East African Scholars Journal of Engineering and Computer Sciences
Abbreviated Key Title: East African Scholars J Eng Comput Sci
ISSN: 2617-4480 (Print) & ISSN: 2663-0346 (Online)
Published By East African Scholars Publisher, Kenya
Volume-3 | Issue-1 | Jan-2020 | DOI: 10.36349/easjecs.2020.v03i01.01

OPEN ACCESS

Research Article

An Interactive Cloud Based User Oriented, Dynamic and Intelligent Text-To-Speech Module

Raiyetunbi, Oladimeji Jude*[1] and Ayeh Emmanuel[1]
[1]Department of Mathematical Sciences, Kogi State University, Anyigba. Kogi State. Nigeria

Article History
Received: 08.01.2020
Accepted: 18.01.2020
Published: 31.01.2020

Journal homepage:
https://www.easpublisher.com/easjecs

Quick Response Code

Abstract: A **text-to-speech** (**TTS**) system which we refer to as Artificial Interpreter, And of course an area of Artificial Intelligence (AI) Natural language processing, converts normal language text into speech; other systems render symbolic linguistic representations like phonetic transcriptions into speech. We have developed a cloud enhanced system that converts texts into audio. The system was developed using JavaScript as the programming language with HTML5 and CSS for the interface design. JavaScript is used because it is a client-side programming language for web applications. JavaScript is robust with a good number of speech synthesis libraries that can be used to build robust text to speech systems a cloud. JavaScript is also light-weight in comparison with other high-level programming languages that support web application development, e.g Java Applet. We have been able to successfully convert plain text files into audio using this system when uploaded into the system via the file upload form or when copied and pasted into the system. This developed software can be accessed by users all over the world ubiquitously and is supported on major desktop and hand-held devices once it's connected to the web. This app will be of great help to lecturers and students in the classroom environment; e-libraries and computer aided learning for physically challenged users.

Keywords: Text to Speech system, Speech synthesis, Web, Interactive, e-libraries, Artificial Interpreter, ubiquitously, Cloud computing.

INTRODUCTION

The first industrial revolution attempted to create machines that could replace man's physical power. Industrialization has transformed the society totally and brought immediate crises in later development. In fact, there are machines that can outperform human beings over the centuries man's working ability and thinking process have seen a sea change. The society is becoming increasingly centered on information handling, processing, storage and dissemination, using microelectronic based technologies, today's computers can stimulate many human capabilities such as reading, grasping, calculating, speaking, remembering, comparing numbers, drawing, making judgments, and even interactive learning. Researchers are working to expand these capabilities and, therefore the power of computers by developing hardware and software that can initiate intelligent human behavior. For example, researchers are working on the systems that have the ability to reason, to learn or accumulate knowledge to strive for self-improvement, and to stimulate human sensory and mechanical capabilities. Experts are convinced that it is now only a matter of time; the present generation will experience the impact and utility of new applications based on artificial intelligence in offices, factories, libraries and homes. This general area of research is known as 'Artificial Intelligence.

Now according to researchers, Classroom use of audio recordings has long been a viable instructional intervention for struggling readers (Carbo 1978; Gilbert, Williams, and McLaughlin 1996). The increased interest in using such an intervention could be directly tied to the increased access and popularity of audiobooks. Technological innovations, combined with the marketability of audiobooks, have led to a drastic increase in the offerings of traditionally print resources through electronic media, including audiobooks (King-Sears *et al;* 2011). The popularity of audiobooks has exploded in the past decade with audiobook publishing expanding into a billion-dollar industry. According to figures released in 2010, consumers purchased 900,000

***Corresponding Author:** Raiyetunbi, Oladimeji Jude Email: raitumbi [at]yahoo.com

more audiobooks in 2009 than in 2008, a 4.7 percent increase in unit sales.

Speech synthesis is the artificial production of human speech. A computer system used for this purpose is called a **speech computer** or **speech synthesizer**, and can be implemented in software or hardware products. A **text-to-speech** (**TTS**) system **(Artificial Interpreter, AI)** converts normal language text into speech; other systems render symbolic linguistic representations like phonetic transcriptions into speech, (*Allen, Jonathan et al; 1987).*

Our text-to-speech system (web based, Artificial Interpreter) is composed of two parts: a front-end and a back-end. The front-end has two major tasks. First, it allows the user to upload the text file that should be read through a file upload dialogue. Secondly, it displays the content of the text file into the Text Screen which serves as evidence of the text being read by the system. The back-end, on the other hand, is where the conversion of the text symbols take place. Through this conversion of text symbols into sound, the output is produced. The developed speech synthesizer; text-to speech module would be housed in a cloud. Cloud computing is a term generally used to describe a system available to many users over the internet (*Ubiquitous Technology)*. Our software is cloud enhanced, which means users from all over the world can access it at their convenience on all major devices. Text-to- speech synthesizer (TTS) is the technology which lets computer speak to you. The TTS system gets the text as the input and then a computer algorithm which is called TTS engine analyses the text, pre-processes the text and synthesizes the speech with some mathematical models. The TTS engine usually generates sound data in an audio format as the output, *Itunuoluwa Isewon et al; (2014).* TTS systems, in contrast, are theoretically capable of "reading" any string of text characters to form original sentences.

