Implantation of a Reading Aid for Bengali-Speaking Individuals Who are Blind

Dr. K. B. Jagadish Kumar[1], J. Hari Babu[2], G. Vamsee Krishna [3], P. Raguraman[4], Shravya Chidurala[5]

[1, 2,3,4]Department of Computer Science and Engineering,
[5]Software Engineer
[1, 2, 3, 4] QIS College of Engineering and Technology, Ongole, Andhra Pradesh, India
[5]Accenture
jagadeesh@qiscet.edu.in[1], haribabu.j@qiscet.edu.in [2], vamsikrishna.g@qiscet.edu.in[3],
raghuraman.p@gmail.com [4], shravya.chidurala@accenture.com[5]
Corresponding Author Mail: qispublications@qiscet.edu.in

Article Info
Page Number: 561 - 569
Publication Issue:
Vol 70 No. 2 (2021)

Article History
Article Received*: 05 September 2021*
Revised*: 09 October 2021*
Accepted*: 22 November 2021*
Publication*: 26 December 2021*

Abstract
The reading aid is a hardware component with the necessary software installed to read printed materials aloud as a human reader would. Reading any printed material—books, documents, or anything else—is impossible for people with visual problems. Digitizing the reading materials, extracting the text from the image, and creating audio files from them will quickly solve this problem. This will allow people to understand the text on the paper merely by hearing the audio files created from the text. For those who speak Bengali and are visually impaired, a device has been implemented. The optical character identification (Oci) engine utilised for text detection is called tesser-act-oci. The Python-gtts module, a text-to-speech engine, is used to convert the words extracted by tesser-act-oci into sound.The entire operation is carried out using a small raspberry pi computer. Words to sound conversion and character detection accuracy from a captured image are both up to 85%. It should be remembered that the proportion of the right words to all the other words in a picture determines correctness.

I. INTRODUCTION

There are many Bengali speakingpeople in Indiawho have vision issues have a difficult time reading printed materials and books in their own language. They find it very difficult to learn using a traditional technique like Braille. Blind or visually impaired persons use the tactile writing system called braille. Embossed paper is typically used to write in Braille. Braille is an analogue learning system that takes time to master and is not a simple way to read. By the way, not all writing is done in the Baille style, and in this day and age, practically all documents are printed.They would find it simpler to read printed materials and any kind of printed Bangla books thanks to this device. The device takes a photo of the printed paperwork. The characters, words, and phrases in the image are extracted using the optical character identification (oci) engine, and the output is structured as text files. The text file is converted into a sound file using a “texttospeech” (tts) converter engine. Sound is played using the open source omx player, a linux-based sound player. The entire procedure is being implemented on a raspberry Pi model B, which provides outstanding computing power in a

tiny form size. oci and sound file conversion are the two major components needed to finish this task. Text can be extracted from images using a process called optical character identification (oci). Making editable documents from picture files in the form of text files is the primary function of an oci.

There has been a significant amount of work put towards Bengali oci. Some people have created their own algorithms, while others have done so based on oci that already exists. [1] Displays a comprehensive printed Bangla oci system. [2] Displays character segmentation. Bangla has a very intricate pattern for word creation. Along with vowels and consonants, it also has compound words [3] that are constructed with two or more distinct characters. In Bengali, the word "Mattra" is formed by a group of letters arranged along a horizontal line. These images [4] show how to identify the Matra region and recognise overlapping letters in printed Bengali scripts. Tesseract is a free, open-source optical character identification system for printed Bangla manuscripts [5]. There have been several ocis proposed, but none of them are as effective as tesseract. To carry out oci for this project, the tesseract-oci engine was used. tesseract is a Google-maintained and developed open source software. Each unique language requires character training data for identification purposes. tesseract has improved Bengali language pre-teaching data. A computer-based system called text to speech (tts) can turn text that can be read by computers into speech. The two fundamental procedures that go into creating a sound are “digitalsignalprocessing (dsp) and naturallanguageprocessing (nlp)”. For the creation of banglatts, a decent lot of effort has been done from several perspectives. Text standardization and diaphone preparation for Bangla TTS are presented in [6]. Epoch Synchronous Non Overlap Add (Esnola) for voice generation is described in detail in [7] and will be used to create a Bengali speech synthesiser for mobile devices. Using free source Festival, a Bangla tts system was created in [8]. In A Model for Bangla “text to speech” Processing [9], a new structure for Bangla tts is presented. All of these initiatives were theoretically suggested, but none of them have been put into practice as an open source package that are used here for a natural voice. espeakng is a portable open source speech synthesizer that runs on Linux, Windows, and other operating systems. Espeak has excellent voice quality for English, however Bengali has poor voice quality and improper pronunciation. Thanks to a text-to-speech API with a Python interface called gtts, Bengali has excellent sound quality. Gtts is managed over a secure internet connection and a command line interface. Although gtts is an online service, it operates quickly and without a hitch.

