Development of GUI for Text-to-Speech Recognition using Natural Language Processing

Partha Mukherjee
Dept of Computer Application
Techno India College of Technology
Kolkata, INDIA
mpartha1194@gmail.com

Soumen Santra
Dept of Computer Application
Techno India College of Technology
Kolkata, INDIA
soumen70@gmail.com

Subhajit Bhowmick
Dept of Computer Application
Techno India College of Technology
Kolkata, INDIA
bhowmicksubhajit96@gmail.com

Ananya Paul
Dept of Computer Application
Techno India College of Technology
Kolkata, INDIA
ananyatisl@gmail.com

Pubali Chatterjee
Dept of Computer Application
Techno India College of Technology
Kolkata, INDIA
chatterjee.pubali99@gmail.com

Arpan Deyasi
Dept of Electronics and Comm. Engg.
RCC Institute of Information Technology
Kolkata, INDIA
deyasi_arpan@yahoo.co.in

***Abstract*—Natural language processing is a widely used technique by which systems can understand the instructions for manipulating text or speech. In the present paper, a Text-to-speech synthesizer is developed that converts text into spoken word, by analysing and processing it using Natural Language Processing (NLP) and then using Digital Signal Processing (DSP) technology to convert this processed text into synthesized speech representation of the text. Here we developed a useful text-to-speech synthesizer in the form of a simple application that converts inputted text into synthesized speech and reads out to the user which can then be saved as an mp3 file.**

Keywords—Text-to-speech; Natural language processing; Speech synthesizer; Speech recognition; Signal transformation

I. INTRODUCTION

Speech is the first primary mode of communication in Human Intelligent System (HIS) where NLP plays a role with many aspects of the field deal with linguistic natures of computation. NLP is a way of research and application that explains how a system (mainly a computer) can be used to understand, identify and manipulates a natural language. TTS is the automatic conversion which configures the concept of speech recognition, speech analysis, speech synthesis, speech tuning, speech alteration etc. Here TTS use to convert a text into speech that resembles, as closely as possible for a native speaker of the language who trying to read that text. TTS is the technology by which a computer can speak to user and give the computed information. TTS system acquires the text as input and then a computer algorithm which 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. This TTS system also worked upon Natural Language Generator (NLG).

Kanisha [1] explained an innovative way for STS for visually impaired people through voice signal. Optical character recognition process is introduced in TTS system by Thu [2] with the intension to develop image to speech conversion system. Pros and cons of interactive voice response system are reviewed by researchers [3] which are used in different day-to-day applications. Different methods are also compared [4] for online systems in terms of user efficiency. Concatenation method is further developed for local language [5] with better accuracy which is an extension of work of Kanisha [1]. Different speech synthesis processes are also discussed by Htun and his co-workers [6]. Text inside any image is recently tracked by camera, and converted to speech by Patil *et. al.* [7] for multilingual language. It is pointed out that this method is very useful for blind persons for detecting currency notes [8]. Finger-mounted camera provides a great relief in these cases [9].

The TTS synthesis is a procedure where first text analysis and then generate the waveform of speech. Here it converts this phonetic and prosodic information into a wave form based upon approximation formula. The amplitude of each signal which forms from the speech waves is measured and creates the proper speech. Those speech waves are linguistic approach of texts. Sometimes these forms are linguistic or non-linguistic in nature. In the present paper, a synthesizer is developed which, apart from TTS conversion, saves the file into mp3 format.

II. OVEREVIEW OF SPEECH SYNTHESIS

Speech synthesis is one of the artificial computations of producing human voice. A TTS system converts any text followed by grammatical language into speech. Synthesized speech is a collection of small pieces of recorded speech which are stored in a knowledge base (KB). This KB System differs in the size based on the stored speech units. That system also maintains the speech quality based upon its algorithm by which it analysed the tree of speech units for better clarity. Alternatively, a synthesizer can be proper in

978-1-5386-5550-4/18/$31.00 ©2018 IEEE

such a way that the system must control the vocal peach and distinguished it from other human voice list to create a completely "different synthetic" voice output. These are the significant features for a good quality of a speech synthesizer.

A TTS system (or "voice generated-engine") is composed of one interface as a front-end and a back-end. The first interface converts raw text containing alpha-numeric symbols like numbers and abbreviations in terms of speech into the equivalent of out words. Here analysis of text includes various features such as recognition of text unit, normalization of text unit pattern, pre-processing, etc. The front interfaces always engage with phonetic conversion to each unit, and divides and marks to form a speech tree or pattern tree using the speech unit which configures the tune and rhythm through phrases, clauses, and sentences. This process of transcriptions is known as text-to-phoneme (TTP) or grapheme-to-phoneme (GTP) conversion [6]. These two conversions are together known as symbolic linguistic representation which is the desired output of TTS. Whereas in the back side the symbolic linguistic representation converts into sound. Sometimes this back end computes pitch analysis, contour analysis, rhythm analysis etc., for output speech.