Reading comprehension and interest are essential in learning. The act of reading permits students to learn new vocabulary and concepts and to access different types of reading materials (Frank Serafini 2004). If students fall behind in reading comprehension for their age/grade level, then students struggle to process new vocabulary and concepts presented in textbooks and other literature. Difficulty in reading may translate into poor school performance due to the inability to process new vocabulary and concepts in a meaningful manner. These difficulties can evolve into students' losing interest in reading and entering a state of learned helplessness. This cycle can lead to students' dropping out of high school and possessing below-average reading comprehension skills as adults. Hence the manual or traditional reading approach that are usually employed in schools, lecture rooms , Libraries and workshop, could be faced with certain challenges, such as

1) The Traditional way of reading lecture notes or studying for examinations, etc is quite tedious and demanding. Most times, readers get stressed up due to other activities they were engaged with before settling down to read.
2) Other health challenges like migraine can be a reason why readers will lose interest to sit down and read.

Developing strong reading skills in students is one of the key goals of every education program. It is through reading that students expand their vocabulary and learn about the world. Not everyone acquires reading skills at the same rate, Meredith Cicerhia extracted from www.read and spell.com three common reading problems for students

The purpose of this paper work is to improve reading skills and improve user's attitude toward reading by creating a web based audio app for struggling readers. Researchers have demonstrated that the use of technology exposes struggling readers to different types of literature and assists them with vocabulary acquisition (Stone-Harris 2008). The significance of the study also exhibits that the use of web based text to speech synthesizer apps can lead to an improvement in struggling readers' skills and attitudes. If use of such apps can be proven to benefit struggling readers, then educators will possess another instructional technique to assist struggling readers improve their reading skills and attitudes.

CONCEPTS AND IDEAS

INTERACTIVE COMPUTER ASSISTED LEARNING

Ever since the introduction of microcomputers, an educational and instructional software; teaching software 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. Different teaching application software 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 software 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.

However, given the rapid development of technology, recent types of courseware were not available in the early CAI research (Kabari L. G and L. F. Atu 2015). Modern implementation of CAI includes more advanced hardware and software technology, and allows for greater student interaction, and greater stores of information.

Several systems have been developed that convert text to speech. Some of them which inspired the development of our software are as follows: (Itunuoluwa Isewon *et al;* 2014) developed the Text To Speech Robot, a simple application with the text to speech functionality. The system was developed using Java programming language. As a result, the software could only function offline when installed on a computer because it was not cloud enhanced. (Kaladharan N, 2015) worked on An English Text to Speech Conversion System; a system which he developed using Microsoft .Net framework. Although the system was successful, it was not cloud enhanced or enabled.

(Frank Serafini 2004) has explained that much research validates the importance of reading aloud to students, positing that the act of reading aloud introduces new vocabulary and concepts, provides a fluent model, and allows students access to literature they are unable to read independently. He adds that audiobooks, sometimes called Text to speech system (TTS) or Artificial Interpreter is an important component of a comprehensive reading program. (Kylene Beers 1998) has said that audiobooks, when used with reluctant, struggling, or second language learners, serve as a scaffold that allows students to read beyond their reading level. Since the reading process develops through oral language experiences, audiobooks benefit struggling readers by increasing comprehension and appreciation of written text (Wolfson 2008). This benefit has long been seen by classroom teachers.

METHODOLOGY AND APPROACH

COMPONENENTS INTEGRATION AND EXPERIMENTAL SECTION

The methodology used in this research work is the **SSADM** (Structured System Analysis and Design Methodology). **SSADM** is another method dealing with information systems design. It was developed in the UK by CCT (Central Computer and Telecommunications Agency) in the early 1980s. It is the UK government's standard method for carrying out the system analysis and design stages of an information technology project of any kind. The SSADM has been traditionally used for the development of medium or large systems.

Below is the waterfall diagram of the SSADM.