II. SYSTEMOVERVIEW

This device is created using a raspberry pi model that includes a pi camera. The main difficulties in applying this technique are word sound and letter retrieval from an image. The process begins by taking a picture of the printed paper. Since the raw material for processing is the image, a clear and full-page capture is crucial. In order to improve the image quality, some preprocessing techniques are used. Tesser-act is used to identify Bengali words. Tsser-act engine receives the processed image. Tesser-act is an oci engine that scans input photos and identifies the letters and characters contained therein.Text files are the product of tesser-act. The same words that form sentences on the image are included in this text file. The Python gtts module converts text files into audio by using the Google Text to Speech CoversionApi. An mp3 audio file is stored in the specified directory. A sound player is need to listen to the sound file..

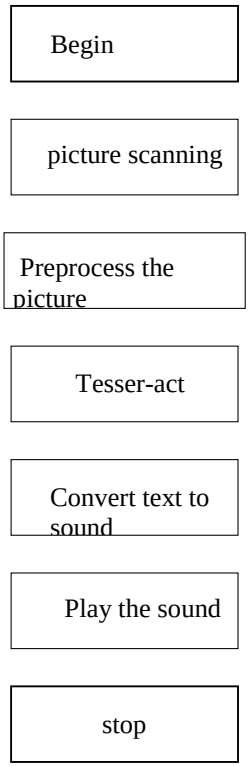


Fig. 1: Flow of the proposed procedure

III. PROCEDURES

This gadget requires multiple techniques to be implemented. The techniques include:

- Developing hardware
- Capturing and processing images
- Performing tesser-act
- Producing sound files
- Playing sound

A. Developing Hardware

The entire process is carried out on a raspberry-pi model. A Linux-based operating system named raspbian is being installed. For the purpose of taking a picture of the printed sheets, a “pi” camera is linked to the raspberry-pi. A decent charger with a high current rating of 1A powers the raspberry-pi. A powerful speaker is connected to the raspberry pi's sound port to play sound. An on/off switch is attached to the general purpose input/output (gpio) pins to turn the device on and off.

B. Capturing andprocessing images

The camera on the Raspberry Pi that is attached to the Usb port takes a picture of the printed document. Remember that the tesseract's output and sound quality are both influenced by the calibre of the photographed scene. A complete, crisp image without any noise produces better results. However, a captured image won't always be crisp and complete. Before sending the image to Tesseract for oci, certain image processing techniques are used to improve it. These processing procedures were used to prepare the image:

- Deskewing
- Despecking
- eliminating lines
- Refining Pictures
- Edge identification

C. *Performing tesser-act*

The identification of Bangla characters will be the project's main implementation problem. Tesser-act-oci performs this. To obtain an editable text file, tesser-act receives the processed image from the previous step. Tesseract is a open-source visual character recognition engine. The instructions from the github source link [10] are used to install it on the Raspberry Pi. In order for the oci Engine to process Bengali inputs, it needs a pretrained data file. Nearly all language-trained data files for character identification are available in Tesseract.The file is called "ben.traineddata" for Bengali and is located in the tess-data directory. This file was created by concatenating a number of different files. When the trained data file is put in a certain spot, Tesseract accurately recognises characters and creates text files. Tesseract's engine operates at a level of accuracy that is typical. For precise matching, this engine may operate extremely well on its library. The source file or test dataset has been carefully incorporated into our system. It is crucial to note that every character in our system must be individually identified so that the oci can recognise it if it appears in the input file. For your convenience, here is a quick explanation of how tesser-act functions [11]:

- Blueprints are examined and kept.
- gathered as Particles of contour data
- Text lines are ordered with particles.
- Words are separated from lines
- The identification process's initial pass tries to identify each word individually.
- Positive feedback given to the adaptive trainer
- Adaptive trainer uses lessons acquired in a second effort to recognize words that were not correctly identified in the initial attempt.
- Text has been verified for small caps and ambiguous spaces fixed.
- Texts produced digitally.

