DEVELOPMENT OF AN INTELLIGENT TEXT-TO-SPEECH MODEL FOR VISUALLY IMPAIRED STUDENTS USING OPTICAL CHARACTER READER

I. JMGBEAFULIKE (1) **C. C.UGORJI (2)**
[1]Department of Computer Science Chukwuemeka Odumegwu Ojukwu University, Uli
[2]Department of Computer Science Abia State Polytechnic, Aba

ABSTRACT
Over the years the visually impaired students have suffered a lot owing to the fact that they could hardly visualize very well what is being typed or written as text. This simply implies that it will be an exigent task for visually impaired students to read text from scene, images and text boards. They compel themselves to see typed text or written text but at the end of the day a lot of mistakes are being committed in the process. The aim of this research is to design and implement an intelligent text-to-speech model for visually impaired students using an optical character reader. With this new model, the visually impaired can scan text or take note which can be voiced out and also saved for a repeat play. The object-oriented analysis and design methodology (OOADM) was used to analyze and design the app. The application was developed using PHP (hypertext processor) and HTML (Hypertext Markup Language) and is deployed in an android mobile device for use.

Keywords: Optical character recognition, Binarization, Text-to-speech conversion, visually impaired students.

1. INTRODUCTION

A text-to-speech synthesis is the automatic conversion of a text into speech that resembles, as closely as possible, a native speaker of the language reading the text. It is a technology which lets computer speak to you (Wasala etal., 2006). The text-to-speech system gets the text as the input and then a computer algorithm analyzes the text, pre-processes the text and synthesizes the speech (Gupta etal., 2007). The TTS engine usually generates sound data in an audio format as the output. The text-to-speech (TTS) synthesis procedure consists of two main phases. The first is text analysis, where the input text is transcribed into a phonetic or some other linguistic representation, and the second one is generation of speech waveforms, where the output is produced from this phonetic and prosodic information (Zhang, 2004). These two phases are usually called high and low level synthesis. The input text might be for example data from a word processor, standard ASCII from email, a mobile text message or a scanned text from a newspaper. Application of text-to-speech synthesizers is numerous. The application can be used with learning difficulties, people who speak the language but cannot read it, people with visual impairments, people who access contents on mobile devices, people who multi-task and even by people with literary difficulties (Itunuoluwa et al., 2014). This paper is concerned with developing an intelligent text-to-speech model using an optical character reader for the visually impaired students which they can use to scan typed or written text that can be voiced out and saved for a repeat play.

2. LITERATURE REVIEW

2.1 Review of Related Works

A presentation of general review of previous work on text-to-speech will be carried out here. There exists some research works for assisting visually impaired persons/students with text-to-speech technology. A useful text-to-speech synthesizer in the form of a simple application has

been implemented for visually impaired people. It was developed in the form of a simple application that possesses the capability of converting inputted text into synthesized speech and reads it out to the user. The text-to-speech (TTS) synthesis consists of two main phases which are text analysis and generation of speech waves (Itunuoluwa et al., 2014).

Another implementation was on text-to-speech system using optical character recognition (OCR) that was able to recognize text character and convert it into speech signal. The text contained in the page was first pre-processed. The pre-processing module prepares the text for recognition. This is followed by the segmentation of the text to separate the character from each other. The segmentation process was followed by extraction of letters and resizing them and storing them in a text file. The software platforms used include MATLAB, LabVIEW and Android (Jisha etal., 2015).

Another set of people implemented a text-to-speech conversion for Hindi language that was able to voice out single text-to-speech system for Indian languages viz: Hindi. The design generally involves two steps: text processing and speech generation. A graphical user interface was designed for converting Hindi text-to-speech in Java Swings (Kaveri etal., 2012).

The next set of people implemented a text-to-speech conversion using Raspberry PI. The combined the feature of optical character recognition and text-to-speech synthesizer in Raspberry PI. Their design consists of two modules: image processing module and voice processing module. Their device was developed based on Raspberry PI V2 with 900MHz processing speed. Their design has a small inbuilt camera that scans the text printed on a paper, converts it to audio format using a synthesized voice for reading out the scanned text (Lakshmi, 2016).

2.2 Concept of Speech Synthesis

Speech synthesis can be described as the artificial production of human speech (Suendermann etal., 2010). The computer system used for this purpose is called speech synthesizer, and can be implemented in software and hardware. A text-to-speech (TTS) system converts normal language text into speech (Allen et al., 1987).

2.3 Speech Synthesis Techniques

A text-to-speech system (or “engine”) is composed of two parts (Van Santen etal, 1997) a front-end and a back-end. The front-end has two major tasks. First, it converts raw text containing symbols like numbers and abbreviations into the equivalent of written-out words. This process is often called text normalization, pre-processing, or tokenization. The front-end then assigns phonetic transcriptions to each word, and divides and marks the text into prosodic units, like phrases, clauses, and sentences.

2.4 Intelligent Systems

An intelligent system is a system that can imitate, automate some intelligent behavior of human being. Expert systems, intelligent agents and knowledge based systems are good examples of intelligent systems. (Vantruc, 2013). Criteria for intelligence includes the capacity to learn from experience, security, connectivity, the ability to adapt to according to current data and the capacity for remote monitoring and management (Vantruc, 2013).

