Implementation of a reading device for bengali speaking visually handicapped people

Md. Mahade Sarkar, Shuvasis Datta, Md. Mahedi Hassan
Department of Electrical and Electronic Engineering
Chittagong University of Engineering and Technology,
Chittagong - 4349 , Bangladesh
mahadesarkareee@gmail.com, shuvasisdatta@gmail.com, m.m.mahedihaasan@gmail.com

***Abstract*—A reading device is a compact hardware setup with necessary programmes coded in it which read out printed documents like a human reader. People having eyesight problem can't read books, papers or any kind of printed reading materials. This problem can be solved simply by taking image of the reading materials, extracting words from the image and converting those words to sound, so by hearing that text converted sound they can understand what is written on that paper. A device is implemented for Bengali speaking visually handicapped people. For character recognition tesseract-ocr is used as optical character recognition(OCR) engine. Python gTTS module, a text to speech engine is used to convert the words extracted by tesseract-ocr to sound. Whole process is implemented on Raspberry Pi based a compact hardware design. Accuracy of character detection from the captured image and words to sound conversion is as high as 85 %. It is to be mentioned that accuracy is calculated as percentage of correct words to total words in an image**

***Keywords*—Tesseract; OCR; gTTS; Raspberry Pi; linux.**

I. Introduction

Large number of Bengali speaking people from Bangladesh and India, having eye sight problem facing much trouble reading printed papers, books in their native language. It's too hard for them to study in conventional system like Braille. Braille is a tactile writing system used by people who are blind or visually impaired. In Braille it is traditionally written with embossed paper. Braille is an analog system to study which needs time to learn and not a easy way to read. However all materials are not written in Baille system, almost all papers are printed in this modern age. This device will help them reading printed papers and any kind of printed Bangla books. The device takes the image of the printed documents. From that image characters, words and sentences are extracted using optical character recognition(OCR) engine and these are formated as a text file. A text to speach(TTS) converter engine is used to convert that text file to a sound file. Sound is played by omxplayer, a linux based open source sound player. The whole process is being implemented in a Raspberry Pi model B which has great processing power in a compact design.To complete this job, two main part is doing OCR and converting the text file to sound. Optical Character Recognition (OCR) is the process of extracting text from an image. The main purpose of an OCR is to make editable documents like a text file from image file. For doing OCR in Bengali language lot of work has been done. Some has developed their own algorithms and some has developed based on existing OCR. A complete printed Bangla OCR system is shown in [1]. Character Segmentation is shown in [2]. Bangla word formation patern is very complex. It has not only vowels and consonants but also compound word which is build using two or more individual character [3]. A cluster of characters forms a word using a horizontal line in Bengali which is called "Mattra". Identification of Matra region and overlapping characters for OCR of printed Bengali scripts are shown here [4]. Tesseract an open source based Optical Character Recognizer for Bangla printed documents is shown [5]. Several OCR was proposed but none of those are so efficient as tesseract. For the implementation of this project tesseract-ocr engine has been used for doing OCR. Tesseract is an open source Optical Character Recognition (OCR) engine which is maintained and developed by google. For recognition purpose each individual language need pretrainned data of the characters. Tesseract has enhanced pretrainned data for Bengali language. A Text to Speech (TTS) is a computer based system which is capable of converting computer readable text into speech. There are two main components creating a sound such as Natural Language Processing (NLP) and Digital Signal Processing (DSP). Some good amount of work has been done from diffrent perspective for creating bangla TTS. Text normalization and diphone preparation for Bangla TTS is presented in [6]. Using Epoch Synchronous Non Overlap Add (ESNOLA) for speech generation is described for implementation of a Bengali speech synthesizer for mobile devices in [7]. A Bangla TTS system was developed using open source Festival in [8]. A Framework for Bangla Text to Speech Synthesis has proposed a new framework for Bangla TTS [9]. All these efforts were proposed theoritically but none of these is implemented as a open source based package which can be used for this project for smooth Bengali voice. eSpeakNG is a compact open source based speech synthesizer for Linux, Windows and other platforms. Espeak gives a nice voice quality for english but for Bengali, pronounciation is not accurate and voice quality is not so good. gTTS is a python interface which is text to speech API, gives excelent voice quality for Bengali language. gTTS operates based on comand line interface with a secure internet connection. Although gTTS works online but it's operation is very fast and smooth.

