



Voice Conversion Technique Used for Text Document Using Supervised Machine Learning

¹Aditya Pai H, ²Jagadish Chandola, ³Abhishek Jain

¹Associate Professor, Graphic Era (Deemed to be university)

² Research Scholar, Graphic Era (Deemed to be university)

³ Department of Computer Applications, Graphic Era Hill University, Dehradun, India.

ABSTRACT

Communication is a basic process while sharing information. Documentation and reports are being generate in a huge manner for individuals as well as groups of individuals. To share a documented information in groups like classes or other offices is seems more efficient when we do it using orally. The sharing of the voice information in the module has certain disadvantages where it is for the people with hearing defect such as deaf people and thus can help them a lot to improve the communications by using sign and text method. They are required to evaluate every aspect for the conversion of the text in the form of the speech this will then help the people with hearing and speaking problem. To achieve this method smoothly we are required to convert the text to voice so that all the problems are solved for the people in the real world scenario.

Key words: automatic text converter, android application, LCD, text extraction.

1. INTRODUCTION

1.1 OVERVIEW

Documentation and file sending receiving have been a part of professional in everywhere. We receive a file and means it with reading and while sending a file we write it.so in the process we look two reading and writing work from persons. But if we need a support to share the document in a group we go through oral explanation. So here oral explanation become an information. So, we can communicate in oral and text. But the voice communication is effective where we want to share information in group. These groups may be in schools, some organization, and offices of some other institutions. Otherwise, we could use this information to help the needed person etc.

We see that text document and voice can share information. But there are some more lacks are there for different places with this information. It means we want it in more interesting manner, would it be possible if we cannot read a text but understand it with voice. It seems to be amazing if it would become possible for us. So, our aspect here is to focus on a requirement of voice creator of a text document of a single page or a spiral document with

hundred pages. So that a reader can get information of a document in voice format. In same fashion we can think of a reader reads a book of his choice with audio. But the most prominent places to use by a visually impaired person who is not capable of reading. This encourages them to actively involve themselves with others, for instance, the braille method is used for reading and writing. Suppose when a visually impaired person wants to read a physical book and it is not available in audio format, which is true for most cases, then he/she needs to get help from a third party to read it. To avoid these inconveniences incurred we propose a suitable model which will assist visually impaired people to read a book without any human assistance. They are come many places where these requested features meet more capabilities of working together or becoming capabilities of visually impaired one. The previously available scanner makes this model more cultured while a scanned image text extract the text and a text to speech synthesizer is helping in voice producing and a turner turning a page. So, our proposed model requires to meet all aspect in our work. We required of Optical Character Recognition (OCR) (text of the images) and a Text to Speech Synthesizer (TTS).

1.2 PURPOSE OF PROJECT

World health organization (WHO) present a data about vision impaired person, there are 2.4 billion individuals who have vision impairment. Numerous intelligent systems have been introduced by our scientists to assist visually impaired people in difficult ways. The previously available scanner makes this model more cultured while a scanned image text extract the text and a text to speech synthesizer is helping in voice producing and a turner turning a page. So, our proposed model requires to meet all aspect in our work. We required of Optical Character Recognition (OCR) (text of the images) and a Text to Speech Synthesizer (TTS). We see that text document and voice can share information. But there are some more lacks are there for different places with this information. It means we want it in more interesting manner, would it be possible if we cannot read a text but understand it with voice. It seems to be amazing if it would become possible for us. So, our aspect here is to focus on a requirement of voice creator of a text document of a single page or a spiral document with hundred pages. So that a reader can get information of a document in voice format. In same fashion we can think of a reader reads a book of his choice with audio. But the most prominent places to use by a visually impaired person who is not capable of reading. This encourages them to actively involve themselves with others, for instance, the braille method is used for reading and writing. Suppose when a visually impaired person wants to read a physical book and it is not available in audio format, which is true for most cases, then he/she needs to get help from a third party to read it. To avoid these inconveniences incurred we propose a suitable model which will assist visually impaired people to read a book without any human assistance. They are come many places where these requested features meet more capabilities of working together or becoming capabilities of visually impaired one. The previously available scanner makes this model more cultured while a scanned image text extract the text and a text to speech synthesizer is helping in voice producing and a turner turning a page. So, our proposed model requires to meet all aspect in our work. We required of Optical Character Recognition (OCR) (text of the images) and a Text to Speech Synthesizer (TTS). This encourages them to actively involve themselves with others, for instance, the

braille method is used for reading and writing. Suppose when a visually impaired person wants to read a physical book and it is not available in audio format, which is true for most cases, then he/she needs to get help from a third party to read it. To avoid these inconveniences incurred we propose a suitable model which will assist visually impaired people to read a book without any human assistance.

