Computing, Information Systems & Development Informatics Journal
Vol. 9 No. 4, 2018 - www.cisdijournal.org

Flexible Innovative Learning System for Visually Impaired People

Ojo Abosede Ibironke & Yusuf Ridwan Lanre
Department of Computer Science
Ogun State Institute of Technology
Igbesa, Ogun State
E-mail: ronkujoyme@gmail.com; yusuflarnrey@gmail.com
Phone: +2348027609915; +23407039597818

ABSTRACT

Natural Language Processing (NLP) and Digital Signal Processing (DSP) are two important areas of artificial intelligence that is concerned with interactions between human natural language and computers to successfully process large natural language data. This paper is an innovation that tends to alleviate the challenge the teachers pass through by passing information to the students that are visually impaired in the following areas; 1) by providing educational content in a way that increases student comprehension. 2) by converting text to speech to know exactly what they are typing and inputting to the computer system 3) by listening to audio lectures and textbooks 4) by helping the auditory learners and people who have difficulty in literacy. More so, some related areas such as Text-To-Speech (TTS) conversion system on an Integrated Circuit, Speech Synthesis and Text To Speech (TTS) and Computer Aided Language Learning (CALL) were delved into. Furthermore, observation and investigation were used as the techniques for finding data. TTS architecture was thoroughly examined where text is taken as an input, analyzed, pre-processed and synthesized the speech using statistical algorithms called TTS algorithm. Use Case Diagram was used to depict each of the Actors in the system and screen shots for user interfaces were justifiably considered. Finally, this paper will bring innovation and productivity to both the teachers and the visual impaired people, not only to teach but, to be taught in a relaxed atmosphere.

Keywords: eaching Assistant, Synthesis, Learning, Application, Text, Speech & System

CISDI Journal Reference Format
Ojo, A.I & Yusuf, R.L. (2018): Flexible Innovative Learning System for Visually Impaired People.
Computing, Information Systems & Development Informatics Journal. Vol. 9 No. 4. Pp 37-46.
Available online at www.cisdijournal.net

1. INTRODUCTION

One of the essential sense organ in the body is the eyes. It is paramount in life for somebody to hear, even if he/she cannot see and be able to communicate to others either verbally, orally or even by sign. The well-known visual impairment types are short sightedness and long sightedness, but there are less known visual impairments such as congenital cataracts (where the lens of the eye is cloudy); Strabismus (where the eye looks at different direction); Glaucoma (where increase pressure inside the eye, brings loss of sight), Retinitis Pigmentosa (where inherited disease destroys the retina) etc. Julian (2012). To have an effective teaching with visual impaired people, there is need for a teaching assistant which will aid the learning. As the motto of UNESCO says 'Education for All' proves how education should be embraced, sponsored and make available to citizenry without neglecting anybody.

In fact, the visual impaired and the disabilities should be the most beneficiaries of adequate learning infrastructure which will make learning enjoyable and convenient to them Belay (2015). A teaching assistant is an application that uses computer technology in teaching an individual or group of individuals in a particular topic, subject or concept; hence assisting the teacher in performing his task and students learning under a relaxed and conducive atmosphere. Before now, there are lots of challenges for both the teacher and the students. The students would have to read and reread instruction repeatedly before the students understand and since the students do not have proper vision to read on their own, they find it difficult to read what they have been taught, at their free time. Another challenge was that students with vision impairment find it difficult to know exactly the word they type even when some of them know the keyboard very well (after proper training), still they just assume that they are correct. At the end of the day, they will find themselves making many mistakes in their typed works. This paper provides solutions to the above challenges;

1. by providing educational content in a way that increases student comprehension.
2. by converting text to speech to know exactly what they are typing and inputting in the computer system
3. by listening to audio lectures and textbooks
4. by helping the auditory learners and people who have difficulty in literacy.