These procedures involve the use of tesser-act:

- Methods for identifying text lines on a page that is slanted
- Algorithms for identifying proportional and non proportional words.
- Methods for disentangling strings of related characters and spotting characters that are broken
- Finding the probably word from a group of characters using linguistic analysis
- Letter classifiers come in two varieties: static classifiers and adaptive classifiers, the latter of which is better at distinguishing between uppercase and lowercase letters since it learns from test examples.

D. *Text toSpeech*

The reading of the documents is the project's primary goal. Therefore, a text file obtained from Tesseract must be converted to sound. gtts handles thatTo create an MP3 file, the Google Text to Speech API's Python interface, gtts, or the command-line utility gtts-cli, are used. By tokenizing lengthy sentences at natural pause points, it enables speech to be uttered indefinitely long. Using

gtts, sound conversion happens quite quickly. Using the correct instructions from the github page, gtts is installed [12].

E. Soundplaying

The sound file found by gtts is played using the free source sound player omx-player. The Raspberry Pi's omx-player is a sound player that runs from the command line It was developed to test the Xbms Raspberry PI implementation, but when used on its own, it proves to be quite helpful. Omx-player utilizes the Open-max (omx) special hardware interface (API), the media API for which the raspberry pi is officially supported [13].

IV. IMPLEMENTATION

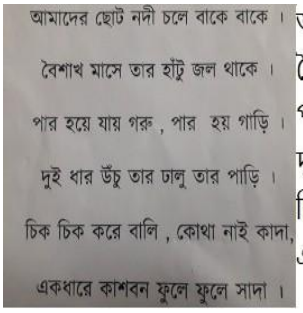
The development of an oci engine requires extensive large-data training and the use of sophisticated detection methods. Therefore, it is preferable to use the most efficient engines, To identify characters and convert text to sound, use gTTS and tesseract-oci. A speaker acting as the raspberry-pi's audio output is linked to the audio port, and a webcam is connected to the Usb connection. Small raspberry-pi based appliances are used to deploy this technology. The operation is initiated by connecting a switch. Using the command line interface is highly advantageous given that the operating system is based on Linux. The commands for the text-to-speech and oci engines are run on a linux terminal.

The Google Word to Voice API is utilised by the Word to Audio Engine Python module. Consequently, a reliable internet connection is needed. The WiFi connectivity of the Raspberry Pi 3 model B resolves that. Raspbian's network configuration file is set up for a particular hotspot provided by an Android phone. An automated startup script is built to begin the entire operation. The startup script runs automatically after the Pi is powered on and the start button is depressed. Start script is a series of successively executed linux command-line commands. The commands in the start script carry out the subsequent actions in order: Check an Android phone's network connectivity, snap a picture, run oci, turn a word document into an audio file, and broadcast the audio file. Following reading the first page of a hard copy, the script repeats the same procedure. The book will continue as long as the audience wishes to read it in a loop. Simply touching the switch will turn the device off if you want to stop reading.

Fig. 2: Gadget made with a raspberry-pi

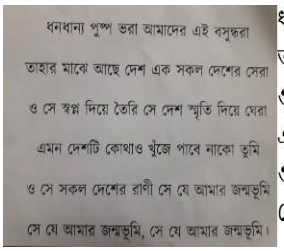
V. RESULT ANALYSIS

The effectiveness of the device's character identification and text-to-sound translation functions determines the results. The clarity and resolution of the image affect the device's performance accuracy. Therefore, increased camera quality, picture capture capability, and image resolution are all necessary for greater accuracy. Several printed paper scripts in Bengali are used to test the device. A text file is produced by the tesser-act engine. The number of right words extracted from that text file's output is used to calculate the character identification system's accuracy. Words that don't match the original printed texts are regarded as being incorrect. Three printed bangla scripts are used in the device's accuracy testing. The text file in Fig. (3-5) contains the tesser-acted output of the written Bengali language characters.