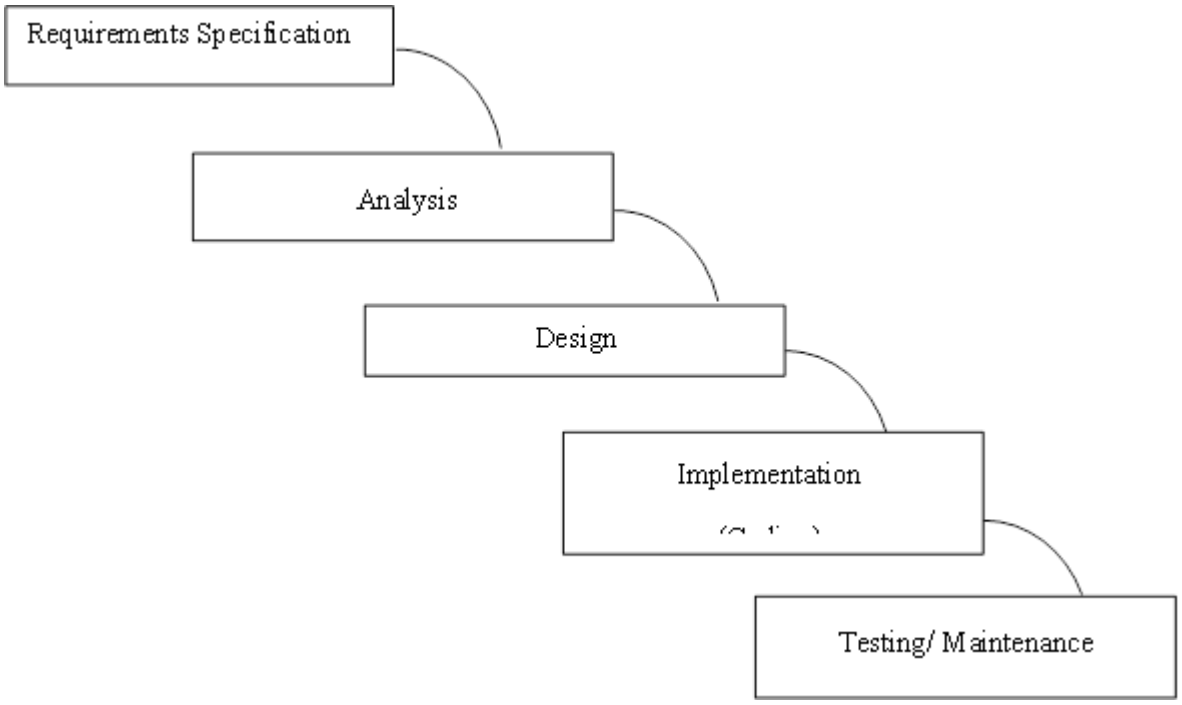

Fig-1: Waterfall Diagram of SSADM (self)

Logical Structure of A Text-To-Speech Synthesizer System

Text-to-speech synthesis takes place in several steps. The TTS systems get a text as input, which it first must analyze and then transform into a phonetic description. Then in a further step it generates the prosody. From the information now available, it can produce a speech signal. The structure of the text-to-speech synthesizer can be broken down into major modules:

- Natural Language Processing (NLP) module: It produces a phonetic transcription of the text read, together with prosody.

- Digital Signal Processing (DSP) module: It transforms the symbolic information it receives from NLP into audible and intelligible speech. The major operations of the NLP module are as follows:

✓ Text Analysis: First the text is segmented into tokens. The token-to-word conversion creates the orthographic form of the token. For the token "Mr" the orthographic form "Mister" is formed by expansion, the token "12" gets the orthographic form "twelve" and "1997" is transformed to "nineteen ninety seven".

✓ Application of Pronunciation Rules: After the text analysis has been completed, pronunciation rules can be applied. Letters cannot be transformed 1:1 into phonemes because correspondence is not always parallel. In certain environments, a single letter can correspond to either no phoneme (for example, "h" in "caught") or several phoneme ("m" in "Maximum"). In addition, several letters can correspond to a single phoneme ("ch" in "rich"). There are two strategies to determine pronunciation:

✓ In dictionary-based solution with morphological components, as many morphemes (words) as possible are stored in a dictionary. Full forms are generated by means of inflection, derivation and composition rules. Alternatively, a full form dictionary is used in which all possible word forms are stored. Pronunciation rules determine the pronunciation of words not found in the dictionary.

✓ In a rule based solution, pronunciation rules are generated from the phonological knowledge of dictionaries. Only words whose pronunciation is a complete exception are included in the dictionary. The two applications differ significantly in the size of their dictionaries. The dictionary-based solution is many times larger than the rules-based solution's dictionary of exception. However, dictionary-based solutions can be more exact than rule-based solution if they have a large enough phonetic dictionary available.

Prosody Generation: after the pronunciation has been determined, the prosody is generated. The degree of naturalness of a TTS system is dependent on prosodic factors like intonation modelling (phrasing and accentuation), amplitude modelling and duration modelling (including the duration of sound and the duration of pauses, which determines the length of the syllable and the tempos of the speech), *Text-to-speech technology* ;*(Retreived February 21, 2014 from http://www.linguatec.net/products/technology).*

Dutoit, T., 1997. High-Quality Text-to-Speech Synthesis: An Overview. Journal Of Electrical And Electronics Engineering Australia 17, 25-36.

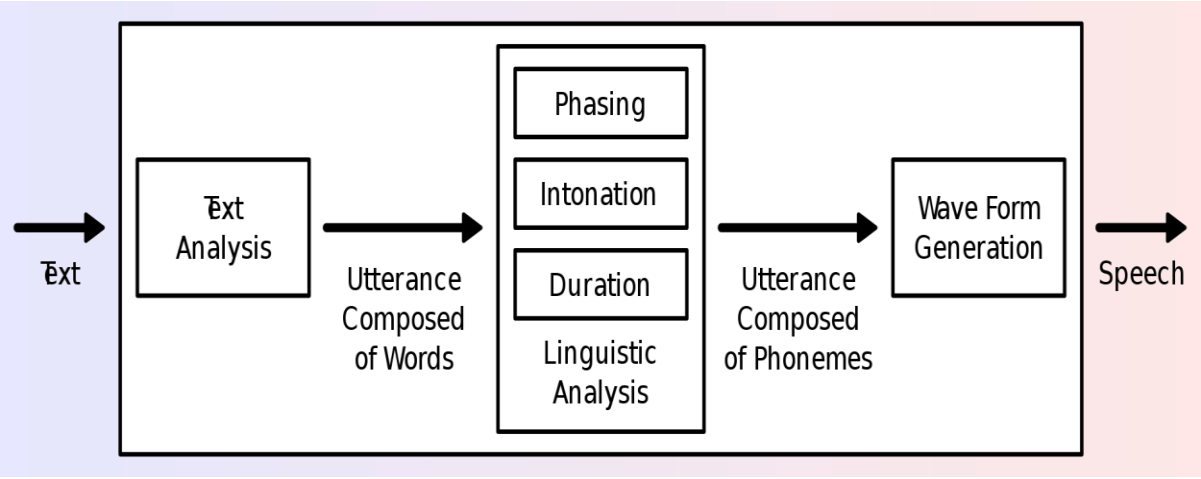

Fig 2: Overview of a typical TTS system (*Wikipedia*)