978-1-5386-2175-2/17/$31.00 ©2017 IEEE

II. System Overview

This device is implemented using a Raspberry Pi model B with a Pi camera. To implement this device main challenges are, extracting characters, words from the image and pronunciation of the words. Process start with capturing the image of the printed paper. Clear and full page captured image in essential as image is the raw ingredient for processing. Some preprocessing techniques is applied to make the image quality better. Detection of bangla words is done by tesseract. Processed image is passed to tsseract engine. Tesseract is an OCR engine which takes image as input and detect the characters, words from that image file. Output of tesseract is a text file. This text file contains the words forming sentences same as on the image. Text file is converted to sound using python gTTS module which is based on google text to speech coversion API. Sound file is saved in working directory in mp3 format. A sound player is used to play the sound file.

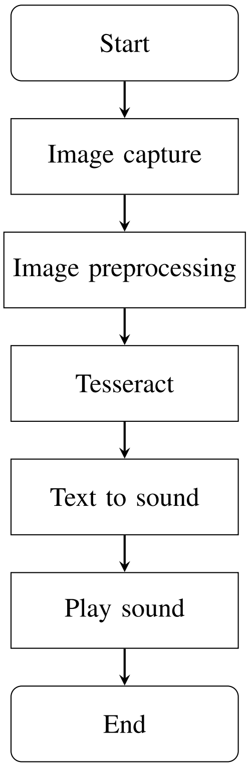


Fig. 1: Flow chart of the whole process

III. Methodology

Implementation of this device needs several processes. Methods used are:

- Hardware designing
- Image capture and processing
- Doing tesseract
- Sound file creating
- Sound playing

A. Hardware designing

Whole process is implmented on a Raspberry Pi model B which has a 1.2 GHz 64/32-bit quad-core ARM Cortex-A53 CPU and 1 GB LPDDR2 RAM at 900 MHz memory. Raspbian a linux based distribution is being installed as operating system. A Pi camera is connected to the Raspberry Pi to capture the image of the printed papers. Raspberry Pi is powered by a charger having good current rating as high as 1A. A loud speaker is connected to Raspberry Pi's sound port for playing sound. An on-off switch is connected to the GPIO(General Purpose Input / Output) pins to start and off the device.

B. Image capturing and processing

Image of the printed document is captured by the camera connected to the usb port of Raspberry Pi. It should be kept in mind that the output of tesseract and the sound quality depends on the quality of the captured image. A clear noise free complete image gives better output. But captured image will not always be a clear and complete . So some image procesing technique is applied to make it a better image before passing it to tesseract for OCR. For preparing image these processing steps has been applied:

- De-skewing
- Despeckle
- Line Removal
- Smoothing Images
- Canny Edge Detection

C. Tesseract

Main challenge for implementing this project is Bangla characcter recognition. This is done by tesseract-ocr. The processed image from previous section is passed to tesseract for getting the editable text file. Tesseract is a optical character recognition engine which maintained under open source project. It is installed in Raspberry Pi using the instruction from github source link [10]. The OCR Engine needs a pretrained data file to work on Bengali inputs. Tesseract has almost all language traineddata file for character recognition. For Bengali it is 'ben.traineddata', which is placed in tessdata directory. This file is consisted of some other files and a concatenation of those files. When the traineddata file placed in specific location tesseract detects character perfectly and generate text file. Tesseract has a level of accuracy in its engine which is standard. This engine can work very efficiently on it's library for accurate matching. In our system, we have implemented the library file or traineddata in a very detailed manner. It is important to mention that we need to uniquely identify each and every character in our system so that if the input file contains the character, the OCR recognizes it. For convenience, the following is a brief overview of how Tesseract works [11]:

- Outlines are analysed and stored
- Outlines are gathered together as Blobs
- Blobs are organized into text lines
- Text lines are broken into words

- First pass of recognition process attempts to recognize each word in turn
- Satisfactory words passed to adaptive trainer
- Lessons learned by adaptive trainer employed in a second pass, which attempts recognize the words that were not recognized satisfactorily in the first pass
- Fuzzy spaces resolved and text checked for small caps
- Digital texts are outputted

During these processes, Tesseract uses:

- Algorithms for detecting text lines from a skewed page
- Algorithms for detecting proportional and non propo rtional words (a proportional word is a word where all the letters are the same width)
- Algorithms for chopping joined characters and for asso ciating broken characters
- Linguistic analysis to identify the most likely word formed by a cluster of characters
- Two character classifiers: a static classifier, and an adaptive classifier which employs training data, and which is better at distinguishing between upper and lower case letters

D. Text to Speech

Main objective of this project is read out the documents. So text file found from tesseracct, need to convert to sound. That is done by gTTS. gTTS (Google Text to Speech) is a Python interface for Google's Text to Speech API. Creates an mp3 file with the gTTS module or gtts-cli command line utility. It allows unlimited lengths to be spoken by tokenizing long sentences where the speech would naturally pause. Sound conversion using gTTS is very fast. gTTS is installed using proper instruction from github link [12].

E. Sound playing

Sound file found by gTTS is played using OMXplayer, an open source based sound player. OMXPlayer is a commandline based sound player for the Raspberry Pi. It was developed as a tested for the XBMC Raspberry PI implementation and is quite handy to use standalone. OMXPlayer uses the OpenMAX (omx) hardware acceleration interface (API) which is the officially supported media API on the Raspberry Pi [13].

IV. IMPLEMENTATION

Building a OCR engine is a lengthy and time consuming process which needs large data training and applying intelligent detection techniques. So using the most effective engine tesseract-ocr for character recognition and gTTS for text to sound conversion is a better option. To implement this device using a Raspberry Pi based compact design peripheral devices, a camera is connected to pi's usb, a speaker as a sound output is connected to the sound port. A switch is connected to start the process. As the operating system is linux based, it's great advantage using comand line interface. The OCR engine and text to speech engine operation comands are operated in a linux termianl. Text to speech engine is a python module which operate based on google text to speech API. So an uninteruptable internet connection is required. That is solved by the wifi connection ability of Raspberry Pi 3 model B. Network configuration file of Raspbian is configured for a specific hotspot that is supplied by a Android phone. To automatecally start the whole process a startup script is written. After pi is powered and start switch is pressed on the startup script executes automatically. Start script is linux comandline based comands, thats executes sequentially. The comands in start script do these operations step by step: check network connectivity which is supplied by an android phone, capture image, do OCR, convert text file to sound file and play the sound file. After reading one page of printed document the script start the same process again. It continues in a loop as long as the audience is willing to read the book. For stop reading simply pressing the switch the device will be off.

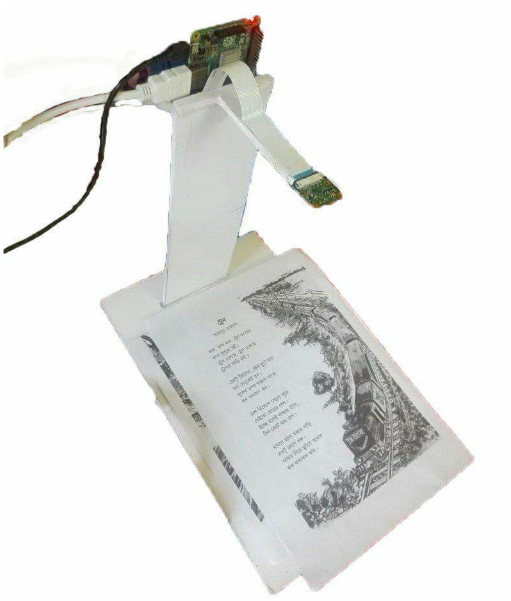
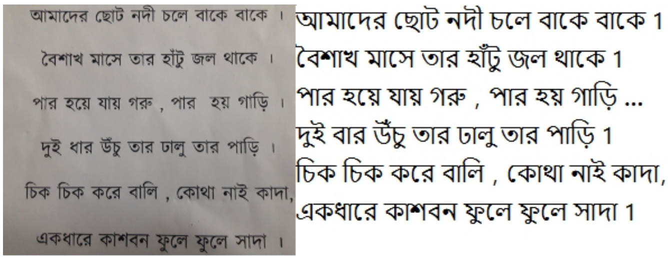