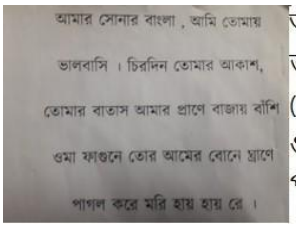
আমাদের ছোট নদী চলে বাকে বাকে 1
বৈশাখ মাসে তার হাঁটু জল থাকে 1
পার হয়ে যায় গরু , পার হয় গাড়ি ...
দুই বার উঁচু তার ঢালু তার পাড়ি 1
চিক চিক করে বালি , কোথা নাই কাদা,
একধারে কাশবন ফুলে ফুলে সাদা 1

Fig. 3: First picture and tesser-acted output



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 জন্মক্টট্ '*

Fig. 4: Second picture and tesser-acted output



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

Fig. 5: Third picture and tesser-acted output

Correct words are manually counted from the text file. The table below shows the number of words from the three figures above:

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 measures the proportion of correctly spelled words to all the words in the table.

6: display of correct and incorrect words using graphics

VI. CONCLUSION

This gadget for visually impaired bengali speakers has been effectively implemented. Although it was specifically designed for those with vision impairments, this device has other uses as well. It is running without a hitch, and accuracy is excellent. There is a slight increase in total execution time, but it does not exceed two minutes. It is as a result of the Raspberry pi's processing power. It is reasonable to assume that as the raspberry-pi's processor develops, the processing time will soon decrease. The versatility of this device can be increased by creating an offline text to speech engine

REFERENCES

[1] B. B. CHAUDHURI and U. PAL, *"A complete printed Bangla ocisystem",Patternidentification,Vol.31,No.,pp.531549,1998*

[2] ShamimAhmed,MohammodAbulKashem,*"EnhancingtheCharacter Segmentation Accuracy of Bangla oci using BPNN", International JournalofScienceandResearch(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 Imagesoci".*

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

[5] Md. AbulHasnat, MuttakinurRahman, ChowdhuryMumit Khan, *"An open source Tesseract*

based Optical Character Recognizer for Bangla script", 10th International Conference on Document Analysis and identification,pp. 671-675 ,2009

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

[7] S.Mukherjee,ShyamalKumarDasMandal,*"ABengaliSpeech*

[8] P Ramprakash, M Sakthivadivel, N Krishnaraj, J Ramprasath. "Host-based Intrusion Detection System using Sequence of System Calls" International Journal of Engineering and Management Research, Vandana Publications, Volume 4, Issue 2, 241-247, 2014

[9] N Krishnaraj, S Smys."A multihoming ACO-MDV routing for maximum power efficiency in an IoT environment" Wireless Personal Communications 109 (1), 243-256, 2019.

[10] N Krishnaraj, R Bhuvanesh Kumar, D Rajeshwar, T Sanjay Kumar, Implementation of energy aware modified distance vector routing protocol for energy efficiency in wireless sensor networks, 2020 International Conference on Inventive Computation Technologies (ICICT),201-204

[11] Ibrahim, S. Jafar Ali, and M. Thangamani. "Enhanced singular value decomposition for prediction of drugs and diseases with hepatocellular carcinoma based on multi-source bat algorithm based random walk." Measurement 141 (2019): 176-183. https://doi.org/10.1016/j.measurement.2019.02.056

[12] Ibrahim, Jafar Ali S., S. Rajasekar, Varsha, M. Karunakaran, K. Kasirajan, Kalyan NS Chakravarthy, V. Kumar, and K. J. Kaur. "Recent advances in performance and effect of Zr doping with ZnO thin film sensor in ammonia vapour sensing." GLOBAL NEST JOURNAL 23, no. 4 (2021): 526-531. https://doi.org/10.30955/gnj.004020 , https://journal.gnest.org/publication/gnest_04020

[13] N.S. KalyanChakravarthy, B. Karthikeyan, K. Alhaf Malik, D.BujjiBabbu,. K. NithyaS.Jafar Ali Ibrahim , Survey of Cooperative Routing Algorithms in Wireless Sensor Networks, Journal of Annals of the Romanian Society for Cell Biology ,5316-5320, 2021

[14] Rajmohan, G, Chinnappan, CV, John William, AD, ChandrakrishanBalakrishnan, S, AnandMuthu, B, Manogaran, G. Revamping land coverage analysis using aerial satellite image mapping. Trans Emerging Tel Tech. 2021; 32:e3927. https://doi.org/10.1002/ett.3927

[15] Vignesh, C.C., Sivaparthipan, C.B., Daniel, J.A. et al. Adjacent Node based Energetic Association Factor Routing Protocol in Wireless Sensor Networks. Wireless PersCommun 119, 3255–3270 (2021). https://doi.org/10.1007/s11277-021-08397-0.