1.3 SCOPE OF THE PROJECT

The scope of this project may follow any of the below:

- to know that Automatic turning of the pages in a document/book and scanning of text is made possible
- To build a voice-over friendly navigation system using Android Application with captured images.
- Voice is built from text to user interface.
- To build a user-friendly application for future.

2. METHODOLOGY AND DESIGN

An automatic page turner is needed. as a helpful device for people who suffer from disabilities and includes people whose upper-body function is limited due to their disabilities. A page turner can be useful for enthusiastic readers as it provides suitable support includes page turning features. We see that text document and voice can share information. But there are some more lacks are there for different places with this information. It means we want it in more interesting manner, would it be possible if we cannot read a text but understand it with voice. It seems to be amazing if it would become possible for us. So, our aspect here is to focus on a requirement of voice creator of a text document of a single page or a spiral document with hundred pages. So that a reader can get information of a document in voice format. In same fashion we can think of a reader reads a book of his choice with audio. But the most prominent places to use by a visually impaired person who is not capable of reading. This encourages them to actively involve themselves with others, for instance, the braille method is used for reading and writing. Suppose when a visually impaired person wants to read a physical book and it is not available in audio format, which is true for most cases, then he/she needs to get help from a third party to read it. To avoid these inconveniences incurred we propose a suitable model which will assist visually impaired people to read a book without any human assistance. They are come many places where these requested features meet more capabilities of working together or becoming capabilities of visually impaired one. The previously available scanner makes this model more cultured while a scanned image text extract the text and a text to speech synthesizer is helping in voice producing and a turner turning a page. So, our proposed model requires to meet all aspect in our work. We required of Optical Character Recognition (OCR) (text of the images) and a Text to Speech Synthesizer (TTS). The pages will be turned automatically by the turner once it is in running mode. An automatic page scanner is useful in scanning a document and compile the scanned image to the android application. An Android Application is a software which is refers to an Android package. Languages used to build the Android based application can be Java and C++. They are then run inside Virtual

Machine. The Google Cloud Vision API, Google Cloud Vision OCR, and Google Cloud Text-to-Speech Are another part of Google Vision API. Ionic open-source UI toolkit which is used for building high-quality desktop and mobile applications.

2.1 DESIGN OF THE SYSTEM

The design of the system is to ensure the different architectures are defined, where the components and the different modules and its interfaces and for the system data to ensure the requirements of the clients are met. The system design can be seen as the ensuring that the system theory is applied to the development of the product. The proposed model comprises an Automatic Page Turner, page Scanner, Android Application. Then OCR and TTS play are used in converting the images to text to speech. The prototype turner can turn the pages of the book smoothly without causing any damage to the page and the scanner scans the pages of the book simultaneously. At the time of scanning the pages of the book, the pages will not get skipped. Also, books or documents can easily be converted into digital form by using the automatic turner. The android application can convert the image set(folder) obtained into an audiobook. We see that text document and voice can share information. But there are some more lacks are there for different places with this information. It means we want it in more interesting manner, would it be possible if we cannot read a text but understand it with voice. It seems to be amazing if it would become possible for us. So, our aspect here is to focus on a requirement of voice creator of a text document of a single page or a spiral document with hundred pages. So that a reader can get information of a document in voice format. In same fashion we can think of a reader reads a book of his choice with audio. But the most prominent places to use by a visually impaired person who is not capable of reading. This encourages them to actively involve themselves with others, for instance, the braille method is used for reading and writing. Suppose when a visually impaired person wants to read a physical book and it is not available in audio format, which is true for most cases, then he/she needs to get help from a third party to read it. To avoid these inconveniences incurred we propose a suitable model which will assist visually impaired people to read a book without any human assistance. They are come many places where these requested features meet more capabilities of working together or becoming capabilities of visually impaired one. The previously available scanner makes this model more cultured while a scanned image text extract the text and a text to speech synthesizer is helping in voice producing and a turner turning a page. So, our proposed model requires to meet all aspect in our work. We required of Optical Character Recognition (OCR) (text of the images) and a Text to Speech Synthesizer (TTS). The controls include capturing the voice input/instruction, and conversion from voice data to text, processing the text. The Voice Over Friendly Navigation Controls includes options like START, STOP, PAUSE, RESUME, CHANGE VOICE and many other commands. It works according to the voice input taken from the user. Hence this model helps the visually impaired person to hear the book in an interactive way.

