ISSN(Online): 2319-8753
ISSN (Print): 2347-6710

International Journal of Innovative Research in Science, Engineering and Technology

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Visit: www.ijirset.com

Vol. 9, Issue 3, March 2020

An Improved Text-to-Speech Synthesizer System for the Physically Impaired using Hidden Markov Model (HMM)

Isodje, Oluseyi Adedoyin[1] and Onuodu, Friday Eleonu[2]

Department of Information Technology, National Open University of Nigeria, Nigeria[1]

Research Scholar, Department of Computer Science, University of Port Harcourt, Rivers State, Nigeria[2]

ABSTRACT: There is a great level of importance attached to Text-to-Speech Synthesizers. This is because; Text-to-speech supports poor writers, students who have dyslexia, those who may confuse words, students with poor eyesight as well as users who have a reading disability. Students can listen to their text after they have typed their work. Some programs colour highlights each word that has been spoken. Users can see each word more clearly, as each word is voiced. In this work, an Improved Text-to-Speech (TTS) Synthesizer System for the physically impaired using Hidden Markov Model (HMM) was developed. Furthermore, Water-fall Model Methodology was adopted in this approach, and also implemented with VB.Net and MySQL as backend. The expected results of the work showed an efficiency rate of 90% in terms of Time Complexity (TC), Life Cycle Assessment (LCA), Benchmarking (B), Multi-Criteria Decision Making (MCDM), Risk Assessment (RA), Cost Benefit Analysis (CBA) and Speed (S). The assessed values for the mentioned parameters were given as 21, 15, 12, 16, 11, 9 and 6 respectively. In addition, the study could be beneficial to physically impaired students, teachers of the physically impaired students and software developers.

KEYWORDS: Dyslexia**,** Hidden Markov Model, Physically Impaired, Synthesizer, Text-to-Speech

I. INTRODUCTION

There is a great level of importance attached to Text-to-Speech Synthesizers. This is because; Text-to-speech supports poor writers, students who have dyslexia, those who may confuse words, students with poor eyesight as well as users who have a reading disability. Students can listen to their text after they have typed their work. Some programs colour highlights each word that has been spoken. Users can see each word more clearly, as each word is voiced. They can slow the voice rate, fast or slow. Most programs also have pitch control. Voices can be heard with a very 'high' pitched voice or with a very deep sounding voice. Struggling writers often do not wish to re-read their assignments or process writing [1]

Once they finish, they typically want to hand it to their teacher immediately and dismiss it. Children thoroughly enjoy hearing their work, especially if the voices are robots, clowns, old women or flies! One program of latter years, TalkAny, has a number of very amusing voices. Given that they can listen to their story, or project with a funny or unusual voice, younger students usually respond and elect to hear the text spoken through the computer speakers, or for privacy in a busy classroom or at home, through headphones. They might speed up the voice or customize it in other ways. Most reluctant writers and readers will participate and have their text voiced over and over again. They see the words and hear them in context. They can listen for spelling and typing errors and quite independently ascertain whether the document is ready for publication, printing, and assessment or requires further editing. They have greater ownership and involvement in their text production and can enjoy the writing process. A voiced application also assists writers as commas are voiced as slight pauses.

ISSN(Online): 2319-8753
ISSN (Print): 2347-6710

International Journal of Innovative Research in Science, Engineering and Technology

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Visit: www.ijirset.com

Vol. 9, Issue 3, March 2020

A full stop (or endpoint, or period) will demarcate the end of a sentence. Some packages can be set to speak each letter or number as it is typed. Users can also opt to have each word voiced i.e. after the space bar entry. Sentences can be read back to the writer after a full stop, exclamation mark or question mark. Users clearly hear the associated inflexion with regard to the latter punctuation marks. The difference is heard and users can determine why and how these conventions are used. This is a powerful learning outcome. For ease of use, most programs have a toolbar, where users can quickly open new or existing documents, print, navigate through their documents, change font size, colour and style, change the background color and have text voiced. Voices are usually selected from a Speech Menu.

According to Yi et al [2], "Text to speech (TTS) and automatic speech recognition (ASR) are two dual tasks in speech processing and both achieve impressive performance, thanks to the recent advance in deep learning and large amount of aligned speech and text data". However, the lack of aligned data poses a major practical problem for TTS and ASR on low resource languages. Speech can be turned off entirely as well if required as there are times when users do not wish to have the content voiced. Passages can be repeated depending upon the program's features as well or just single words.

Volume and speed can be altered to suit individual needs with a test facility sometimes included in the speech settings dialogue box. Educators, trainers, lecturers and tutors as well as teaching staff must begin to offer school-based research, class notes, instruction and work sheets electronically in a word processing format. Educators and trainers must promote a change of culture in our schools and Universities so that students who struggle, who find reading difficult and who may have dyslexia, can cope better and be more independent. If all class work from teaching staff was produced and provided in programs such as MS Word, then students could use text to speech technologies to enlarge text quickly, listen to passages, sentences or key words and become more capable and confident.