2. LITERATURE REVIEW

Earlier to the discovery of electronic signal processing, there were many efforts to create machines that use human voices. Some of those that made attempts were Albertus Magnus (1198–1280), and Roger Bacon (1214–1294). In 1779, models of human vocal tract that could generate five long vowel sounds were built by Christian Kratzenstein. According to Charles (1857), Wheatstone produced a "speaking machine" based on von Kempelen's design, and in 1857, M. Faber built the "Euphonia". Paget resurrected Wheatstone's design in 1923. Bell Labs designed and developed a vocoder in the 1930s. The vocoder automatically analyzes speech into its fundamental tone and resonances. From his work on the vocoder, Homer Dudley developed a keyboard-operated voice synthesizer called The Voder (Voice Demonstrator), which he exhibited at the 1939 New York World's Fair. The Pattern playback was built by Dr. Franklin S. Cooper and his colleagues at Haskins Laboratories in the late 1940s and completed in 1950.

Ever since the introduction of microcomputers, an educational and instructional softwares; teaching softwares or computer-assisted instruction (CAI) has provided an alternative and supplemental instructional method which are used in teaching students in schools. CAI includes more complex programs which incorporate tutorial instruction (Parr & Fung, 2000). Different teaching application softwares have functionalities of record keeping and management systems. However, they are also referred to by a variety of other names, such as Computer Based Instruction (CBI), Computer Assisted Learning (CAL) etc. CBI softwares include tutorials, practice and integrated learning systems and it places more emphasis on individual learning process to accommodate the needs, interests, current knowledge, and learning styles of the students. (Parr & Fung, 2000).

1.1 Speech Synthesis

Speech synthesis is the process of creating speech signals from text by a Text to Speech system. Earlier to the synthesis process, the text is processed using a lexicon, letter-to-sound rules and models that disambiguate part of speech and define supra-segmental features. The result is synthesized using one of the following methods which are; concatentive, articulatory and formant synthesis. The articulatory synthesis is based on a model of the vocal tract while an acoustic model of a human voice is used to generate the speech in the formant synthesis. The articulatory and formant types of synthesis are more flexible than concatentive synthesis due to the fact that they bring together all units of a human speech from an indexed and labeled speech database. According to Black & Lenzo, (2007), the database must be designed in such a way that it will possess enough coverage of all possible phonetic combinations that will appear in the language that is to be synthesized.

Text-To-Speech Conversion System on an Integrated

The translation of text into speech becomes important with the advent of data communication and other new applications. Data is saved in a digital format that can be easily processed and manipulated by digital processing means. There is some data that may need to be converted analog format for speech that is intelligible to humans. Text-to-speech conversion is a process of converting a particular digital format text (e.g., ASCII text) into a particular analog format (e.g., speech) suitable for reception by humans.

Text to Speech (TTS) and Computer Aided Language Learning (CALL)

An attempt to marry two different technologies together gave rise to the integration of text to speech on a computer aided language learning device. The TTS requires less storage space, other than large recorded speech systems such as a voice recorder which must be properly recorded and free from undesired utterances. Text to speech offers lots of benefits over prerecorded human speech. A large volume of unparalleled utterances that do not need to be in any specific preplanned can be created by a TTS. Isewon et al (2014) defined Text-to-speech synthesis as the automatic conversion of a text into speech that resembles, as closely as possible, a native speaker of the language reading that text. Some of the factual features of TTS are to develop new technology advancements that increase students' motivation to learn the phonetic form of words they read. Similarly, it can also be used to develop exercises that require students to take dictation or to match written and auditory forms of words (Carol, 2007). Finally, it is significant that TTS be built into the system so that students' usage can be tracked and combined into the student models.

2. METHODOLOGY

2.1 Method of Data Collection

There exist various methods of gathering facts about a situation. This includes interview, questionnaire, record inspection and observation. Each of these methods has advantages and disadvantages. For the purpose of this study, the following facts finding methods were employed for data collection:-

Observation: Observation method provides firsthand information about the system. It involves systematically watching the operations, procedures and personnel of the school over a period of time and recording things as they happen. This method was adopted in other to see how activities were been done in the research area.

Investigation: - Books, Journals. Online materials, School websites of visual impaired and software diagnosis were consulted for the purpose of this paper and they all serve as a guide to the study.