Fig. 2: Device implemented using Raspberry Pi

V. RESULT

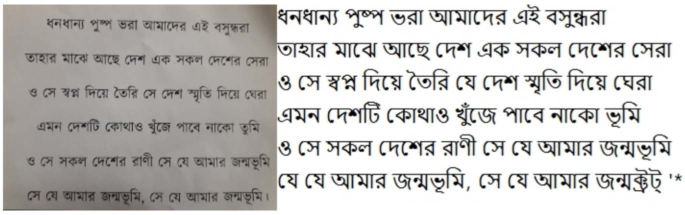
Result is calculated on the performance of character recognition and text to sound conversion ability of the device. Device performance accuracy depends on the quality of image and on the image resolution. That means higher accuracy depends on better quality of camer, ability of image capturing and image resolution. Device is tested on several bangla printed paper script. The output of tesseract engine is a text file. Accuracy of character recognition is calculated based on the number of correct words found from the ouput of that text file. Any mismatch of words to the original printed documents is considered as wrong word. For calculating accuracy the device is testing on three printed bangla script. Fig.(3-5) shows the image of three printed Bengali script and tesseracted output text file.



আমাদের ছোট নদী চলে বাকে বাকে ।
বৈশাখ মাসে তার হাঁটু জল থাকে ।
পার হয়ে যায় গরু , পার হয় গাড়ি ।
দুই ধার উঁচু তার ঢালু তার পাড়ি ।
চিক চিক করে বালি , কোথা নাই কাদা,
একধারে কাশবন ফুলে ফুলে সাদা ।

আমাদের ছোট নদী চলে বাকে বাকে 1
বৈশাখ মাসে তার হাঁটু জল থাকে 1
পার হয়ে যায় গরু , পার হয় গাড়ি ...
দুই বার উঁচু তার ঢালু তার পাড়ি 1
চিক চিক করে বালি , কোথা নাই কাদা,
একধারে কাশবন ফুলে ফুলে সাদা 1

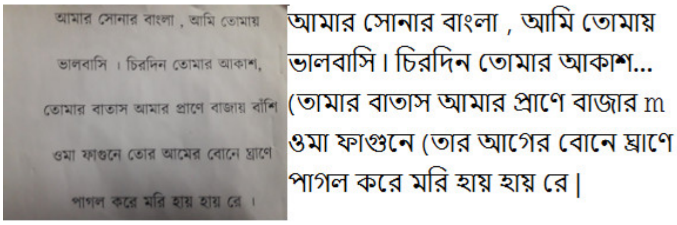
Fig. 3: First image and tesseracted output



ধনধান্য পুষ্প ভরা আমাদের এই বসুন্ধরা
তাহার মাঝে আছে দেশ এক সকল দেশের সেরা
ও সে স্বপ্ন দিয়ে তৈরি সে দেশ স্মৃতি দিয়ে ঘেরা
এমন দেশটি কোথাও খুঁজে পাবে নাকো তুমি
ও সে সকল দেশের রাণী সে যে আমার জন্মভূমি
সে যে আমার জন্মভূমি, সে যে আমার জন্মভূমি।

ধনধান্য পুষ্প ভরা আমাদের এই বসুন্ধরা
তাহার মাঝে আছে দেশ এক সকল দেশের সেরা
ও সে স্বপ্ন দিয়ে তৈরি যে দেশ স্মৃতি দিয়ে ঘেরা
এমন দেশটি কোথাও খুঁজে পাবে নাকো ভূমি
ও সে সকল দেশের রাণী সে যে আমার জন্মভূমি
যে যে আমার জন্মভূমি, সে যে আমার জন্মক্কট্ '*

Fig. 4: Second image and tesseracted output



আমার সোনার বাংলা , আমি তোমায়
ভালবাসি । চিরদিন তোমার আকাশ,
তোমার বাতাস আমার প্রাণে বাজায় বাঁশি
ওমা ফাগুনে তোর আমের বোনে ঘ্রাণে
পাগল করে মরি হায় হায় রে ।

আমার সোনার বাংলা , আমি তোমায়
ভালবাসি। চিরদিন তোমার আকাশ...
(তামার বাতাস আমার প্রাণে বাজার m
ওমা ফাগুনে (তার আগের বোনে ঘ্রাণে
পাগল করে মরি হায় হায় রে |

Fig. 5: Third image and tesseracted output

Correct words are counted from the text file manually. Counted words from the above three figures are shown in the table bellow:

TABLE I: Words count from each image

Page No	Total words	Correct words	Accuracy in percentage
01	38	37	97.3%
02	48	44	91.6%
03	27	22	81.4%

Accuracy is the total number of correct words to the total number of words found in the table .

$$\begin{aligned}\text{Accuracy in percentage} &= \frac{\text{correct words}}{\text{total words}} * 100\\ &= \frac{103}{113} * 100\\ &= 91.15\%\end{aligned}$$

91.15% accuracy if calculated based on three bangla printed paper. Testing the device on several other script accuracy is above 85% . The text to sound conversion accuracy is above 90%

Comparison of right words and wrong words can be easily unserstood by a bar plot. The figure below shows that for three printed parer.