Hidden Markov Model (HMM) is a statistical Markov model in which the system being modelled is assumed to be a Markov process with unobservable (i.e. hidden) states. The hidden Markov model can be represented as the simplest dynamic Bayesian network. The mathematics behind the HMM were developed by L. E. Baum and co-workers. HMM is closely related to earlier work on the optimal nonlinear filtering problem by Ruslan L. Stratonovich, who was the first to describe the forward-backward procedure. In simpler Markov models (like a Markov chain), the state is directly visible to the observer, and therefore the state transition probabilities are the only parameters, while in the hidden Markov model, the state is not directly visible, but the output (in the form of data or "token" in the following), dependent on the state, is visible.

Each state has a probability distribution over the possible output tokens. Therefore, the sequence of tokens generated by an HMM gives some information about the sequence of states; this is also known as pattern theory, a topic of grammar induction. The adjective hidden refers to the state sequence through which the model passes, not to the parameters of the model; the model is still referred to as a hidden Markov model even if these parameters are known exactly. Hidden Markov models are especially known for their application in reinforcement learning and temporal pattern recognition such as speech, handwriting, gesture recognition, part-of-speech tagging, musical score following, partial discharges and bioinformatics. A hidden Markov model can be considered a generalization of a mixture model where the hidden variables (or latent variables), which control the mixture component to be selected for each observation, are related through a Markov process rather than independent of each other. Recently, hidden Markov models have been generalized to pairwise Markov models and triplet Markov models which allow consideration of more complex data structures and the modelling of non-stationary data.

II. RELATED WORK

Venkateswarlu et al [3] researched on Text-to-Speech conversion. The work presented a robust approach for text extraction and converting it to speech. Testing of device was done on raspberry pi platform. The Raspy is initially connected to the internet through Video Local Area Network (VLAN). The software is installed using command lines.

ISSN(Online): 2319-8753
ISSN (Print): 2347-6710

International Journal of Innovative Research in Science, Engineering and Technology

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Visit: www.ijirset.com

Vol. 9, Issue 3, March 2020

The authors did a good work. However, performance evaluation showed that their developed system could not be useful for bigger businesses that already have big servers.

Ali [4] looked at English Voices in Text-to-Speech Tools: Representation of English users and their varieties from a perspective. The author investigated a corpus of online Text-to-Speech tools and software to discuss their suitability for teaching English according to the plurithic view of English, which throws focus on various users and uses of English. The author did a good work. However, a major drawback of the work was the inability of the author to implement the discussed tools to a model.

Fadi et al [5] looked at Parrotron: An End-to-End Speech-to-Speech Conversion Model and its applications to hearing-impaired speech and speech separation. The authors looked at Parrotron, an end-to-end-trained speech-to-speech conversion model that maps an input spectrogram directly to another spectrogram, without utilizing any intermediate discrete representation. The network is composed of an encoder, spectrogram and phoneme decoders, followed by a vocoder to synthesize a time-domain waveform. The authors did a good work. However, they could not test the developed model on order speech-disorders.

Kaveri et al [6] researched on a Review: Translation of Text-to-Speech Conversion for Hindi Language. According to the authors, language is the ability to express one's thoughts by means of a set of signs (text), gestures, and sounds. Text-to-speech (TTS) convention transforms linguistic information stored as data or text into speech. It is most widely used in the audio reading devices for the deaf and dumb people nowadays. TTS is one of the major applications of NLP (Natural Language Processing). The authors did a good work. However, they could not implement the discussed issue to a model in order to shed light on the discussed issue.

Ming et al [7] looked at Argoverse: 3D Tracking and Forecasting with Rich maps. The authors developed Argoverse a dataset designed to support autonomous vehicle perception tasks including 3D tracking and motion forecasting. Argoverse includes sensor data collected by a fleet of autonomous vehicles in Pittsburgh and Miami as well as 3D tracking annotations, 300k extracted interesting vehicle trajectories, and rich semantic maps. The sensor data consists of 360◦ images from 7 cameras with overlapping fields of view, forward-facing stereo imagery, 3D point clouds from long range LiDAR, and 6-d of pose. However, they could not carry out performance evaluation of their developed Argoverse with other existing system.

Masoud et al [8] researched on the combination of weather stations for electric load forecasting. The authors compared the performance of seven alternative methods with simple averaging as the benchmark using the data of the Global Energy Forecasting Competition 2012. Furthermore, their results demonstrate that some of the methods outperform the benchmark in combining weather stations. In addition, averaging the forecasts from these methods outperforms most individual methods. However, the authors could not implement their developed model with a programming language.

Emmanuel et al [9] looked at Googling Fashion: Forecasting Fashion Consumer Behavior using Google Trends. The authors discussed the current state of Google Trends as a useful tool for fashion consumer analytics; and also showed the importance of being able to forecast fashion consumer trends. However, the authors could not implement their developed model with a programming language.

Files [10] Forecasting and Operational Research: A Review. The study identified particular topics of OR interest over the past 25 years. After a brief summary of the current research in forecasting methods, the study also examined those topic areas that have grabbed the attention of researchers: computationally intensive methods and applications in operations and marketing. Applications in operations have proved particularly important, including the management of inventories and the effects of sharing forecast information across the supply chain. The second area of application

ISSN(Online): 2319-8753
ISSN (Print): 2347-6710

International Journal of Innovative Research in Science, Engineering and Technology

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Visit: www.ijirset.com

Vol. 9, Issue 3, March 2020

is marketing, including customer relationship management using data mining and computer-intensive methods. We could not site any model developed by the author, which would have further shed light on the discussed issue.