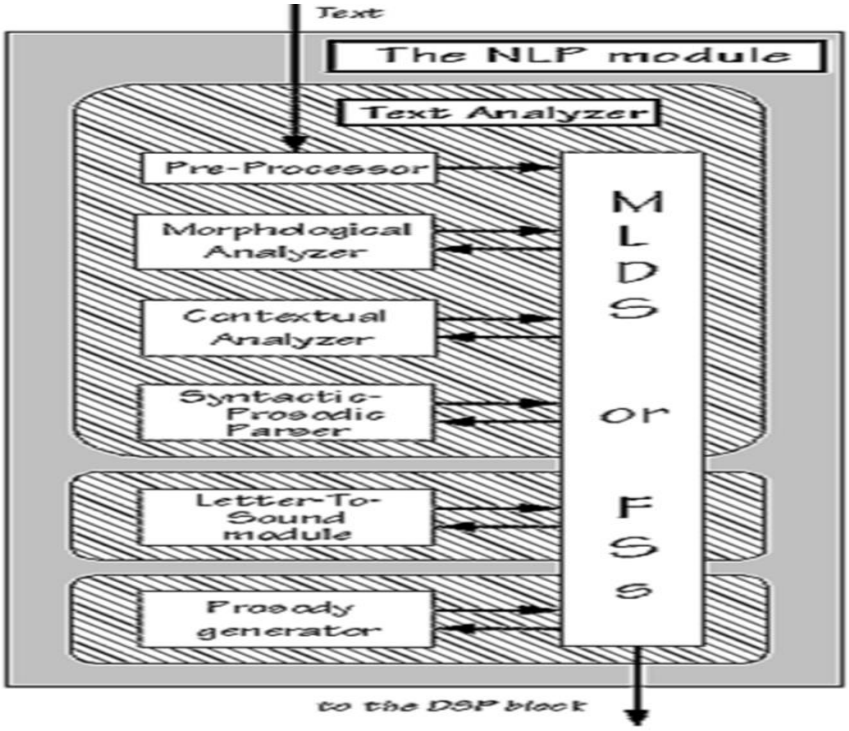

Fig 3: Operations of the natural Language processing module of a TTS synthesizer. (Itunuoluwa Isewon et al; 2014)

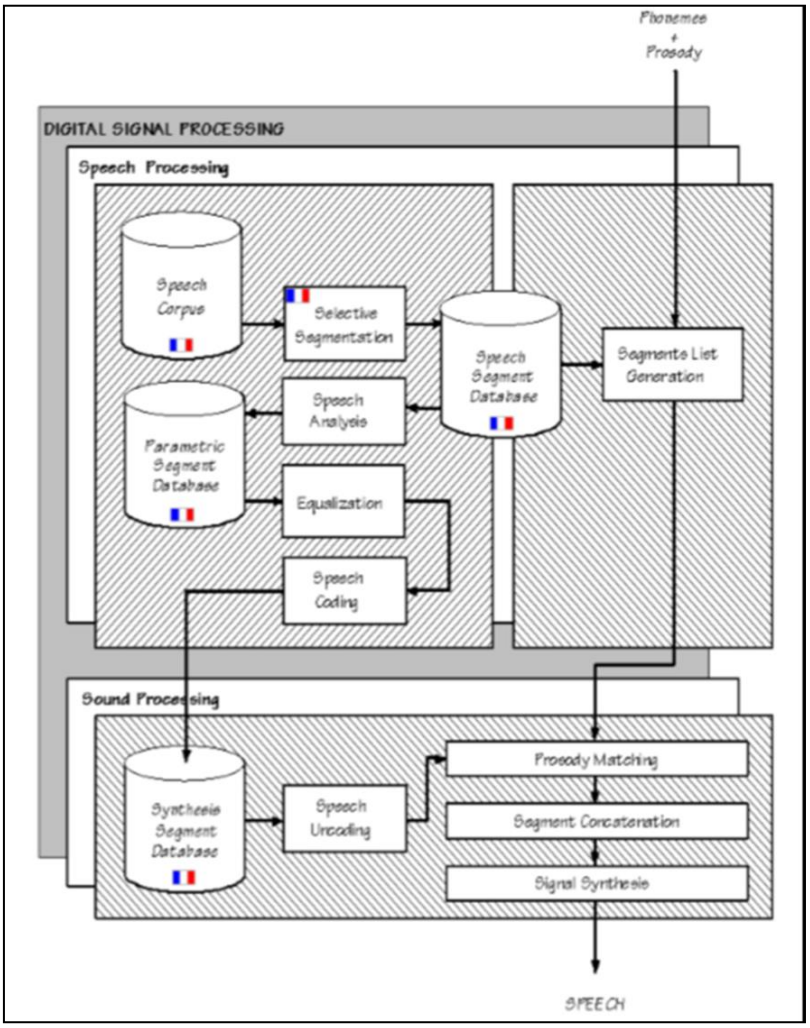

Fig 4: The DSP component of a general concatenation- based synthesizer. (*Dutoit, T., 1997*)