2.2 Text to Speech Architecture

The text to speech architecture comprises of the Natural Language Processing (NLP) and the Digital Signal processing. The figure 3.1 shows the text-to-speech synthesizer. TTS is a technology that allows the computer speaks to its user by taking text as input and then analyses the text, pre-processes the text and synthesizes the speech using statistical algorithms called TTS algorithm. A text to speech Architecture is depicted in Fig. 1:

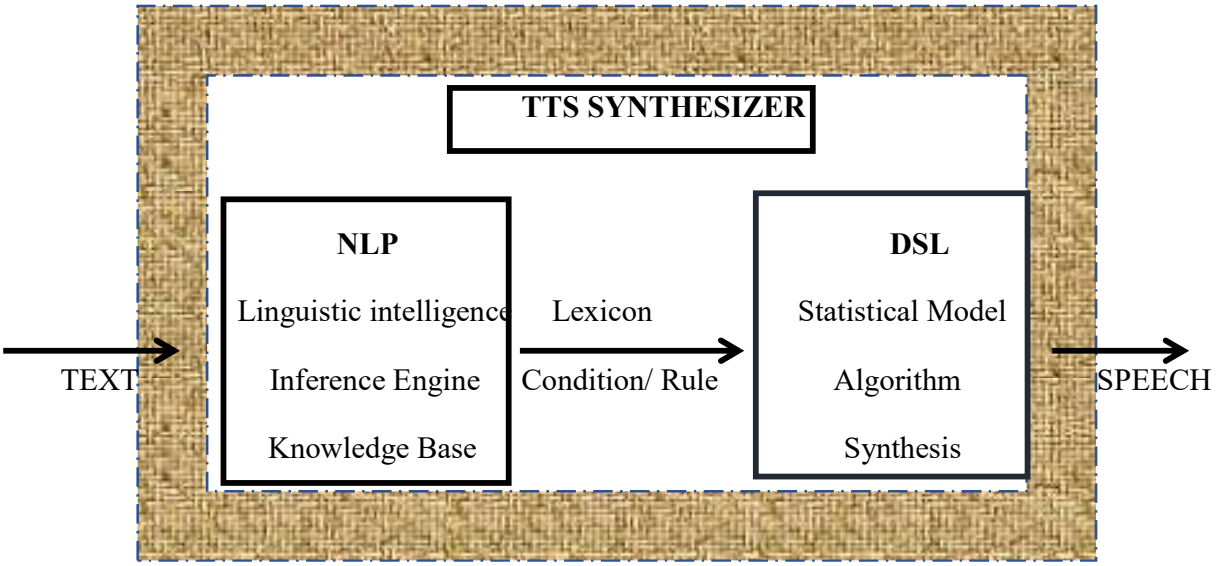


Fig 1: Text to speech Architecture
(Source: Isewon et al (2014))

The architecture depicts how a text is stored, processed and converted into spoken word, by analyzing and processing the text and then converting this processed text into synthesized speech representation of the text in audio formats for the students to listen to at their own convenience.

The linguistic Intelligence is the ability to speak, recognize, and use mechanisms of phonology (speech sounds), syntax (grammar) and semantics (meaning).

The Inference engine is a machine that uses an efficient procedures and rules to deduce a correct, flawless solution from the knowledge base. It also acquires and manipulates the knowledge from the knowledge base to arrive at a particular solution.

The knowledge base is a database of specific domain knowledge that contains rules and facts and the aim is to represent knowledge in a simple form that is understandable by humans.

Statistical algorithms most natural language processing methods use statistical algorithms that is based or combined with a Classical Deep Learning approach which are superficial in their analysis and later converted into algorithms. These mathematical models could be bilinear transform which is also called Tustin's method. This is used in digital signal processing and discrete-time control theory to transform continuous-time system representations to discrete-time and vice versa. Another mathematical model used for signal processing is the Z-transform that converts a discrete-time signal- which is a sequence of real or complex numbers into a complex frequency domain representation. This theory uses Z-transform and it can be defined as either a one-sided or two-sided transform with the below definition: Z-transform of a discrete-time signal x[n] is the formal power series X(z) defined as:

$$X(z) = \mathcal{Z}\{x[n]\} = \sum_{n=-\infty}^{\infty} x[n] z^{-n}$$ Where n is an integer and z is, in general, a complex number:

$$z = Ae^{j\phi} = A \cdot (\cos\phi + j\sin\phi)$$

where
A is the magnitude of z, j is the imaginary unit, and ϕ is the complex argument (also referred to as angle or phase) in radians. And, Unilateral Z-transform an alternative has in cases where x[n] is defined only for n ≥ 0, the singlesided or unilateral Z-transform is defined as:

$$X(z) = \mathcal{Z}\{x[n]\} = \sum_{n=0}^{\infty} x[n] z^{-n}.$$

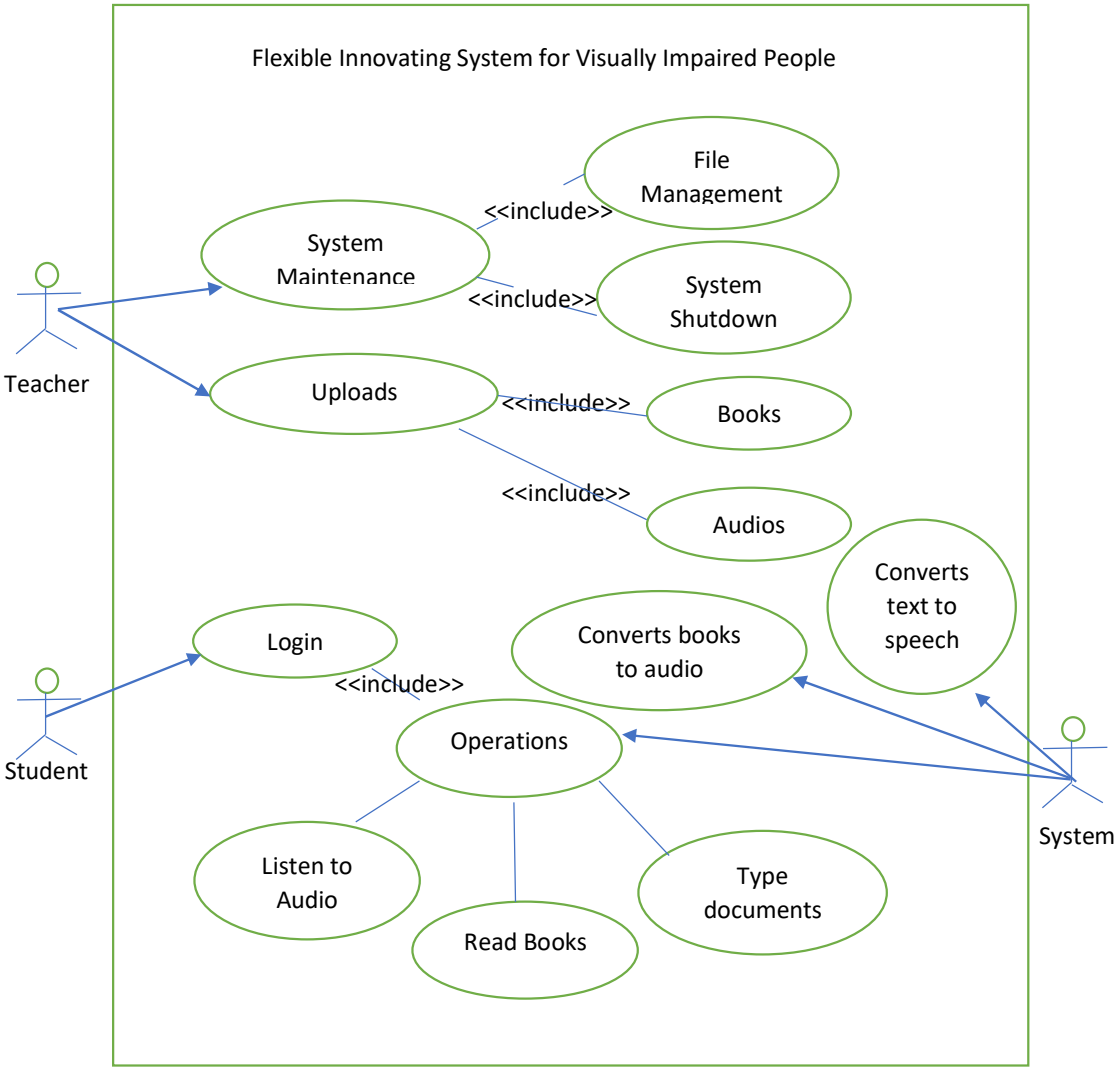
In signal processing, this definition can be used to evaluate the Ztransform of the unit impulse response of a discrete-time causal system (Wikipedia, 2018).