Fan [11] looked at Internet-of-Things (IOT) Network System for connected safety and Health Monitoring Applications. The study addressed a hybrid wearable sensor network system towards the Internet of Things (IoT) which is associated with safety and health monitoring applications. The system was aimed at improving safety in the outdoor workplace. However, their developed system was lacking in Multi-Criteria Decision Making.

Rob [12] researched on a Brief History of Forecasting Competitions. The work reviewed history of forecasting competitions, and discussed about their design, implementation, and further information about forecasting. The authors also provided a few suggestions for potential future competitions, and for research about forecasting based on competitions. However, a major drawback observed in the study is the failure to implement the discussed issue with a model.

Leila et al [13] looked at Bayesian Network Model for Flood Forecasting Based on Atmospheric Ensemble Forecasts. The study proposed a Bayesian Network (BN) model to estimate flood peak from 10 Atmospheric Ensemble Forecasts (AEFs). The Weather Research and Forecasting model was used to simulate historic storms using five cumulus parameterization schemes. The BN model was trained to forecast flood peak from AEFs. Mean Absolute Relative Error was calculated as 0.076 for validation data while it was calculated as 0.39 in artificial neural network (ANN) as a widely used model. However, the calculated error of the work showed that there is need for further improvement of their work.

Tanguy et al [14] researched on Improving Wind Power Forecasting through cooperation: A case study of operating farms. The authors proposed a method for wind power forecasting that focuses on interactions between neighbouring wind turbines. The model is a multi-agent system based on a cooperative approach to improve an initial forecast. The work was carried out jointly with meteo-swift; a company specialized in wind power forecasting. The model was evaluated under real conditions on five wind farms currently operated by power producers. An improvement in forecast accuracy was observed compared to the model initially used by the company. However, a major drawback observed in the study is the failure to implement the discussed issue with a model.

Kasun [15] looked at Sales Demand Forecast in E-commerce using Long Short-Term Memory Neural Network Methodology. The authors trained a Long Short-Term Memory network (LSTM) that exploits the non-linear demand relationships available in an E-commerce product assortment hierarchy. The authors also proposed a systematic pre-processing framework to overcome the challenges in the E-commerce business. Furthermore, they also evaluated their forecasting framework on a real-world online marketplace dataset from Walmart.com. Their adopted method achieved a competitive result on category level and super-departmental level datasets, outperforming state-of-the-art techniques. The authors did a good job. However, their model was not designed to process and visualize digital signals used in forex forecasting which is as a result of their model not being able to accept real-time unstructured datasets.

Abdelkarim [16] looked at Short-Term Load Forecasting using Machine Learning and Periodicity Decomposition. The authors proposed a load forecast through the decomposition of the historical time series in relation to the historical evolution of each hour of the day. The output of this decomposition was served as input to different algorithms of machine learning. Furthermore, the authors also tested our model by five machines learning methods, their achieved results are examined with three of the most commonly used evaluation measures in forecasting. However, a major drawback observed in the study is the failure to implement the discussed issue with a model.

Emily [17] researched on Digital Storytelling to Enhance Adults speaking skills in Learning Foreign Languages, a case study: the study examined the relation between adults' engagement in digital storytelling (scaffolded by an interactive learning environment) and their speaking skills and motivation when learning a foreign language.

ISSN(Online): 2319-8753
ISSN (Print): 2347-6710

International Journal of Innovative Research in Science, Engineering and Technology

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Visit: www.ijirset.com

Vol. 9, Issue 3, March 2020

Furthermore, the study used a pre-test, post-test control group design with two groups of 20 Russians who were beginners in learning Greek as a foreign language (n = 40). The 12-h intervention was technology-supported only for the experimental group. Even though the comparison of participants' recorded speech pre- and post-intervention revealed a statistically significant decrease of mistakes made during speech from pre- to post-intervention for both groups, an independent samples t-test to compare the groups' post-intervention speaking performance revealed a statistically significant difference in favour of the experimental group ($t(38) = 4.05$, $p < 0.05$). However, a major drawback observed in the study is the failure to implement the discussed issue with a model.

Ohad [18] looked at Text-based editing of talking-head video. According to the work, editing talking-head vide o to change the speech content or to remove filer words is challenging. We proposed a novel method to edit talking-head video based on its transcript to produce a realistic output vide o in which the dialogue of the speaker has been modified, while maintaining a seamless audio-visual flaw (i.e. no jump cuts). Our method automatically annotates an input talking-head video with phonemes, visemes, 3D face pose and geometry, reflectance, expression and scene illumination per frame. To edit a video, the user has to only edit the transcript, and an optimization strategy then chooses segments of the input corpus as base material.

Yi et al [19] looked at FastSpeech: Fast, Robust and Controllable Text-to-Speech. The authors proposed a novel feed-forward network based on Transformer to generate mel-spectrogram in parallel for text-to-speech conversion. The authors did a good job. However, performance evaluation of their developed model showed latency in the conversion process due to the accumulated complex components of their implementation tools.