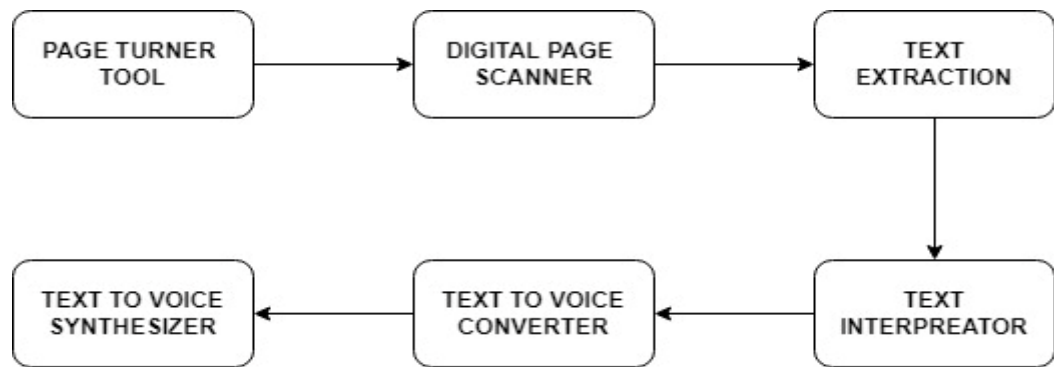


Fig. 1 System Design

2.2 IMPLEMENTATION

These are the step by step implementation of process can be proceed in the following sequence:

- The steps involved in the procedure are as follows:
- The user must keep a spiral document in the Automatic page turner and turn on the camera device to open the scanner in the application.
- Automatic page-turner will turn exactly one page and send a signal to android application.
- Camera takes the pictures of the document page and save the images in the folder of android application.
- OCR is used for text extraction
- TTS is used in text to speech producing
- Users can work with different voice commands to navigate through the document of interest.

There are some Modules also used for building our component:

- Turning Module
- Scanning Module
- Relay Module switches which can open and close the circuits
- LCD, the display module