Algorithm is a sequential method for doing something or a set of instructions to be executed over and over until certain condition(s) is/are met.

Synthesis is the process of using a lexicon, letter-to-sound rules and models that is clear and definite to convert text to speech using any of the following methods concatentive, articulatory and formant synthesis.

Computing, Information Systems & Development Informatics Journal
Vol. 9 No. 4, 2018 - www.cisdijournal.org

Use Case Diagram



The following Actors featured in the system

1. **The Teacher:** The also act as the administrator of the system, he /she upload books, videos etc. that the visual impaired people used for learning.
2. **The Student:** This is the people that use the system by listening to the Audios or videos uploaded by the teacher
3. **The System:** This is the computer system that will convert all input texts to audio speech or video speech. It synthesis the text to speech so that the visual impaired people can listening to it

4. RESULT DEMONSTRATION INTERFACE

The results interface are depicted in the figures that follow:

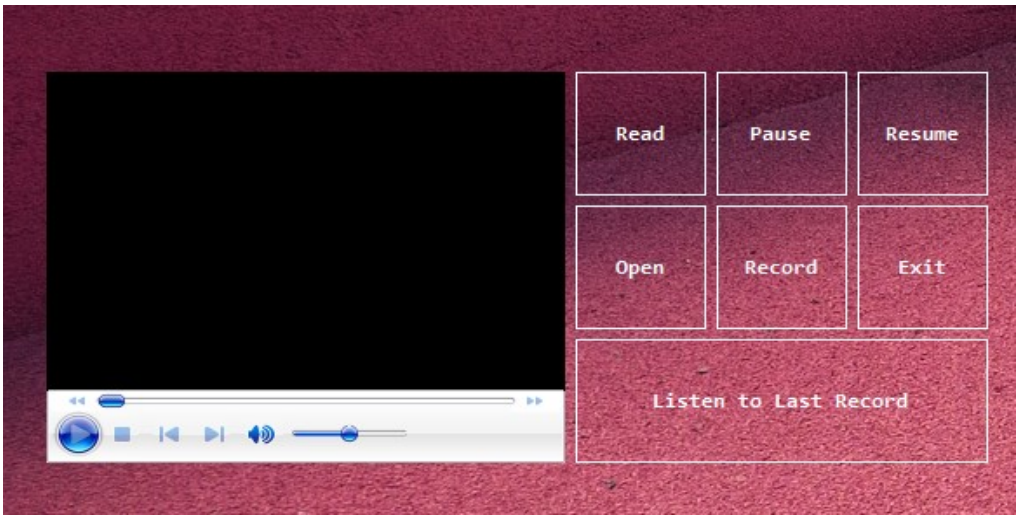


Figure 4a: Add books. The above figure illustrates the interface to convert books into audio formats.

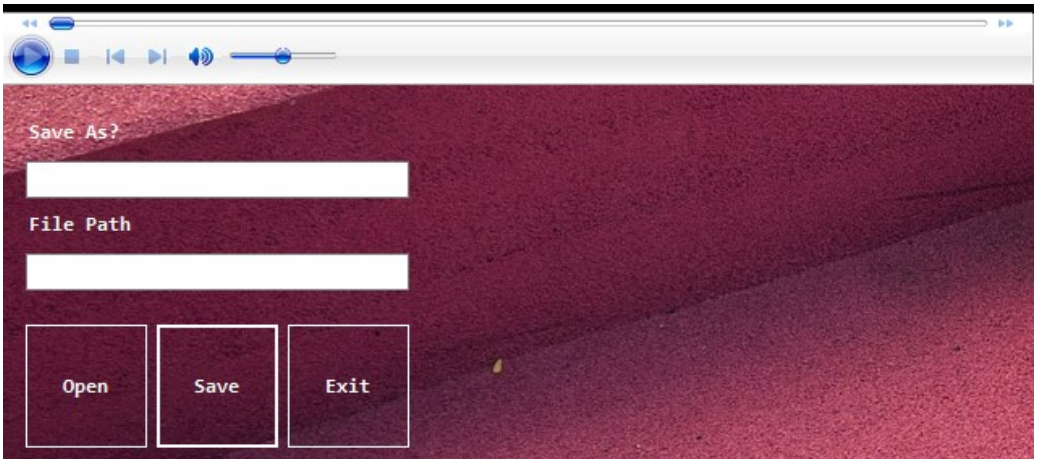


Figure 4b: Add Audio. The abpve figure illlustrates the add audio interface.