III. TEXT-TO-SPEECH CONVERSION

Text to speech, abbreviated as TTS, is a form of speech synthesis that converts text into spoken voice output. Text to speech systems were first developed to aid the visually impaired by offering a computer-generated spoken voice that would "read" text to the user. Optical character Recognition (OCR) is a process that converts scanned or printed text images, handwritten text into editable text for further processing. Testing of device was done on raspberry pi platform. The Raspy is initially connected to the internet through Video Local Area Network [3]

IV. MATERIALS AND METHODS

A. Methodology

The adopted methodology for the proposed system design is the Water-fall Model. Waterfall Model is a sequential model that divides software development into different phases. Each phase is designed for performing specific activity during SDLC (Software Development Lifecycle Methodology) phase.

B. Analysis of the Existing System

The Existing System is a FastSpeech: Fast, Robust and Controllable Text-to-Speech Synthesizer as developed by Yi et al (see figure1). Synthesized speech is usually not robust. Due to error propagation and the wrong attention alignments between text and speech in the autoregressive generation, the generated mel-spectrogram is usually deficient with the problem of words skipping and repeating. Synthesized speech is lack of controllability. Previous autoregressive models generate mel-spectrograms one by one automatically, without explicitly leveraging the alignments between text and speech.

As a consequence, it is usually hard to directly control the voice speed and prosody in the autoregressive generation. Considering the monotonous alignment between text and speech, to speed up mel-spectrogram generation, in the existing system, it is a novel model, FastSpeech, which takes a text (phoneme) sequence as input and generates mel-

ISSN(Online): 2319-8753
ISSN (Print): 2347-6710

International Journal of Innovative Research in Science, Engineering and Technology

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Visit: www.ijirset.com

Vol. 9, Issue 3, March 2020

spectrograms non-auto-regressively. It adopts a feed-forward network based on the self-attention in Transformer and 1D convolution. Since a mel-spectrogram sequence is much longer than its corresponding phoneme sequence, in

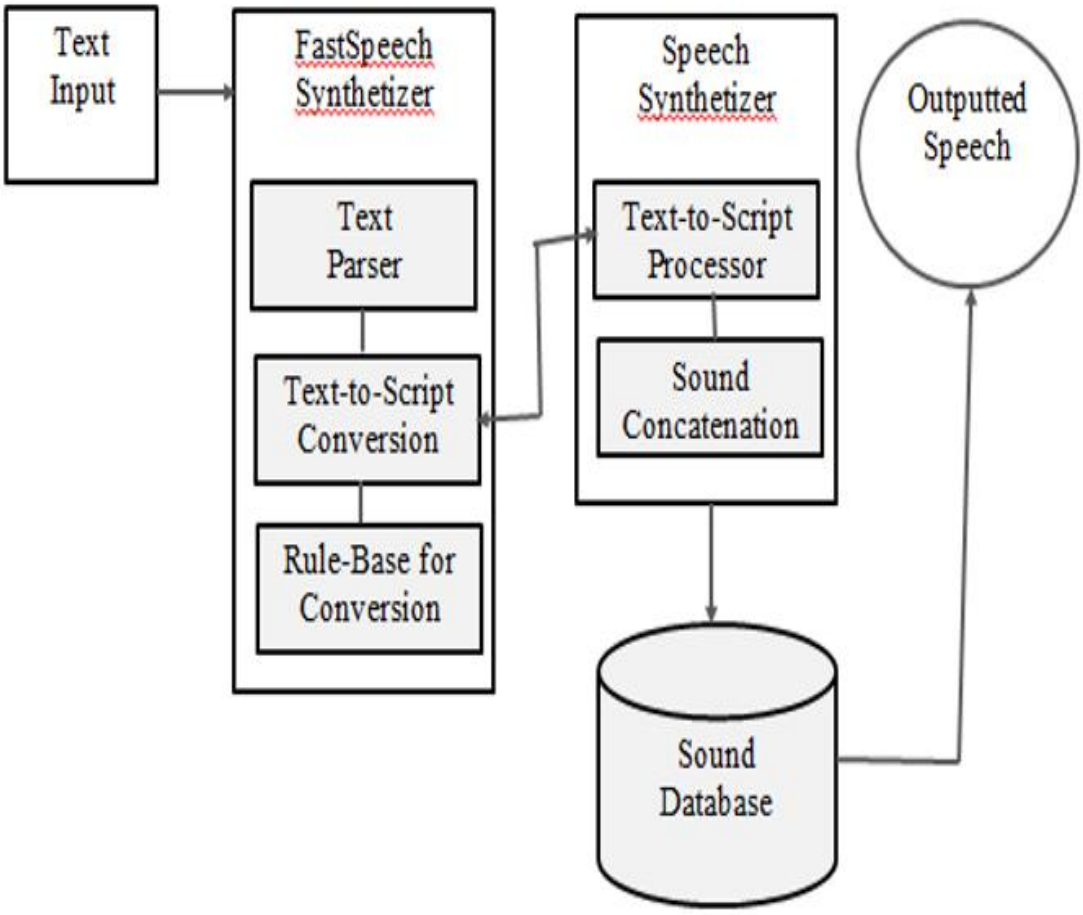

Fig. 1. Existing System Architecture of a FastSpeech Synthesizer
(**Source:** [2])

order to solve the problem of length mismatch between the two sequences, FastSpeech adopts a length regulator that up-samples the phoneme sequence according to the phoneme duration (i.e., the number of mel-spectrograms that each phoneme corresponds to) to match the length of the mel-spectrogram sequence. The regulator is built on a phoneme duration predictor, which predicts the duration of each phoneme.