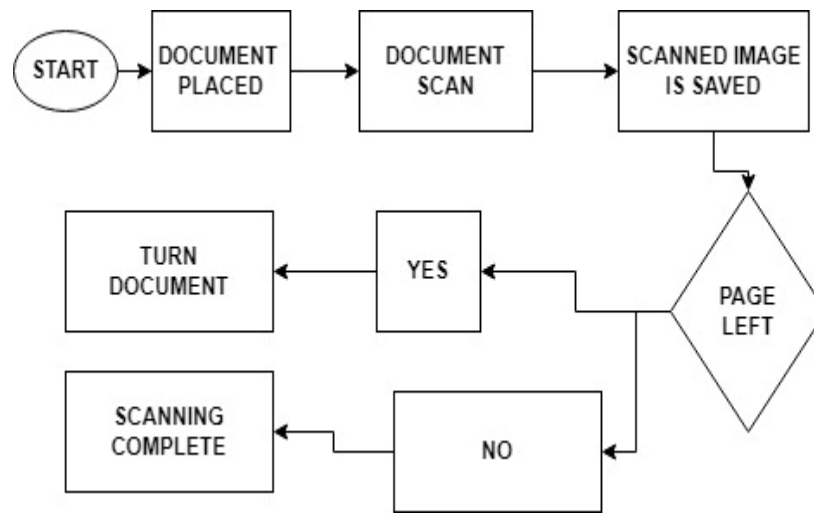


Fig. 2 Process flow of Scanner and turner

Login page is created for the end users to use the system. User Manual is also presented to guide the new users for the use. The manual is presenting the status of all inside processes that are needed to operate the system in working time. Document Folder on Screen is also presented to view to user about their working file document if exist initially. Load Camera Command open the Scanner to scan page starting from. New created folder containing images captured using load command. Text Extraction is done by synthesizer by providing token to text. Some important command, Start Command, Pause Command, Search Command, Read Slow/Read Fast Command are also given to operate in a controlled manner as requested.

3. LITERATURE REVIEW

Yoshihiro Watanabe had developed an Automatic page-turner which is used for turning books at very high speed. It can turn the pages at very high definition at over 250 pages/min using the original media format. This system has been in practical use since 2013[1]. Ellaine Abueg has developed a switch-activated page-turner that has two gear motors of which one activates the page-lifting arm while the other activates the timing belt controlled wiping arm[2]. Daniel Reetz has developed the Big Finger Little Finger scanner prototype. It uses one rod with an eraser at the end to push the page (and possibly separates it) and then another rod near the spine of the book pushes the page over, then a picture is taken. The book lies in a V-shaped stand[3]. Treventus presents ScanRobot which pulls both facing pages to a document scanner bar that then lifts up and scans the pages. They are then both swept in one direction, turning the page and allowing the device to scan another pair[4]. The Page Flipper is a cost effective, simple device that can flip pages of a book one page at a time in both directions which was visualized by Bhavingawali[5]. Google developers proposed a method to search for a book on the internet and download it[6]. Clem has proposed a method for Sending images from Raspberry Pi to Android Applications with code[7]. Dominic Mazzoni proposed a method to describe the images to visually impaired people using AI[8]. Voice commands were used in the development of a prototype at the digital library of the Department of Mathematics in Usmanu Danfodiyo University, Sokoto for reading books[9].

N Bano worked on an Automatic Page scanner to preserve the books and documents using digitalization[10]. S Venkateswarlu worked on Text to Speech conversion which used an innovative method that helps the user to hear the contents of text images instead of reading through them. It makes use of OCR and TTS in Raspberry pi[11]. Kaz Sato designed a method for converting PDF to Audiobook using machine learning [12]. Aditya Dhapola in his article introduced a method for voice recognition to perform tasks done on a daily basis[13]. Bashier Suliman Elbashier Ahmed and Huwaida Tagelsir Elshoush had worked on Biometrics Solutions for Visually impaired persons in their Conference Paper[14]. Mandeep Singh Kandhari developed Voice-controlled E-Commerce Web Application[15]. Scan.it web application is a Natural Language Processing based system that is used for text recognition. The code was implemented using Node.js. Tesseract OCR is used to recognize the text and In Translator is used to translate this extracted text to speech [16].

4. RESULTS

The automatic page-turning and scanning system has high accuracy which is up to 90%. The automatic turner lifts only one page at a time, and simultaneously clicks pictures, to save operational time. The manual is presenting the status of all inside processes that are needed to operate the system in working time. Document Folder on Screen is also presented to view to user about their working file document if exist initially. Load Camera Command open the Scanner to scan page starting from. New created folder containing images captured using load command. Text Extraction is done by synthesizer by providing token to text. Some important command, Start Command, Pause Command, Search Command, Read Slow/Read Fast Command are also given to operate in a controlled manner as requested. The most important thing to note is that this system is fully automated. This device is capable of precisely flipping the page automatically, one at a time without any misbehaviour.

- 1) Text is extracted from the images and is converted to audio format. It recognizes both capital and small letters, numbers, and different symbols. The text extraction has achieved an accuracy rate of up to 90% in our application.
- 2) A voice assistant is used to take voice input from the user and perform the tasks accordingly.
- 3) Many features are provided like authentication facility, searching for a particular book on the internet, instruction manual feature and identifying and recognizing the text and voice up to 42 languages which are supported by Google Vision API.