2.4 Applications of Synthesized Speech

The applications of speech processing (study of speech signals and the processing methods of these signals) are quite diverse. The application can be used by people with learning difficulties,

people who speak the language but cannot read it. People with visual impairment, people who access content on mobile devices, people who multi-task and by people with literary difficulties.

3. PROPOSED SYSTEM

The proposed system is android and mobile based. It will convert text-to-speech using optical character recognition which simply means the recognition of printed or written text character. It follows the following steps:

- Photo scanning of the text character-by-character
- Analysis of the scanned-in image
- Translation of the character image into character codes, such as ASCII

The new proposed system can also allow the visually impaired to say any given word and it will be converted to text.It has the following as the components of the main menu:

- Speak to take note
- Tap microphone to speak
- Convert voice to text
- Scan to take note
- Read scanned text and save

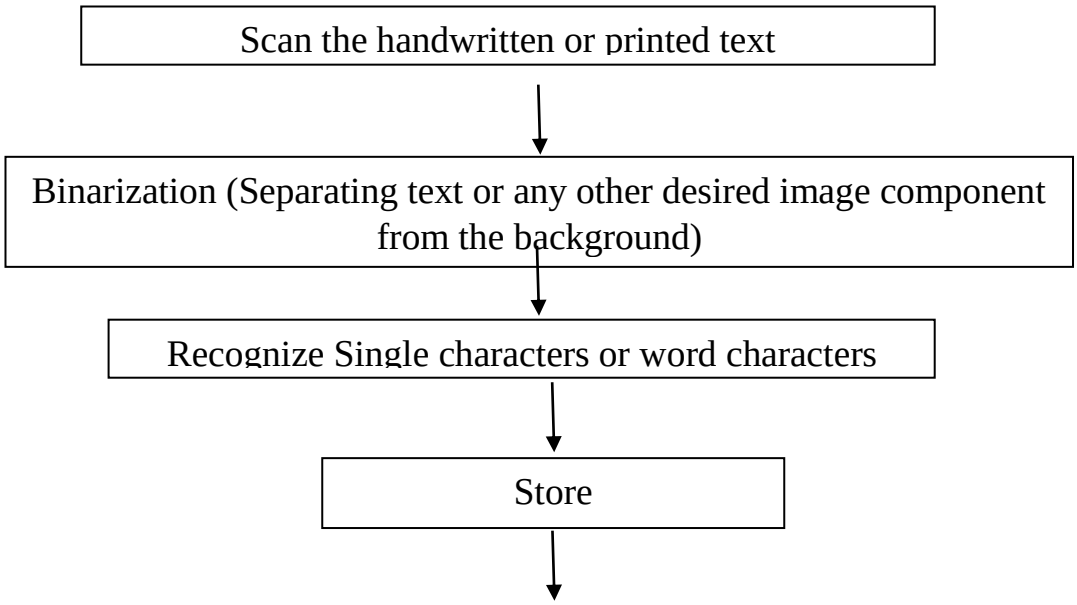
3.1 System Implementation

The system to achieve the set objectives was developed using Java programming language as the front-end. Java was used because it is a portable language and can easily run in all platforms. Sqlite-3 database management system was used as the backend. Android studio was also used along with java.

3.2 Methodology

The methodology used in this research work is the object oriented analysis and design methodology (OOADM). OOADM models “the real world” requirements, independent of the implementation environment. The design applies object-oriented concepts to develop and communicate the architecture and details of how to meet requirements.

3.3 Architecture of the System



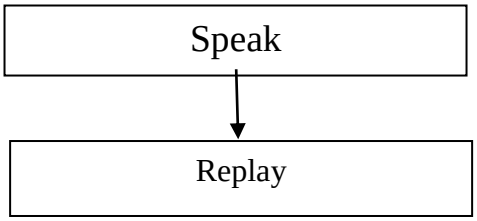


Figure 1: Architecture of the Proposed System

In figure 1, the proposed system will convert text-to-speech by using optical character recognition which is the recognition of printed or written text character. It normally will start by photo scanning of the text character-by-character, analysis of the scanned-in image and then translation of the character image into character codes, such as ASCII, commonly used in data processing. The new proposed system can also allow the user to say any given word and it will be converted to text. The text can be saved for a repeat play.
The following modules make up the new system:

- Take note module: Here, you can either speak or scan to take note
- Tap mic to speak module: This enables you to tap on the microphone to voice out the note you have taken.
- Convert voice to text module: Here, the spoken voice can be converted to text
- Scan to take note module: Note taken can be scanned.
- Scan text module: Text can also be scanned
- Read scanned text and save module: The text is saved after scanning
- Repeat play module: the saved text can be replayed.

The new proposed system can serve dual purposes: both as text-to-speech and speech-to-text for the visually impaired persons.