ISSN(Online): 2319-8753
ISSN (Print): 2347-6710

International Journal of Innovative Research in Science, Engineering and Technology

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Visit: www.ijirset.com

Vol. 9, Issue 3, March 2020

1) Explanation of the Existing System Components

The following components of the Existing System Architecture are:

- **The Text Input:**This component receives the inputted text that is sent to the system for conversion.
- **FastSpeech Synthesizer:**This component takes care of Text conversion in its first phase (i.e. script). Furthermore, it contains other sub-components such as the text-parser, text-to-script conversion, and a rule-base for the conversion. The text-parser accepts the text input and breaks it up into meaningful components which are further converted to scripts. In addition, the rule-base supports the conversion processes.
- **The Speech Synthesizer:**These components receives the first phase of conversion and further concatenates the scripts and compare them to pre-existing sounds in the sound database before integrating the scripts and sounds together before producing a unique output.
- **Sound Database:**This is an organized collection of related sounds in digital format
- **Speech:**This is the final output of the text-to-speech conversion

2) Disadvantages of the Existing System

The following disadvantages of the Existing System are:

- Latency in the conversion process due to the accumulated complex components of their implementation tools.
- Improper word recognition issue that affects the physically impaired.
- Inefficient error detection, recognition and correction process.

C. Analysis of the Proposed System

The proposed improvement to the existing system encompasses an improved text-to-speech synthesizer component using HMM and VB.Net as illustrated in figure 2. Visual Basic (VB) is an event-driven programming language and environment from Microsoft that provides a graphical user interface (GUI) which allows programmers to modify code by simply dragging and dropping objects and defining their behaviour and appearance. VB is derived from the BASIC programming language and is considered to be event-driven and object-oriented. VB is intended to be easy to learn and fast to write code with; as a result, it is sometimes called a rapid application development (RAD) system and is used to prototype an application that will later be written in a more difficult but efficient language.

Forms are created using drag-and-drop techniques. A tool is used to place controls (e.g., text boxes, buttons, etc.) on the form (window). Controls have attributes and event handlers associated with them. Default values are provided when the control is created, but may be changed by the programmer. Many attribute values can be modified during run time based on user actions or changes in the environment, providing a dynamic application. For example, code can be inserted into the form resize event handler to reposition a control so that it remains centred on the form, expands to fill up the form, etc. By inserting code into the event handler for a key-press in a text box, the program can automatically translate the case of the text being entered, or even prevent certain characters from being inserted. Visual Basic can create executable (EXE files), ActiveX controls, or DLL files, but is primarily used to develop Windows applications and to interface database systems.

ISSN(Online): 2319-8753
ISSN (Print): 2347-6710

International Journal of Innovative Research in Science, Engineering and Technology

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Visit: www.ijirset.com

Vol. 9, Issue 3, March 2020

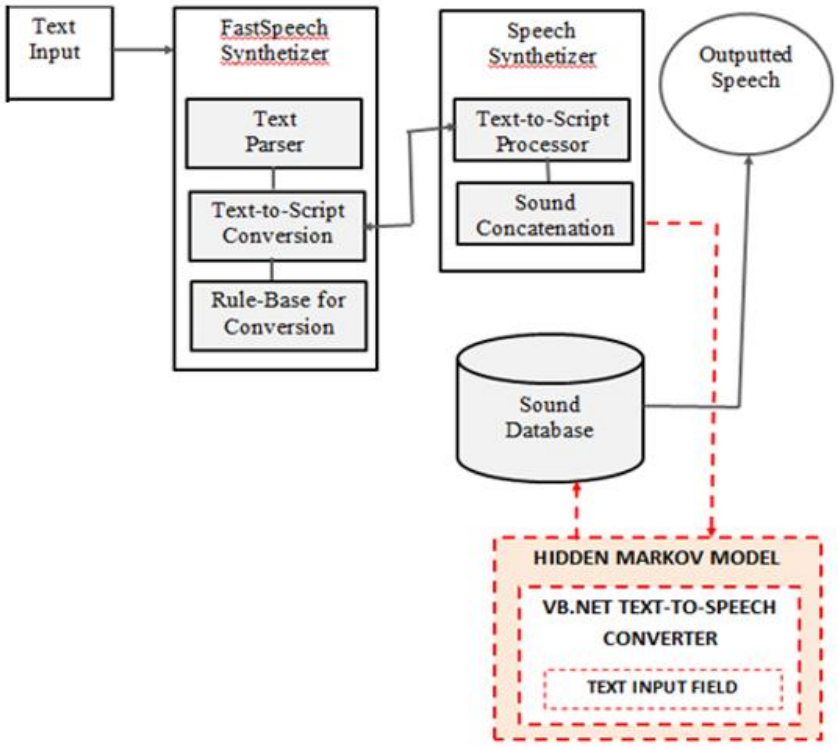

Fig. 2. Proposed System Architecture for an Improved FastSpeechSynthetizer using Hidden Markov Model and VB.Net

1) Explanation of the Proposed System Components

The following component(s) of the Proposed System Architecture is:

- **Hidden Markov/VB.Net Converter:**This component enables latency reduction and further improves the speed of conversion for the Text-to-Speech process. The VB.Net converter has a text-input field in which the text is entered, and further interfaces with the sound database for the final concatenation and conversion.