Figure 4c: All Uploads. In this interface, all uploaded books is displayed

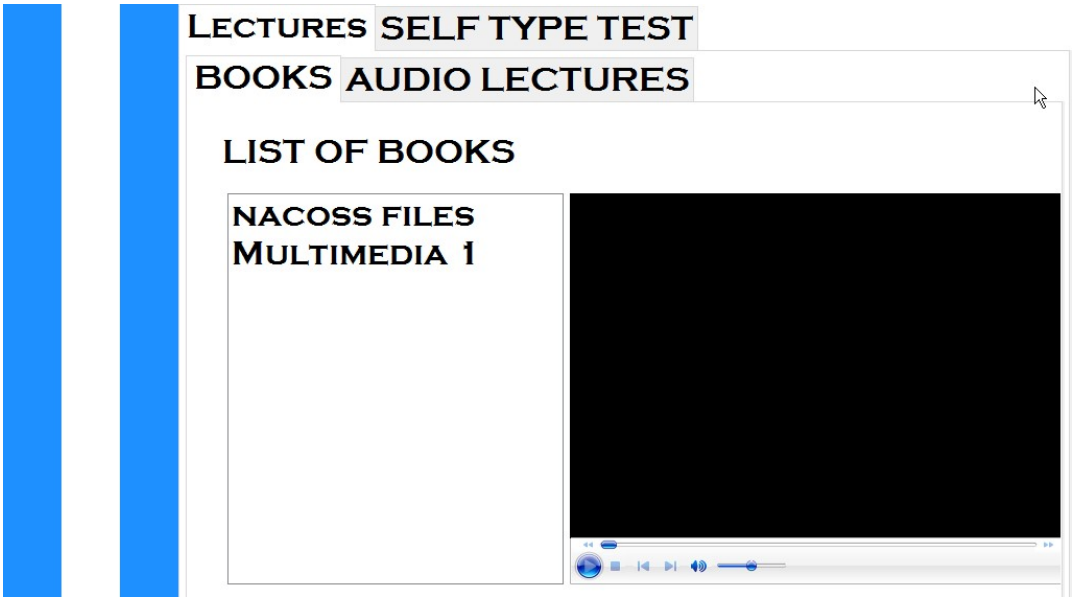


Figure 4d: Listen to lectures. In this interface, a visually impaired student can listen to textbooks that has already been synthesized into audio formats.

