure 2

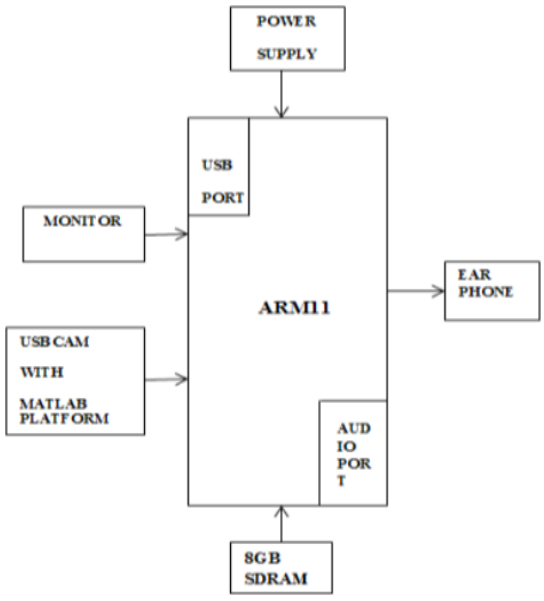

Figure 2 Block Diagram of Proposed System

2.2 Character Recognition System

Images for Character Recognition (CR) framework have been obtained by catching the photo of an archive or item or by checking the transcribed text or by

immediately composing into a PC framework. The noise accessible in the examined object is eliminated at pre-preparing stage at the hour of text creation, proper channels, for example, min-max channel, Gaussian channel and so on are utilized to wipe out the noise. "Binarization" technique changes highly contrasting or hued image to double image which is as dark or white. By and large dark image is the combination 0s and 1s and the double image is the blend of 0s or 1s. The positive qualities over the limit level are taken as 1s and the foundation with negative qualities are taken as 0s. Nearly image handling strategy yield is as double image. On thresholding, a dim level image with pixel esteems running somewhere in the range of 0 and 255 represent the paired image10. While filtering the text, it might possibly be actually on a level plane put, by utilizing the inclination edge revision it is actually adjusted. In the event that the information image is excessively enormous, at that point it is resized to lessen the measurements to update the expediency of handling. During segmentation all lines are isolated by applying line histogram and by utilizing section histogram, words are separated from each column and last characters are removed from words. Feature extraction is a basic aspect of any example recognition application and features of individual characters are extricated, wavelet based multiresolution method for feature extraction has been utilized. Classifiers coordinate the info characteristics with accumulated example and recognize impeccable proportional information. Post handling improves the precision of recognition and in this manner should be utilized at times. The bit by bit measure utilized for perceiving the characters

2.2.1 Espeak:

ESpeak is an order line device for Linux that changes over text to speech. This is a reduced speech synthesizer that offers help to English and numerous different dialects. It is written in C. ESpeak peruses the text from the standard info or the information record. The voice produced, be that as it may, is no place near a human voice. Yet, it is as yet a smaller and helpful device in the event that you need to utilize it in your tasks.

Some of the main features of eSpeak are:

1. An order line device for Linux and Windows
2. Talks text from a record or from stdin
3. Mutual library variant for use by different projects
4. SAPI5 adaptation for Windows, so it very well may be utilized with screen-readers and different projects that support the Windows SAPI5 interface.
5. Ported to different stages, including Android, Mac OSX and so forth.
6. A few voice characteristics to look over
7. Speech yield can be spared as .WAV record
8. SSML (Speech Synthesis Markup Language) is supported halfway alongside HTML
9. Small in size, the total program with language support and so on is under 2 MB.
10. Can make an interpretation of text into phoneme codes, so it could be adjusted as a front end for another speech amalgamation motor.
11. Advancement instruments accessible for delivering and tuning phoneme information

2.2.2 Install eSpeak

To install eSpeak in Ubuntu based system, use the command below in a terminal:

```
sudo apt-get install espeak
```

eSpeak is an old instrument and I assume that it ought to be accessible in the archives of other Linux dispersions, for example, Arch Linux, Fedora and so on. You can introduce eSpeak effectively utilizing dnf, pacman and so forth. To utilize eSpeak, simply use it like: espeak and press enter to hear it resoundingly. Use Ctrl+C to close the running system.