In the system analysis and design stage there are some tools that are used in order to construct the model. These system developmental tools that are used for this application are the Unified Modeling Language (UML).Several different notations for describing object oriented designs were proposed in the 1980s and 1990s. The Unified Modeling Language is an integration of these notations. It describes notations for a number of different models. The Unified Modeling Language (UML) is used to specify, visualize, modify, construct and document the artifacts of an object-oriented software-intensive system under development. Unified Modeling Language. Unified Modeling Language (UML) is a standardized general-purpose modeling language in the field of software engineering. The standard is managed, and was created by, the Object Management Group. UML includes a set of graphic notation techniques to create visual models of software-intensive systems UML offers a standard way to visualize a system's architectural blueprints, including elements such as activities, actors, business processes, database schemas, (logical) components, programming language statements, reusable software components. The UML models that were used in designing the text to speech software are: Sequence diagram, Use case and Flow chart diagrams.

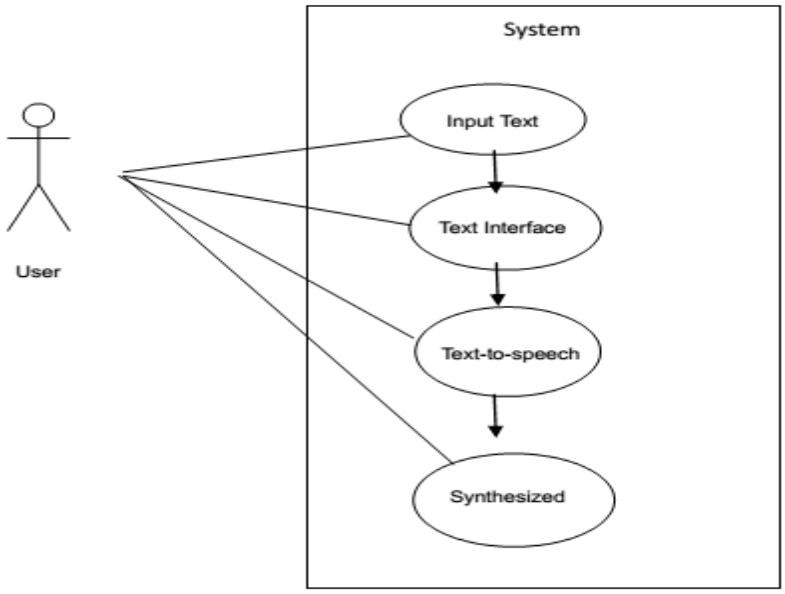

Fig 5. Use case diagram for the voice synthesizer

Behavior diagrams emphasize what must happen in the system being modeled. Since behavior diagrams illustrate the behavior of a system, they are used extensively to describe the functionality of software systems. An example is the use case diagram which was used in this paper. This model describes the functionality provided by a system in terms of actors, their goals represented as use cases, and any dependencies among those use cases. Use case diagrams graphically depict the interactions between the system and external system and users. In other words, they graphically decide who will use the system and in what ways the user expects to interact with the system. Figure above shows the use case diagram in this work. **.** From the case diagram the user is required to type in / input his/her preferable text in the right text format, or can even load a file from his/her computer or a storage device. The precondition is that user must type in a valid text document for the interface to convert the text to synthesized speech. The post conditions are either a success end where an audible speech is produced or a failure end where the system display error message and prompts the user to re- input a new text. The user can do either of the following:

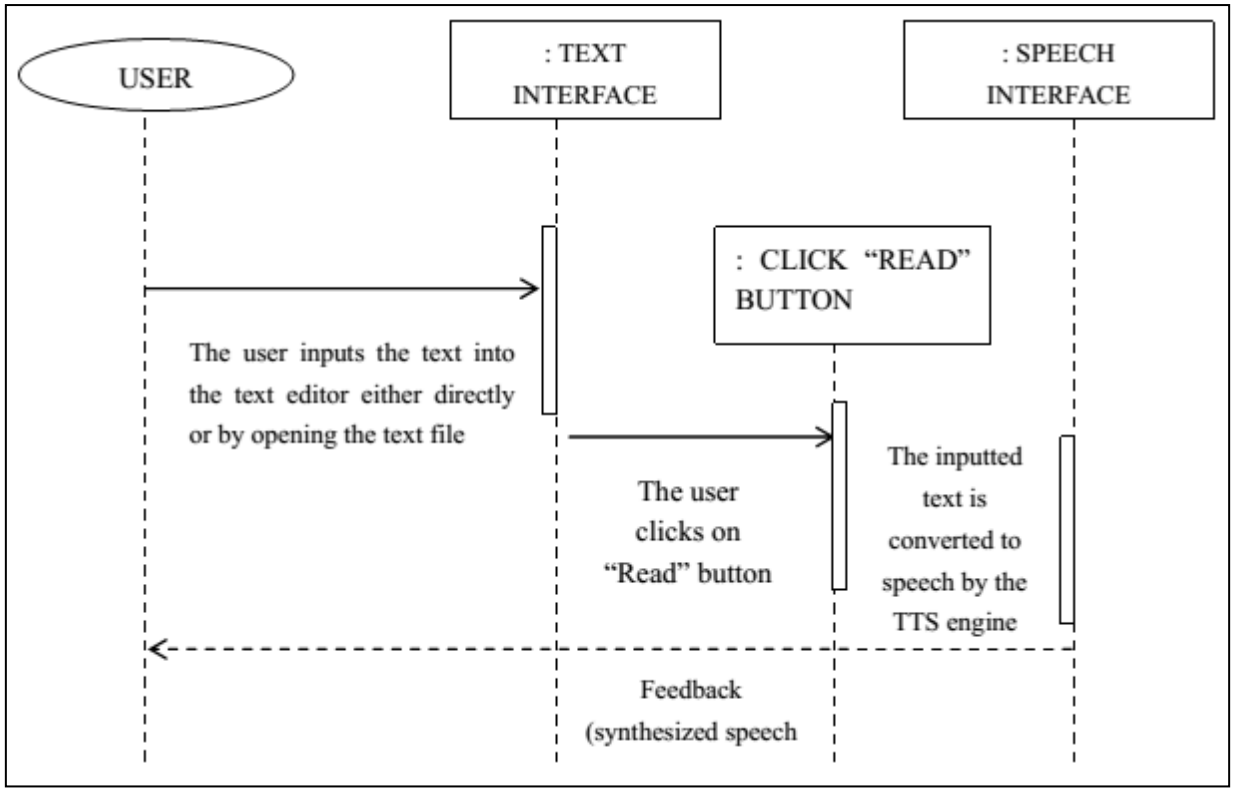

Fig 6. Sequence diagram for the TTS voice synthesizer

- ✓ Type the text and click read, the system responds by converting the typed text to audible synthesized speech.
- ✓ Open a text file in the text interface, the system responds by converting the opened text to audible synthesized speech.

The Sequence diagrams above, a subset of behavioral diagrams, emphasize the flow of control and data among the things in the system being modeled. An example is sequence diagram. Sequence diagram is an example interaction diagram that shows how objects communicate with each other in terms of a sequence of messages. Also indicates the lifespan of objects relative to those messages. It graphically depicts how objects interact with each other via messages in the execution of a use case or operation. They illustrate how messages are sent and received between objects and in what sequence.

SYSTEM REQUIREMENTA AND FUNCTIONS

Our application is a web-based application that can translate texts into audio. The system was developed using JavaScript as the programming language. JavaScript is used because it is a client-side programming language for web applications. JavaScript has a good number of speech synthesis libraries that can be used to build robust text to speech systems for the web.

The application is divided into two main modules - the main application module which includes the basic GUI components which handles the basic operations of the application such as input of texts and text files for conversion. The second module, the main conversion engine which is integrated into the main module is to convert the loaded or typed text into audio through the *artyom.js* library.

Software Requirements

The following are the softwares requirement specification for our Ultimate Text Reader based on the Artyom Text to Speech Js Library:

- **Google Chrome**
  Chrome 4 to 32 does not support Speech Synthesis API property. Chrome 33 to 67 supports Speech Synthesis API property.

- **Mozilla Firefox**
  Speech Synthesis API is not supported by Mozilla Firefox browser version 2 to 30. Speech Synthesis API is not supported by Mozilla Firefox browser version 31 to 48 by default but can be enabled in Firefox using the media. Web speech. Synth enabled about config flag. Speech Synthesis API is supported by Mozilla Firefox browser version 49 to 61.

- **Internet Explorer**
  IE browser version 6 to 11 doesn't support Speech Synthesis API.

- **Safari**
  Safari browser version 3.1 to 6.1 doesn't support Speech Synthesis API. Safari browser version 7.1 to 11.1 doesn't support Speech Synthesis API.

- **Microsoft Edge**
  Microsoft Edge browser version 12 to 13 does not support this property speech-synthesis-api. Microsoft Edge browser version 14 to 17 supports this property speech-synthesis-api.

- **Opera**
  Opera version 10.1 to 26 doesn't support Speech Synthesis API. Opera version 27 to 53 supports Speech Synthesis API.