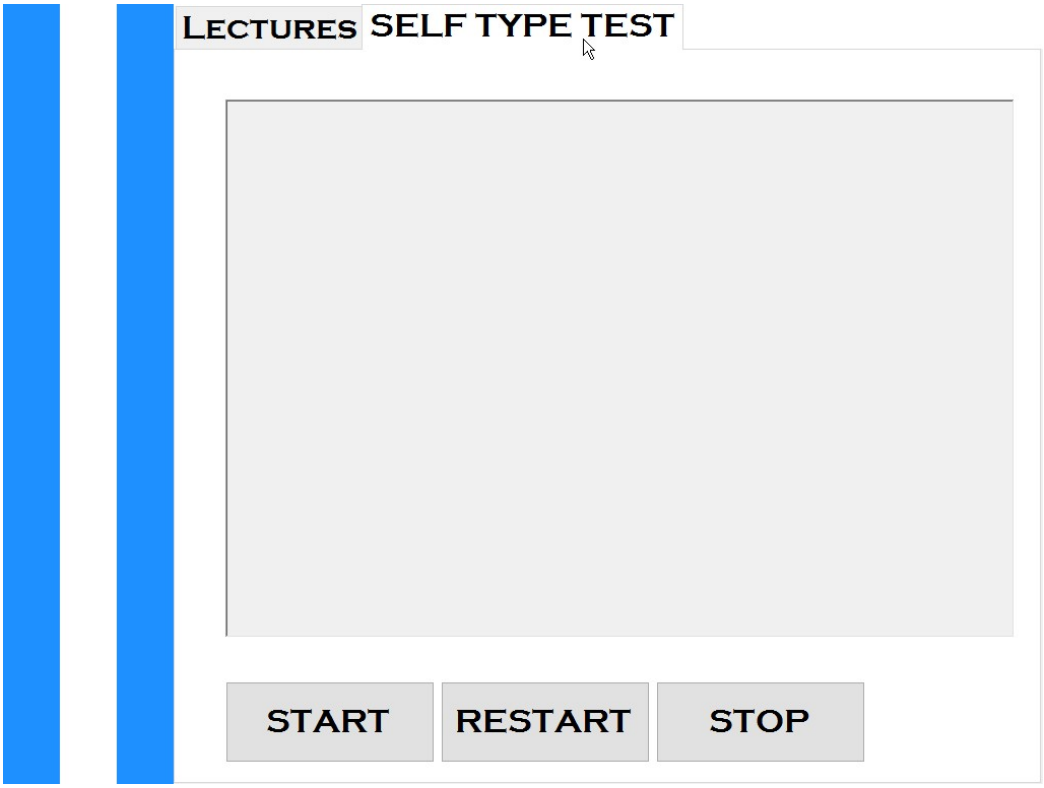


Figure 4e: self type test. This interface illustrates a self type test where the students types and the system speaks out what is been typed.

5. SUMMARY, CONCLUSION AND RECOMMENDATION

5.1 Summary

This system underpins some of the activities of teaching assistant software for visually impaired students. It synthesizes books in electronic formats to .mp3 audio format to ease students in learning and the teachers, in teaching their students unlike the previous system. This research work developed software that can successfully handles and operates as a teaching assistant for visually impaired students. In the process of the design, information on software development process was obtained from the internet, journals and observation.

5.2 Conclusion

Text to speech synthesis is a rapidly growing aspect of Natural Language Processing (NLP) which is increasingly playing a more important role in the way we interact with the system and interfaces across a variety of platforms. This paper has identified the various operations and processes involved in text to speech synthesis with an attractive graphical user interface that allows the user to type in his/her text.

5.3 Recommendation

We therefore recommend that this application should be made available for every school for blind and vision impaired student. That there should be plan to create engines for localized Nigerian languages so as to make text to speech technology more accessible to a wider range of Nigerians which is already exists in some native languages such as Swahili, Konkani, the Vietnamese synthesis system and the Telugu language. That there should be an implementation of a text to speech system on other platforms, such as telephony systems, ATM machines, video games and any other platforms where text to speech technology would be an added advantage and increase functionality.

REFERENCES

1. Belay, T. E (2005). African Perspective on Visual Impairments ICT's and Policies: A Personal Experience, A Variety of Perspectives and Technological Solution, A Paper Presented To: UNESCO at the World Summit on the Information Society, Workshop on ICT and persons with disabilities Kram Exhibitions Park, Room Zaghouan, Tunis, Tunisia
2. Black, A. & Lenzo, K. (2000). Building voices in the Festival Speech Synthesis System. Retrieved from https://www.researchgate.net/publication/243669217_Building_voices_in_the_Festival_speech_synthesis_system April 15, 2018
3. Julian (2012).Visual Impairments, *including Blindness. A publication of NICHCY, Disability Fact Sheet #13,1825 Connecticut Avenue N.W.Washington,*
4. Isewon I. (2014) Design and Implementation of Text To Speech Conversion for Visually Impaired People. International Journal of Applied Information Systems (IJAIS),Foundation of Computer Science FCS, New York, USA Volume 7– No. 2, April 2014 – www.ijais.org
5. Parr M & Fung I, (2000) A Review of the Literature on Computer-Assisted Learning, Particularly Integrated Learning Systems, and Outcomes with Respect to Literacy and Numeracy. School of Education, University of Auckland.
6. Sisson, C (2007). Text-to-speech in Vocabulary Acquisition and Student Knowledge Models: Language Technologies Institute, Carnegie Mellon University
7. Wikipedia (2018). Digital Signal Processing. Retrieved from https://en.m.wikipedia/wiki/Digital_signal_processing on August 16, 2018
8. Wikipedia ((2018). Z-transform. Retrieved from https://en.m.wikipedia/wiki/Z-transform on August 16, (2018