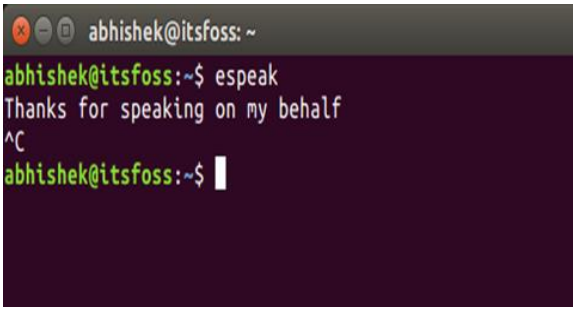

Figure 3.2 Espeak System

There are a few different choices accessible. You can peruse through them through the assistance segment of the program.3

2.2.3 GUI version: Gespeaker

On the off chance that you favor the GUI adaptation over the order line, you can introduce Gespeaker that gives a GTK front end to eSpeak.

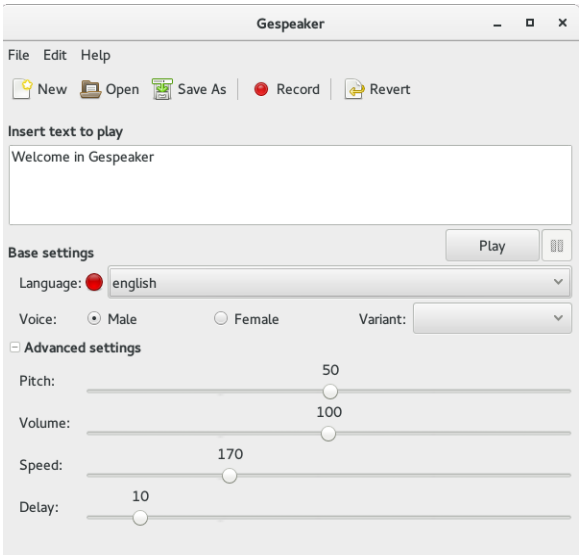

Figure 3 Gespeaker

While such devices probably will not be valuable for general processing need, it could be helpful in the event that you are chipping away at certain activities where text to speech change is required. I let you choose the utilization of this speech synthesizer.

2.3 Object Detection

In model framework introduced, object is distinguished by utilizing ultrasonic sensor; it is associated with the Raspberry Pi board. At the point when anyone's hand is put before that ultrasonic sensor, e-talk motor alarms the blind individual that an individual is ahead. An Ultrasonic sensor is a gadget that can quantify the separation to an object by utilizing sound waves. It allots separation by sending a sound wave at a particular recurrence and tuning in for that sound wave to bob back. By recording the slipped by time between the sound wave being produced and the sound wave bobbing back, it is conceivable to ascertain the separation between the sonar sensor and the object.

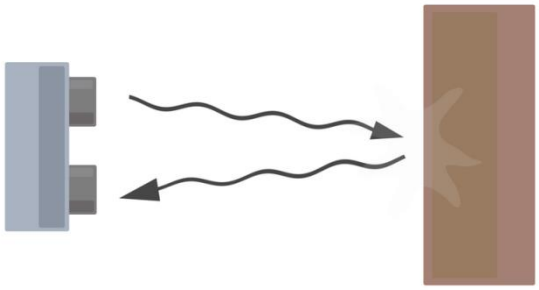

Figure 4 Diagram of the basic ultrasonic sensor operation

Since it is known thggat sound goes through air at around 344 m/s (1129 ft/s), you can set aside the effort for the sound wave to return and duplicate it by 344 meters (or 1129 feet) to locate the absolute full circle separation of the sound wave. Full circle implies that the sound wave voyaged multiple times the separation to the object before it was distinguished by the sensor; it incorporates the 'trip' from the sonar sensor to the object AND the 'trip' from the object to the Ultrasonic sensor (after the sound wave ricocheted off the object). To discover the separation to the object, essentially partition the full circle separation fifty-fifty.