3.4 Sample Input Sample

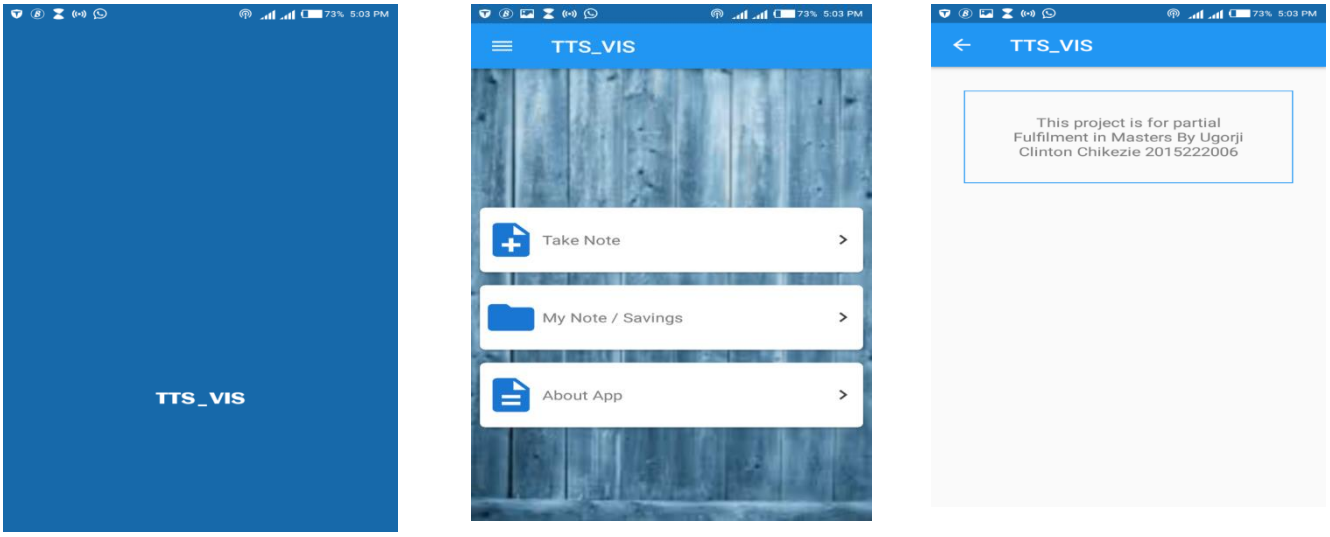


Figure 2: Sample System outputs

4. RESULTS AND DISCUSSION

The proposed system "an intelligent text-to-speech for visually impaired students" converts text-to-Speech which can be saved for a repeat play for visually impaired students.

4.1 Conclusion

Text-to-speech synthesis is a rapidly growing aspect of computer technology and is increasingly playing a more important role in the way we interact with the system and interfaces across a variety of platforms. In this paper the various operations and processes involved in text-to-speech synthesis were identified. The principal aim of this research is to develop a very simple and interactive text-to-speech model using optical character recognition which can scan typed text, voice out typed text and saved it for a repeat play. This intelligent text-to-speech model can run on android and mobile devices

.

REFERENCES

Allen, J., Hunnicutt, M.S., Klatt D., (1987). From Text to Speech: the MITalkSystem, Cambridge University Press.

Gupta. M and Garcia, E. (2007). OCR binarization and image pre-processing for searching historical documents. Pattern Recognition (2), 369. Retrieved 2nd may, 2015

Itunuoluwa, I and Jelili O. (2014). Design and Implementation of Text –To-Speech Conversion for Visually Impaired People. International Journal of Applied Information System (IJAIS) Foundation of Computer Sciences FCS, New York; USA

Jisha, G, Aravid, S., Pooja, C., Saranya, S.S (2015) "Text-to-Speech Conversion using OCR" International journal of Emerging Technology and Advanced Engineering

Kaveri, K., Ramesh K (2012) "translation of Text-to-Speech Conversion for Hindi language. International Journal of Science and Research (IJSR)

Lakshmi, K., SekharRao T.C (2016) "Design and Implementation of Text-to- Speech Conversion using Raspberry Pi" International Journal of Innovative Technology and Research.

Suendermann, D., Hoge, H., and Black, A., (2010). Challenges in Speech Synthesis. Chen, F., Jokinen, K., (eds.), Speech Technology, Springer Science + Business Media LLC.

Vantruc, T., Le Quang, P., Van Thuyen, V., Hieu, L.T., Tuan, N.M., and Hung P.D., (2013). Vietnamese SynthesisSystem, Capstone Project Document, FPT UNIVERSITY.

Van Santen, J.P.H., (1997). Assignment of Segmental duration in text-to-speech synthesis. Computer Speech & Language, Pages 95 – 128

Wasala, A., Weerasinghe R., and Gamage, K., (2006) Sinhala Grapheme-to- Phoneme Conversion and Rules for Schwaepenthesis. Proceedings of the 2006 main Conference Poster Sessions, Sydney, Australia, pp. 890 – 897.

Zhang, J., (2004). Language Generation and Speech Synthesis in Dialogues for Language Learning. Masters Dissertation, Massachusetts Institute of Technology.