2) Advantages of the Proposed System Components

The following advantages of the Proposed System are:

i) Latency reduction in the text-to-speech conversion process
ii) Correct read out with the help of high linguistic accuracy.
iii) Ability to read out long texts with a fluent and natural speaking.
iv) Multi-lingual support with different languages and TTS voices.

ISSN(Online): 2319-8753
ISSN (Print): 2347-6710

International Journal of Innovative Research in Science, Engineering and Technology

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Visit: www.ijirset.com

Vol. 9, Issue 3, March 2020

3) Existing System Algorithm

STEP 1: START
STEP 2:DECLARE VARIABLES
LS, TTSS, TI, SC, SO
WHERE:
LS = LAUNCH SYSTEM
TTSS = TEXT-TO-SPEECH SYNTHETIZER
TI = TEXT INPUT
SC = SOUND CONCATENATION
SO = SOUND OUTPUT
QS = QUIT SYSTEM
STEP 3:LS
STEP 4:INITIALIZE TTSS
STEP 5:TTSS = TI + SC
STEP 6:LISTEN TO SO
STEP 7:STOP
STEP 8:QS

4) Proposed System Algorithm

STEP 1: START
STEP 2:DECLARE VARIABLES
LS, TTSS, TI, SC, SO, HMM, FSS, LC
WHERE:
LS = LAUNCH SYSTEM
TTSS = TEXT-TO-SPEECH SYNTHETIZER
TI = TEXT INPUT
SC = SOUND CONCATENATION
SO = SOUND OUTPUT
HMM = HIDDEN MARKOV MODEL
FSS = FAST SPEECH SYNTHETIZER
LC = LANGUAGE CONVERSION
QS = QUIT SYSTEM
STEP 3:LS
STEP 4:INITIALIZE TTSS
STEP 5:TTSS = TI + SC
STEP 6:LAUNCH HMM
STEP 7:
HMM = TTSS + FSS + LC
STEP 8:LISTEN TO SO
STEP 9:STOP
STEP 10:QS

ISSN(Online): 2319-8753
ISSN (Print): 2347-6710

International Journal of Innovative Research in Science, Engineering and Technology

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Visit: www.ijirset.com

Vol. 9, Issue 3, March 2020

V. RESULTS AND DISCUSSION

A. Hardware Requirements

The following hardware components were required for the proposed system design:

- Processor: at least 600-MHZ Pentium III-class processor recommended.
- Hard disk: at least 1GB of available space required on system drive, 3.3 GB of available space required on installation drive.
- Display super VGA (1024 x 768) or higher resolution display with 256 colours
- RAM: Minimum requirement of 512 MB
- CD-ROM Drive
- USB port enabled
- Sound card and speakers (optional)

B. Software Requirements

The following software components were required for the proposed system design:

- Operating system: windows vista, Microsoft windows 7 and 8
- IDEs and Browsers: notepad++/Sublime text, Firefox/Chrome, Opera or UC browser
- VB.Net
- Internet Service Provider (ISP)

C. Outputs and Discussion

Figure 3, shows the Graphical User Interface of the VB.Net Text-to-Speech Synthesizer. It is seen that the more formats of material people can access, the higher their employment opportunities are. There is a higher need for technical skills amongst people who are blind or have low vision. Blind people require supporting tools that meet their specific needs. The programming tool is designed not only for blind users but also for vision impaired and normal vision users. The interface should be designed in a way that complies with standards for vision impaired users and should be user friendly. The programming tool should be able to help a blind user edit, save, compile, debug and run a program. Moreover, the tool should have program templates and intelli-sense (auto-completion) options for user convenience. In order to achieve these objectives, an iterative approach was used. Each part was developed, tested then improved upon and tested again. This meant that usability issues were always found and improved. The tool has been designed to provide voice for blind users and display suitable font, font sizes and colour scheme for vision impaired and normal vision people. Visual Basic (VB) is an event-driven programming language and environment from Microsoft that provides a graphical user interface (GUI) which allows programmers

ISSN(Online): 2319-8753
ISSN (Print): 2347-6710

International Journal of Innovative Research in Science, Engineering and Technology

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Visit: www.ijirset.com

Vol. 9, Issue 3, March 2020

Fig. 3. Graphical User Interface of the VB.Net Text-to-Speech Synthesizer.

Fig. 4. VB.Net GUI Field for Text-to-Speech Conversion

ISSN(Online): 2319-8753
ISSN (Print): 2347-6710

International Journal of Innovative Research in Science, Engineering and Technology