[16] C ChandruVignesh, S Karthik, Predicting the position of adjacent nodes with QoS in mobile ad hoc networks, Journal of Multimedia Tools and Applications, Springer US,Vol 79, 8445-8457,2020

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

[18] K. M. AzharulHasan, Muhammad Hozaifa,SanjoyDutta, RafsanZaniRabbi, "A Framework for Bangla Text to Speech Synthesis", 16th International Conference on Computer and Information Technology (ICCIT),pp.60-64,2013

[19] https://github.com/tesseract-oci, Last accessed: August 14,2017.

[20] https://tesseract-oci.repairfaq.org/downloads/saltcymrudocument5.pdf, Last accessed: August

14, 2017.
[21] https://github.com/pndurette/gTTS,Lastaccessed:August14,2017.
[22] https://www.raspberrypi.org/documentation/raspbian/applications/omxplayer.md, Last accessed: August 14, 2017.
[23] M. Quwaider and Y. Jararweh, "Cloudlet-based efficient data collection in wireless body area networks," Simulation Modelling Practice and Theory, vol. 50, pp. 57–71, 2015.
[24] K. Dongre, R. S. Thakur, A. Abraham et al., "Secure cloud storage of data," in Computer Communication and Informatics (ICCCI), 2014 International Conference on. IEEE, 2014, pp. 1–5.
[25] M. S. Hossain, G. Muhammad, M. F. Alhamid, B. Song, and K. Al- Mutib, "Audio-visual emotion recognition using big data towards 5g," Mobile Networks and Applications, pp. 1–11, 2016.
[26] J. Chen, K. He, R. Du, M. Zheng, Y. Xiang, and Q. Yuan, "Dominating set and network coding-based routing in wireless mesh networks," IEEE Transactions on Parallel and Distributed Systems, vol. 26, no. 2, pp.423–433, 2015.
[27] L. M. Kaufman, "Data security in the world of cloud computing," Security & Privacy, IEEE, vol. 7, no. 4, pp. 61–64, 2009.
[28] R. Lu, X. Lin, and X. Shen, "Spoc: A secure and privacy-preserving opportunistic computing framework for mobile-healthcare emergency," Parallel and Distributed Systems, IEEE Transactions on, vol. 24, no. 3, pp. 614–624, 2013.
[29] J.-J. Yang, J. Li, J. Mulder, Y. Wang, S. Chen, H. Wu, Q. Wang, and H. Pan, "Emerging information technologies for enhanced healthcare," Computers in Industry, vol. 69, pp. 3–11, 2015.
[30] K. A. Khan, Q. Wang, C. Luo, X. Wang, and C. Grecos, "Comparative Study of internet cloud and cloudlet over wireless mesh networks for realtime Applications," in SPIE Photonics Europe. International Society for Optics and Photonics, 2014, pp. 91 390K.
[31] Rajendran, P. K., Muthukumar, B., &Nagarajan, G. "Hybrid Intrusion Detection System for Private Cloud": A Systematic Approach. Procedia Computer Science, 48,pp.325–329, (2015).
[32] Raj Scholar, A. P., & Rani Assistant professor, S. M. "Behavior Rule Specification-based Intrusion Detection for Safety Critical Medical Cyber PhysicalSystems : A Review. International Journal of Computer Applications.
[33] Shi, Y., Abhilash, S., & Hwang, K. "Cloudlet Mesh for Securing Mobile Clouds from Intrusions and Network Attacks". In 2015 3rd IEEE International Conference on Mobile Cloud Computing, Services, and Engineering (pp. 109–118).(2015).
[34] Sajjad, S. M., Bouk, S. H., &Yousaf, M. Neighbor Node Trust based Intrusion Detection System for WSN. Procedia Computer Science, 63, pp.183–188, (2015).
[35] Vasilomanolakis, E., Karuppayah, S., Mühlhäuser, M., & Fischer, M. "Taxonomy and Survey of Collaborative Intrusion Detection". ACM Computing Surveys, 47(4), pp.1–33. (2015).