

# See and Speak Using Raspberry Pi

GURUNATH REDDY M.

Imagine a machine that can see and speak, and is fully portable. It is surprising, right? In this article, we present a system based on Raspberry Pi, or Raspi, that can see and speak. It takes pictures of text content around its vicinity from the webcam attached to Raspi, converts it to speech and speaks out the text through a headphone or speaker connected to its audio jack.

This portable device can be used in many applications in robotics, automation, hobby projects and more. For example, you can focus your webcam to a text, such as English alphabets, on a signboard, followed by pressing a pushbutton switch connected to Raspi. It will capture the text and convert it to speech and read it out aloud to you. When you get bored of reading books, just click a picture of the textbook page and make it read the same aloud to you.

## Circuit and working

The system uses a webcam, Raspi and pushbutton switch S1 to take pictures as shown in the block diagram in Fig. 1 and the circuit diagram in Fig. 2.

The webcam (we used Logitech C270) is connected to Raspi through one of its USB ports and pushbutton switch S1 to its GPIO pin 16 (or GPIO23) through resistor R2 (1-kilo-ohm) as shown in the circuit diagram.

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

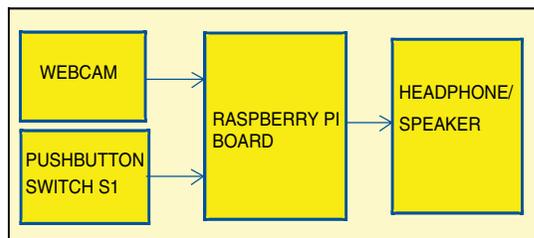


Fig. 1: Block diagram of the See and Speak system

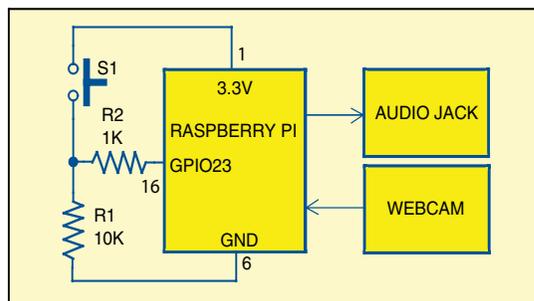


Fig. 2: Circuit connection to Raspi board

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 recognise



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 coloured). 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.

## Software installation

Update and upgrade Raspi-related software using the commands below and reboot your Raspi:

```
$ sudo apt-get update
$ sudo apt-get upgrade
```

Install Tesseract OCR system by issuing following command:

```
$ sudo apt-get -s install tesseract-ocr
```

Install image-manipulation tool Imagemagick using the command:

```
$ sudo apt-get install imagemagick
```

Install fswebcam to get pictures from the webcam using the command:

```
$ sudo apt-get install fswebcam
```

To check whether the webcam is installed properly, issue the command:

```
$ fswebcam example.jpg
```

An image by the name `example.jpg` will get saved in the home directory. If the resolution of this image is not up to the mark, change it by using `-r` option in `fswebcam`. One example of 1280x720 resolution capturing is shown below. Set this according to your webcam.

```
$ fswebcam -r 1280x720 example.jpg
```

To install sound on Raspi, install also sound utilities using the command below:

```
$ sudo apt-get install alsa-utils
```

Edit the modules file at `/etc/modules` using nano editor.

```
$ sudo nano /etc/modules
```

Add the line `snd_bcm2835`. If `snd_bcm2835` is already present, leave the file as it is.

Then, save the file by clicking `ctrl+o` and exit with `ctrl+x`.

Now, install `mplayer` audio movie player using the command:

```
$ sudo apt-get install mplayer
```

Once you have completed all the steps mentioned above, install Festival text-to-speech software using the command:

```
$ sudo apt-get install festival
```

You may try Festival installation using the command below in the terminal and you will hear Hello EFY in the earphones.

```
$ echo "Hello EFY" | festival --tts
```

Once all the above software are installed, copy `see.py` Python code, which is included in this month's DVD available with Electronics For You Plus version of the magazine in Home folder.

Run `see.py` by issuing the following command:

```
$sudo python see.py
```

`see.py` runs indefinitely to get input from the user.

**Note.** If the resolution of your camera is not good, OCR performance will be poor and the speech output will also degrade.

We have used Logitech C270 camera for testing this project. The camera resolution by default is 720x340, which is

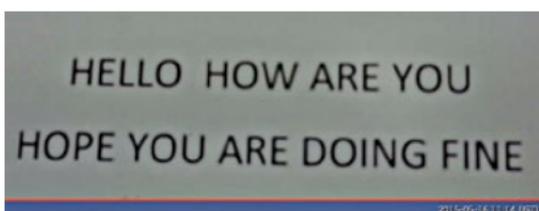


Fig. 3: Text image (`example.jpg`) captured by the camera during testing

### EFY Note

The source code of this project is included in this month's EFY DVD and is also available for free download at [source.efymag.com](http://source.efymag.com)

the maximum resolution supported by this webcam. If the camera is unable to capture the text properly, you will either hear distorted sounds from the speaker or no sound at all.

The text image (`example.jpg`) being captured by this camera during testing is shown in Fig. 3.

You can find `example.jpg` and `speech.txt` files under Home directory. ●

*Gurunath Reddy M. is an MS student at IIT Kharagpur*