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Visit: www.ijirset.com

Vol. 9, Issue 3, March 2020

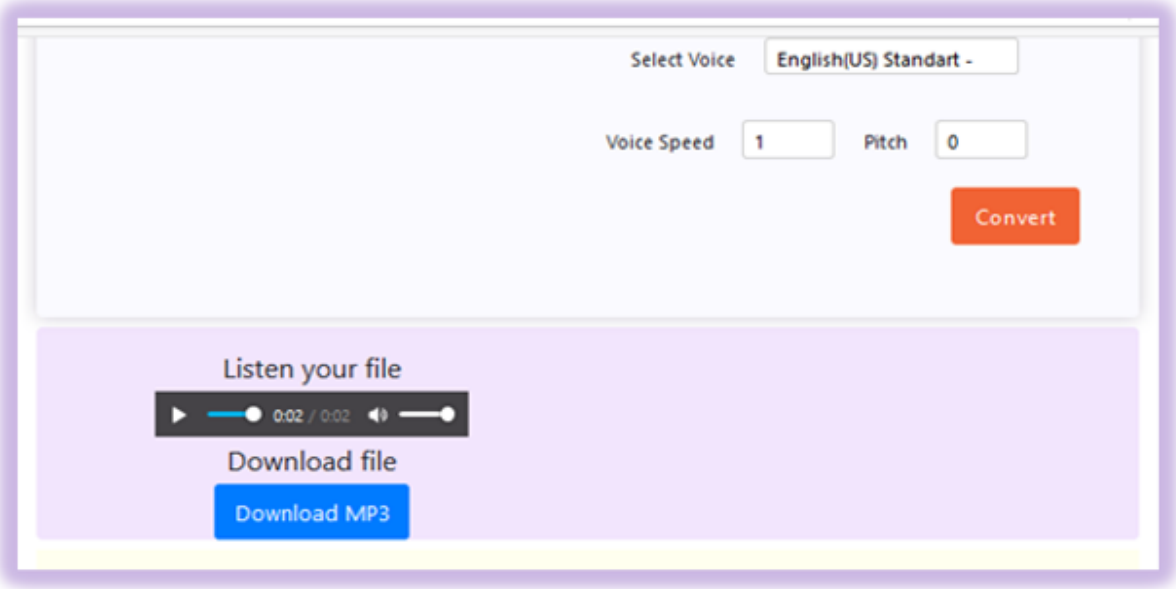

Fig. 5.VB.Net GUI Platform for listening to converted text

to modify code by simply dragging and dropping objects and defining their behaviour and appearance. VB is derived from the BASIC programming language and is considered to be event-driven and object-oriented. From figure 4, the users enters a text in the TTS field, converts and plays or download the text sound in MP3 format (figure 5) A user starts editing a program or loading an existing program using audio code editor. The program on the editor can be saved to a file or can be compiled, debugged and run. For each character entered, the code editor can speak it out. The user can use left, right, up and down arrow key to check any character in the program by voice. The code compiler uses the visual basics software development toolkit (SDK) to compile the program. However, to have voice output, we add code for voice accordingly to the current program using a code modifier then use the visual basics SDK to compile the modified program. For Console application, adding code for voice can be performed by identifying code for text output then add code for voice accordingly. For Windows program, adding code for voice is more complex. Mouse and key event handlers will be added for the user to use mouse or keyboard to design a Windows form. Voice will be output when a control on the Windows form is focused to let the user know what the control is. The compiler also lets the user know if the compilation is successful or if there is a compiling error.

ISSN(Online): 2319-8753
ISSN (Print): 2347-6710

International Journal of Innovative Research in Science, Engineering and Technology

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Visit: www.ijirset.com

Vol. 9, Issue 3, March 2020

D. Performance Evaluation

TABLE I
COMPARATIVE ANALYSIS OF THE EXISTING AND PROPOSED SYSTEMS

SN	YI ET AL (2019)	%	%	PROPOSED SYSTEM (2019)
1.	Time Complexity (TC)	15	21	Time Complexity (TC)
2.	Life-Cycle Assessment (LCA)	7	15	Life-Cycle Assessment (LCA)
3.	Benchmarking (B)	7	12	Benchmarking (B)
4.	Multi-Criteria Decision Making (MCDM)	15	16	Multi-Criteria Decision Making (MCDM)
5.	Risk Assessment (RA)	8	11	Risk Assessment (RA)
6.	Cost Benefit Analysis (CBA)	8	9	Cost Benefit Analysis (CBA)
7.	Speed	5	6	Speed
TOTAL (%)		**65**	**90**	**TOTAL (%)**