5. CONCLUSIONS

The prototype turner can turn the pages of the book smoothly without causing any damage to the page and the scanner scans the pages of the book simultaneously. At the time of scanning the pages of the book, the pages will not get skipped. Also, books or documents can easily be converted into digital form by using the automatic turner. The android application can convert the image set(folder) obtained into an audiobook. The application will have a built voice assistant which can be used to perform many tasks given by the user. The manual is presenting the status of all inside processes that are needed to operate the system in working time. Document Folder on Screen is also presented to view to user about their working file document if exist initially. Load Camera Command open the Scanner to scan

page starting from. New created folder containing images captured using load command. Text Extraction is done by synthesizer by providing token to text. Some important command, Start Command, Pause Command, Search Command, Read Slow/Read Fast Command are also given to operate in a controlled manner as requested. Users are allowed to navigate throughout the application using voice commands which are implemented through Voice Over Friendly Navigation System. It has features like an instruction manual, and provides features where it can identify the text and output the voice in different languages, and a search engine for finding the book already available on the internet.

- The scanner can possess extra features in the coming future like: i) The users can assign the range for scanning the number of pages. ii) Voice assistance can be added to the scanner and turner as well, which takes instructions from the user directly.
- The application could even provide a small description whenever an image is encountered.
- In the future, the application can even have in-built features such as setting timers, highlighting special sentences, and providing vocabulary to find the meaning of the word.
- The device could even read comic books with different voice features.

REFERENCES

- [1] Aditya Pai, VS Veeram, BS Babu, Piyush Kumar Pareek. Six sigma approaches used in implementing in supply chain management: A Review. Research and Applications of Web Development and Design 1 (2): pp. 12-16, 2018.
- [2] A Pai, VS Veeram, Piyush Kumar Pareek, BS Babu , Challenges in SME's ANOVA ANALYSIS PART-2 in Bengaluru , Research and Reviews: Advancement in Robotics 1 (1), 9-15, 2018.
- [3] A Pai, VS Veeram, BS Babu, Piyush Kumar Pareek, ANOVA Analysis Part One of Challenges in SME'S in Bengaluru , Research and Reviews: Advancement in Robotics 1 (1), 1-8, 2018.
- [4] K Venkata Rao, R Balakrishna, H Aditya Pai, PK Pareek , Data Mining for Healthy Tomorrow with the Implementation of Software Project Management Technique , Artificial Intelligence and Evolutionary Computations in Engineering Systems, 345-355, 2016.
- [5] SG Gollagi, N Murthy, A Pai, K Swathi, PK Pareek, An Empirical Analysis of security in SME'S in Bengaluru, International Journal of Management (IJM), 11, pp. 2170-2177, DOI: 10.34218/IJM.11.11.2020.222, 2020.
- [6] S. Gollagi, K. Piyush, Pareek, H Pai, Aditya, P. Pareek, Study of software security measures adaption in SME'S in Bengaluru, International Journal of Advanced Research in Engineering and Technology (IJARET), 11 (11), pp. 2170-2177, 2020.
- [7] Aditya Pai H, Sameena HS, Sandhya Soman, Piyush Kumar Pareek, Analysis of causes and effects of longer lead time in software process using FMEA, (May 17, 2019). Proceedings of the Second International Conference on Emerging Trends in Science & Technologies For Engineering Systems (ICETSE-2019), Available at SSRN: <https://ssrn.com/abstract=3508574> or <http://dx.doi.org/10.2139/ssrn.3508574>

- [8] A Pai, BN Ramesh, PK Pareek, P Prasoon, Failure mode effective analysis for software processes, Int. J. Res. Eng. IT Soc. Sci 8 (10), 2250-0588, 2018.
- [9] Aditya Pai, Venkata Srinivasu Veeram, Bandaru Satish Babu, Piyush Kumar Pareek, Understanding the Adaptability of SME'S in Bengaluru, Research and Reviews: Advancement in Robotics, 1 (1), pp. 16-22, 2018.
- [10] Aditya Pai, Venkata Srinivasu Veeram, Bandaru Satish Babu, Piyush Kumar Pareek, Comparative analysis of conventional approach and automated approach used in lean six sigma while implementing in supply chain management: a review, Research and Applications of Web Development and Design, 1 (1), pp. 7-13, 2018.