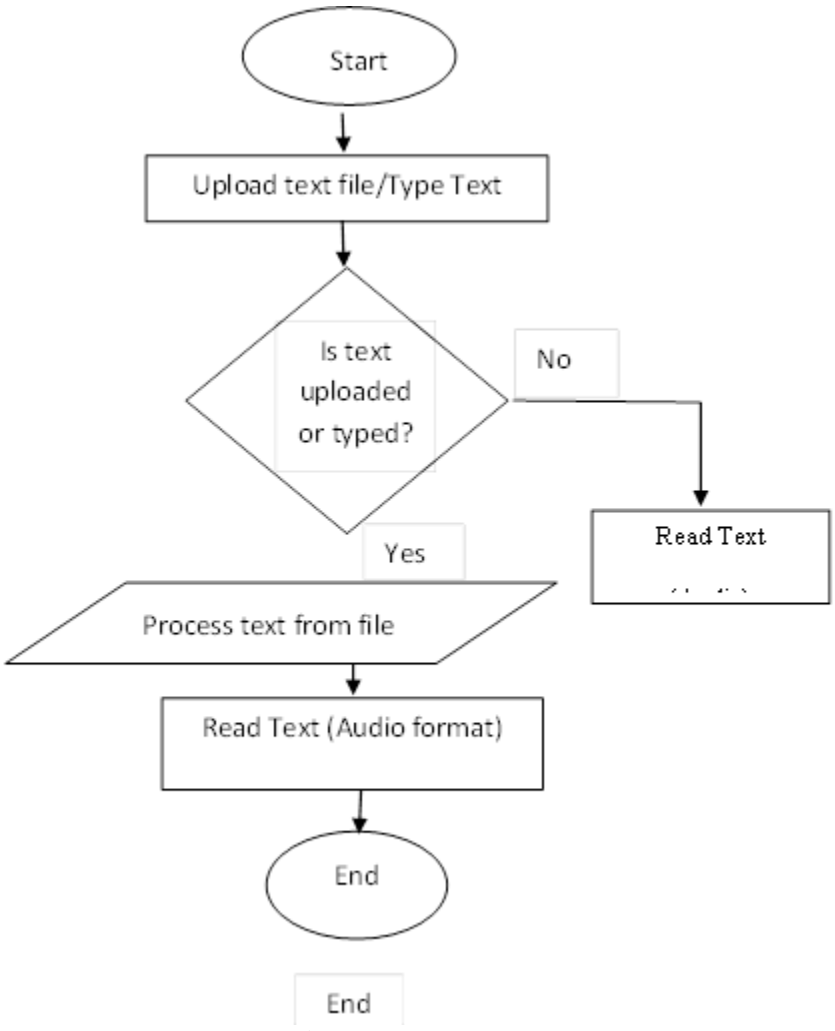

Fig 7. System overview

Interface Design

Below are the screenshots of the Web based TTS

Fig 8: System Launch Interface

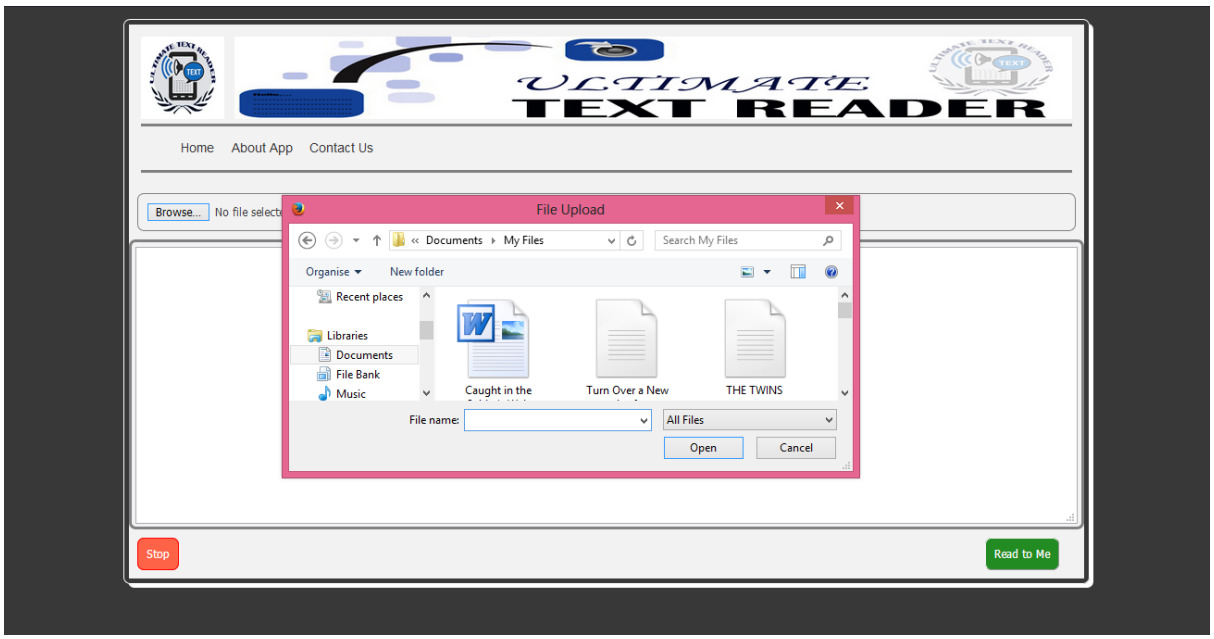

Fig 9: Text File Upload Interface

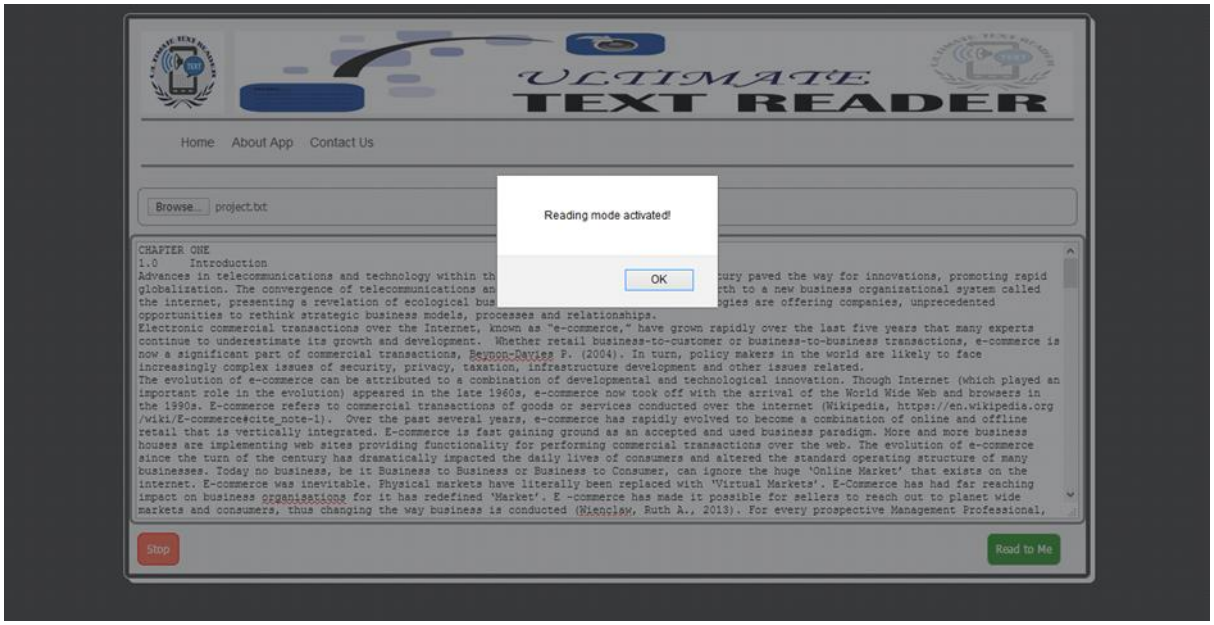

Fig 10: Reading Mode Interface

RESULT, DISCUSSION AND CONCLUSION

Text to speech synthesis is a rapidly growing aspect of computer technology, The Natural Language processing of the Artificial Intelligence. And is increasingly playing a more important role in the way we interact with the system and interfaces across a variety of platforms. We have identified the various operations and processes involved in text to speech synthesis. We have also developed a very simple and attractive graphical user interface which allows the user to type in his/her text provided in the text field in the application. Our system interfaces with a text to speech engine developed for American English. In future, we plan to make efforts to create engines for localized Nigerian language so as to make text to speech technology more accessible to a wider range of Nigerians.

REFERENCES

1. Jonathan,A., Hunnicutt, M. S., & Dennis, K.(1987). From Text to Speech: The MITalk system. Cambridge University Press.
2. Kylene, B. (1998). "Listen While You Read: Struggling Readers and Audiobooks." School Library Journal 44 (4), 34–35.
3. Carbo, M. (1978). "Teaching Reading with Talking Books." Reading Teacher (2005) 32 (3): 267–73. "What Principals Need to Know about Reading Instruction." Principal 85 (1): 46–49. <www.naesp.org/resources/2/Principal/2005/S-Op46.pdf> (accessed March 21, 2013).
4. Dutoit, T. (1997). High-Quality Text-to-Speech Synthesis: An Overview. *Journal Of Electrical And Electronics Engineering Australia 17, 25-36.*