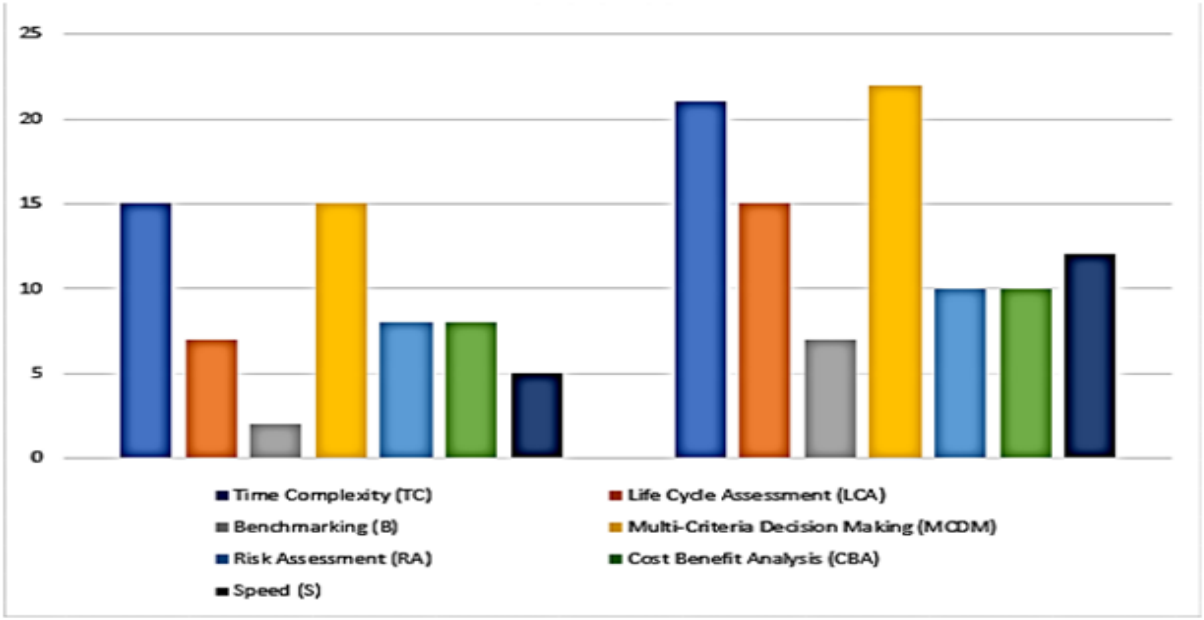

Fig. 6.Performance Evaluation Chart

ISSN(Online): 2319-8753
ISSN (Print): 2347-6710

International Journal of Innovative Research in Science, Engineering and Technology

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Visit: www.ijirset.com

Vol. 9, Issue 3, March 2020

VI. CONCLUSION

In this study, we have developed an Improved Text-to-Speech Synthesizer for the physical impaired using Hidden Markov Model. The study described the functionality, purpose and application of a range of freely available TTS programs or extensions/plug-ins for popular web browsers. They provide many solutions that assist teachers, aides, therapists and parents. In present industry, communication is the key element to progress. In addition, the study concluded on the importance of Text-to-Speech programs in communication. This is because; phone calls, emails, text messages etc. have become an integral part of message conveyance in this tech-savvy world.

REFERENCES

[1] Gerry R., Design and Implementation of an Automatic Speech recognition system and its Application to Information Extraction, *International Journal of Computer Applications, 10(15), 0010 – 0111, 2019*

[2] Yi O, C. Bern, M. Pingh, Optimizing the Functionality of a Voice Recognition System for Assistive Technology, *International Journal of Engineering Technology, 4(4), 1710 – 1718, 2019*

[3] Venkateswarlu J., H. Gomes, P. Dow, Parrotron: An End-to-End, Speech-to-Speech Conversion Model and its application to hearing-impaired speech and speech separation; *arXiv:1904.04169v3[eess.AS], 1-8, 2016*

[4] Ali S., Comparative study of the new generation agile, scalable, high performance Relational Database Management System, *International Journal of Computer Application (IJCA), 48(20), 14 – 21, 2017*

[5] Fadi O., J. Buod, N. Fernilo, Design and Implementation of an Automatic Speech recognition system and its Application to Information Extraction, *International Journal of Computer Applications, 10(15), 0010 – 0018, 2013*

[6] Kaveri R, V. Amadi, L. Ajah, Design and Implementation of Bank Locker Security System Based on Fingerprint Sensing Circuit and RFID Reader, *International Journal of Scientific and Technology Research, 4(7), 116 – 123, 2019*

[7] Ming K., G. Oluoyo, O. Tim, Leveraging the power of Software Integration in online processing and Management, *International Journal of Pure and Applied Science (IJPAS), 10(3), 20-25, 2019*

[8] Masoud K., S. Tome, W. West, Comparative study of the new generation agile, scalable, high performance Relational Database Management System, *International Journal of Computer Application (IJCA), 48(20), 14 – 21, 2019*

[9] Emmanuel O., Y. Liop, D. Gertrude, a rigid approach to online processing in the organization of large datasets Used for online processing, *International Journal of Software Engineering Technology (IJSET) 2(3), 1000 – 1009, 2019*

[10] Files T., Voice Recognition System: Speech to Text, *Journal of Applied and Fundamental Sciences, 4(7), 110 – 200, 2017*

[11] Fan N., An Unsupervised approach for co-channel speech separation using Hibert-Huang Transform and Fuzzy C-means clustering, *International Journal of Speech Technology, 10(16), 1381 – 2416, 2018*

[12] Rob T., Speech Recognition System for Windows Commands, International Journal of Computer Applications, 0975 – 0980, *International Conference on Recent Trends in Engineering Technology, 2012*

[13] Leila G., H. Onman, F. Desl, Voice Recognition System: Speech to Text, *Journal of Applied and Fundamental Sciences, 4(7), 110 – 120*

[14] Tanguy L. U. Geowqu, N. Nibbs, Design and Implementation of Biometric Access Control System Using Fingerprint For Restricted Area Based on Gabor Filter, *International Arab Journal of Information Technology, 8(4), 34-56, 2019*

[15] Kasun N., A Review: Translation of Text-to-Speech Conversion for Hindi Language, *International Journal of Science and Research (IJSR), 3(11), 1027 – 1031, 2014*

[16] Abdelkarim O., Design and Implementation of Commercial Applications of Speech Interface Technology: An Industry at the Threshold, *International Journal of Computer Applications, 92(4), 45 - 55, 2017*

[17] Emily K., Googling Fashion: Forecasting Fashion Consumer Behaviour Using Google Trends, Social Sciences, *MDPI Journal, 2018*

[18] Ohad M., English Voices in Text-to-Speech Tools: Representation of English users and their varieties from world Englishes perspective: Advances in Language and Literary *Studies ISSN: 2203 – 4714, www.alls.aiac.org.au, 2014*