Understand that a few objects probably will not be recognized by ultrasonic sensors. This is on the grounds that a few objects are formed or situated so that the sound wave ricochets off the object, however, are redirected away from the Ultrasonic sensor. It is likewise feasible for the object to be too little to even consider reflecting enough of the sound wave back to the sensor to be distinguished. Different objects can retain the sound wave all together (fabric, covering, and so on), which implies that it is extremely unlikely for the sensor to identify them precisely. These are important variables to consider when planning and programming a robot utilizing a ultrasonic sensor.

2.4 Circuit and working

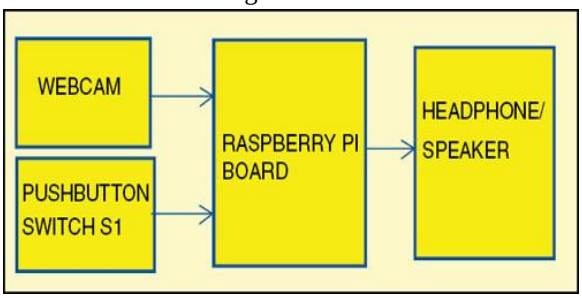

Figure. 5.: Block diagram of the See and Speak system

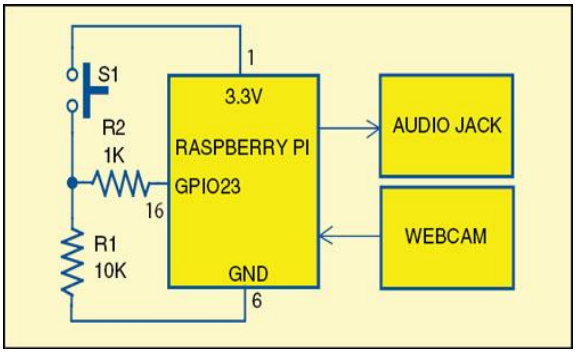

Figure. 6: Circuit connection to Raspi board

The system uses a webcam, Raspi and gui to take pictures as shown in the block diagram in Fig. 5 and the circuit diagram in Fig. 6. First, focus the webcam manually towards the text. Then, to take a picture, press pushbutton switch S1. A delay of around ten seconds is provided, which helps to focus the webcam if you accidentally disturb the webcam and defocus it while pressing the button.

After ten seconds, a picture is taken and processed by Raspi to provide the spoken words of the text through the earphone or speaker plugged into Raspi through its audio jack. When the GPIO pin is set as input, it is floating and has no defined voltage level. For you to be able to reliably detect whether the input is high or low, you need to have some simple resistive circuit so that it is always connected and reads either high or low voltage.

One of the terminals of switch S1 is connected to ground (GPIO pin 6) through pull-down resistor R1 of 10-kilo-ohm. The other terminal is connected to 3.3V of GPIO pin 1. When S1 is pressed, a high voltage is read on GPIO pin 16. When S1 is released, GPIO pin 16 is connected to ground through R1, hence a low voltage is read by GPIO pin 16. When pushbutton S1 is pressed, the webcam takes a picture of the text (after some delay). This text picture is sent to an optical character recognition (OCR) module such as Tesseract. Tesseract is an open source OCR that can be used to recognize the text present in the image. It supports many languages. Here, we have used it for English alphabets.

Before feeding the image to the OCR, it is converted to a binary image to increase the recognition accuracy (to check if the image is colored). Image binary conversion is done by using Imagemagick 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). Here, Festival software is used to convert the text to speech. Festival is an open source text-to-speech (TTS) system, which is available in many languages; in this project, English TTS system is used for reading the text.

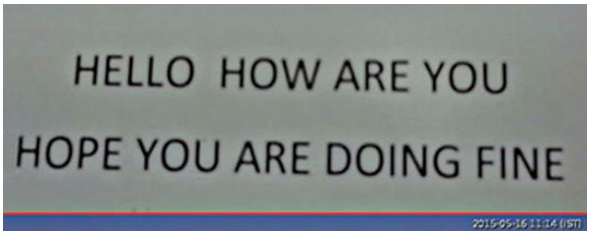

Figure. 7: Text image (example.jpg) captured by the camera during testing

In the event that the goal of your camera isn't acceptable, OCR execution will be poor, and the speech yield will likewise corrupt. We have utilized Logitech C270 camera for testing this venture. The camera goal naturally is 720×340, which is the most extreme goal supported by this webcam. On the off chance that the camera cannot catch the text appropriately, you will either hear mutilated sound from the speaker or no solid by any means. The text image (example.jpg) being caught by this camera during testing is appeared in Fig. 3.7 You can discover example.jpg and speech.txt documents under Home index.