5. Gilbert, L. M., Williams, R. L., & McLaughlin, T. F. (1996). Use of assisted reading to increase correct reading rates and decrease error rates of students with learning disabilities. *Journal of Applied Behavior Analysis, 29(2), 255-257.*
6. King-Sears, M. E., Swanson, C., & Mainzer, L. (2011). TECHnology and literacy for adolescents with disabilities. Journal of Adolescent & Adult Literacy, 54(8), 569-578.
7. Isewon, I., Oyelade, O. J., & Oladipupo, O. O. (2012). Design and implementation of text to speech conversion for visually impaired people. *International Journal of Applied Information Systems, 7(2), 26-30.*
8. Kabari L. G., & Atu, L. F. (2015). "Assisting the Speech Impaired People Using Text-to-Speech Synthesis" 221 International Journal of Emerging Engineering Research and Technology, 3 (I8)
9. Kaladharan, N. (2015). An English Text to Speech Conversion System, International Journal of Advanced Research in Computer Science and Software Engineering, vol 5, Issue 10, October-2015.
10. Meredith Cicerchia, (extracted from www.readandspell.com: three common reading problems for students).
11. Text-to-speech technology: In Linguatec Language Technology Website. Retrieved February 21, 2014, from http://www.linguatec.net/products/tts/information/techno logy.
12. Serafini, F. (2004). "Audiobooks and Literacy: An Educator's Guide to Utilizing Audiobooks in the Classroom." New York: Listening Library. <http://www.frankserafini.com/classroom-resources/audiobooks.pdf> (accessed March 21, 2013).

13. Stone-Harris, S. (2008). *The benefit of utilizing audiobooks with students who are struggling readers*. Walden University.
14. Wolfson, G. (2008). Using audiobooks to meet the needs of adolescent readers. *American Secondary Education*, 105-114.