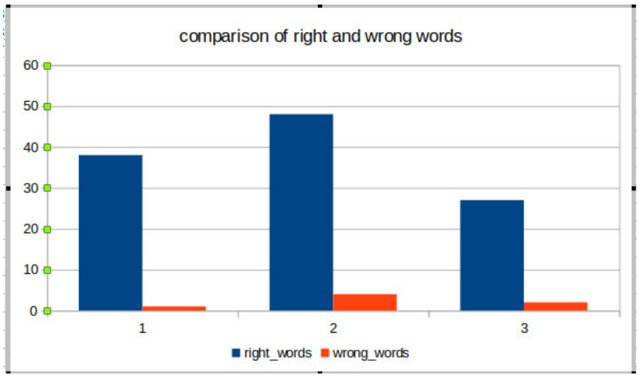


Fig. 6: Graphical presentation of right and wrong words

VI. CONCLUSION

A reading device for bengali speaking visually handicapped people is implenetated successfully. This device is implemented specially for visualy handicapped people but can be used in other purposes. It is operating smoothly and accuracy is excelent. Total execution time is a little longer but not more than two minute. It's beccause of the processing power of Raspberry pi. It can be expected that with the further development of Raspberry Pi's processor will reduce the execution time in near future. Text to speech engine is google text to speech API based, so a offline text to speech engine development can give a better flexibility for this device.

REFERENCES

[1] B. B. CHAUDHURI and U. PAL, *"A complete printed Bangla OCR system",Pattern Recognition, Vol. 31, No. 5, pp. 531549, 1998*

[2] Shamim Ahmed , Mohammod Abul Kashem , *"Enhancing the Character Segmentation Accuracy of Bangla OCR using BPNN", International Journal of Science and Research (IJSR) ISSN (Online): 2319-7064.*

[3] *M. K. Shukla, T. Patnaik, S. Tiwari, S. K. Singh, "Script Segmentation of Printed Devnagari and Bengali Languages Document Images OCR".*

[4] U. Pal and B. B. Chaudhuri, *"Identification of Matra Region and Overlapping Characters for OCR of Printed Bengali Scripts", Intelligent Computing and Information Science Communications in Computer and Information Science, vol. 135, pp. 606-612, 2011.*

[5] Md. Abul Hasnat, Muttakinur Rahman, Chowdhury Mumit Khan, *"An open source Tesseract based Optical Character Recognizer for Bangla script", 10th International Conference on Document Analysis and Recognition,pp. 671-675 , 2009*

[6] M. Masud Rashid , Md. Akter Hussain, M. Shahidur Rahman, *Text Normalization and Diphone Preparation for Bangla Speech Synthesis, Journal of Multimedia, 5:6, 2010.*

[7] S. Mukherjee, Shyamal Kumar Das Mandal, *" A Bengali Speech Synthesizer on Android OS", Proceedings of the 1st Workshop on Speech and Multimodal Interaction in Assistive Environments, pp. 4346, 2012.*

[8] F. Alam, S.M. Murtoza Habib, Mumit Khan, *"Bangla Text to Speech using Festival", Conference on Human Language Technology for Development, pp. 154-161, 2011.*

[9] K. M. Azharul Hasan, Muhammad Hozaifa,Sanjoy Dutta, Rafsan Zani Rabbi, *"A Framework for Bangla Text to Speech Synthesis", 16th International Conference on Computer and Information Technology (ICCIT),pp. 60 - 64 , 2013*

[10] *https://github.com/tesseract-ocr, Last accessed: August 14, 2017.*

[11] *https://tesseract-ocr.repairfaq.org/downloads/saltcymru_document5.pdf , Last accessed: August 14, 2017.*

[12] *https://github.com/pndurette/gTTS, Last accessed: August 14, 2017.*

[13] *https://www.raspberrypi.org/documentation/raspbian/applications/omxplayer.md , Last accessed: August 14, 2017.*