3. CONCLUSION

The proposed Text to speech (TTS) framework was planned so as to create a proportional acoustic sign which goes in synchronization with the text which is given as a contribution to the Raspberry-Pi framework. The programming methodology included the GPIO lines, the information base availability, and the sound enhancers. This exploratory approval can be stretched out to conjoined sounds and expressions of higher intricacy too. The extent of the Raspberry-Pi in speech combination and feature extraction is likewise very limitless.

REFERENCES

[1] Ahmad, Z.H.; Int. Islamic Univ. Malaysia, Kuala Lumpur; Khalifa, O., “Towards designing a high intelligibility rule based standard malay textto-speech synthesis system” IEEE, International Conference on Computer and Communication Engineering. pp 89-94, May 2008.

[2] Chungurski, S.; Fac. of Commun. & Inf. Technol., FON Univ., Skopje; Kraljevski, I.; Mihajlov, D.; Arsenovski, S. “Concatenative speech synthesizers and speech corpus for Macedonian language” International Conference on Information Technology Interfaces. ITI 2008., pp 669-674, June 2008.

[3] Damadi, M.S., Azami, Bahram Zahir ; Electr. & Comput. Eng. Dept., Univ. of Kurdistan, Sanandaj, Iran, “Design and Evaluation of a TexttoSpeech System for Azerbaijani Turkish Language and Database Generation” Digital

Telecommunications, 2009.ICDT'09., pp 111 - 116, July 2009

[4] Falas.T., Stafylopatis, A-G;Dept. Of comput. Sci. & Eng. , Cyprus coll., Lefkosia, Cyprus; ; Graduate Sch. of Eng., Tokyo Univ. ; Hirose, K. ; Takada, S. ; Minematsu, N., "Neural networks in text-to-speech systems for the Greek language", Electrotechnical Conference, 2000. MELECON 2000. 10 th Mediterranean, pp 574-577, 2000.

[5] Lu Gao, Hongzhi Yu, Yonghong Li, Jie Liu.; Key Lab of China's Nat. Linguistic Inf. Technol., Northwest univ. for Nat., Lanzhou, China. "A Research on Text analysis in Tibetan Speech sysnthesis"Information And Automation(ICIA),2010, IEEE International Conference , pp 817 - 822, June 2010.

[6] Pornpanomchai, C., Soontharanont, N.; Langla, C.; Fac. Of inf. & Commun. Technol., Mahidol Univ., Bangkok, "A Dictionary-Based Approach for Thai Text to Speech (TTTS)" Third International Conference on Measuring Technology & Mechatronics Automation, IEEE, International Conference, pp 40-43, Jan 2011.

[7] Raitio, T.; Dept. of Signal Process. & Acoust., Aalto Univ., Helsinki, Finland ; Suni, A. ; Yamagishi, J. ; Pulakka, H. , "HMM-Based Speech Synthesis Utilizing Glottal Inverse Filtering" IEEE Transactions on Audio, Speech, and Language Processing, pp 153-165, Oct 2010.

[8] Sadeque, F.Y.; Dept. Of Comput. Sci. & Eng., Yasar, S.; Islam, M.M., "Bangla text to speech conversion: A syllabic unit selection approach" Cyberworlds, IEEE, International Conference, pp 1-6, May 2013.

[9] www.raspberrypi.org.

[10] Yagi, Y. ; Graduate Sch. of Eng., Tokyo Univ. ; Hirose, K. ; Takada, S. ; Minematsu, N., "Improved concept-to-speech generation in a dialogue system on road guidance" Cyberworlds ,IEEE, International Conference , pp 8 - 436, May 2005.