Speech synthesis is done in many ways such as Concatenative Synthesis (Unit-selection Synthesis, Diphone Synthesis, and Domain Specific Synthesis), Formant Synthesis, Articulatory Synthesis, HMM-based Synthesis, sinewave Synthesis [6-8] etc.

III. TTS Synthesis Model

TTS synthesis takes place in several steps. The TTS systems gets a phrase or collection of phrases as input, which it first must analyses and then converts into a phonetic description. Then in a further step it generates the set of tune, pitch and rhythm (known as prosody). From the following data which are produced in the form unit of speech tree or speech space, it can form a speech signal.

The structure of the TTS synthesizer can be broken down into major modules:

NLP module: It produces a phonetic transcription of the text read, together with prosody.

DSP module: It generates the symbolic representation which receives from NLP into audible and intelligible speech known as NLG.

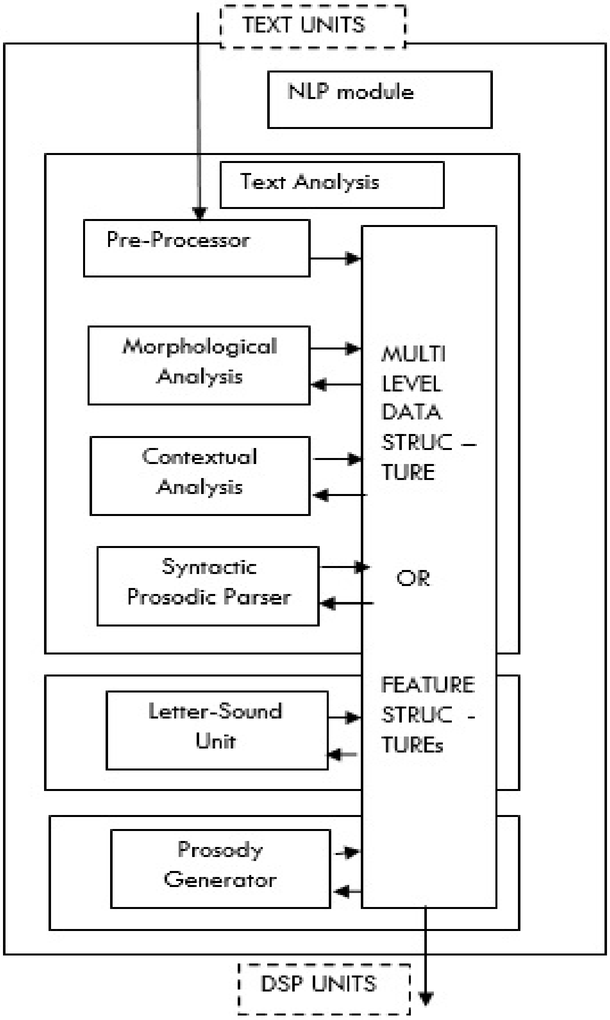


Fig. 1. TTS synthesis model

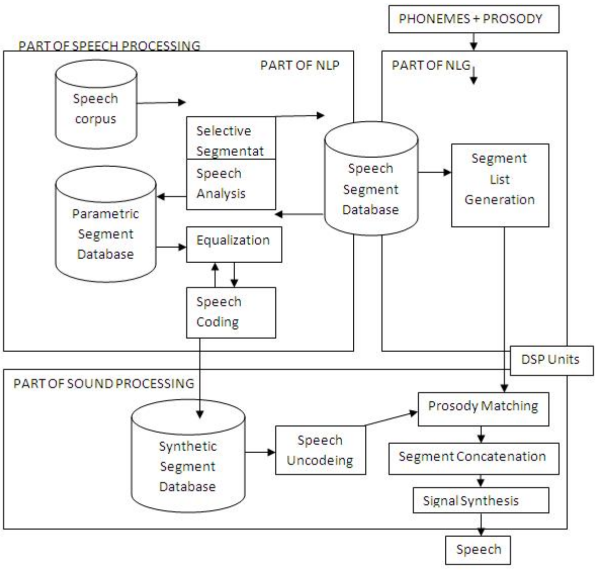


Fig. 2. Speech Synthesis Model

IV. Design and Implementation

Our designed software is called the TTS Gramaty, a simple application with the text to speech functionality. The system was developed using C# language upon .Net Framework 3.5.

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 parameters for conversion either via file or direct keyboard input.

The second module, the main conversion engine which integrated into the main module is for the acceptance of data hence the conversion.

TTS Gramaty (TTSG) converts text to speech either by typing the text into the text field provided or by coping from an external document in the local machine and then pasting it in the text field provided in the application. It also provides a functionality that allows the user browse and open a text document in the machine. TTSG then loads the document's text in the text area of the application and the reading procedure starts automatically.

TTSR contains an exceptional function that gives the user the choice of saving its already converted text to any part of the local machine in an audio format; this allows the user to copy the audio format to any of his/her audio devices, so that they can hence forth treat it as an audio book.

The following figure depicts the loading procedure of the TTS Gramaty.

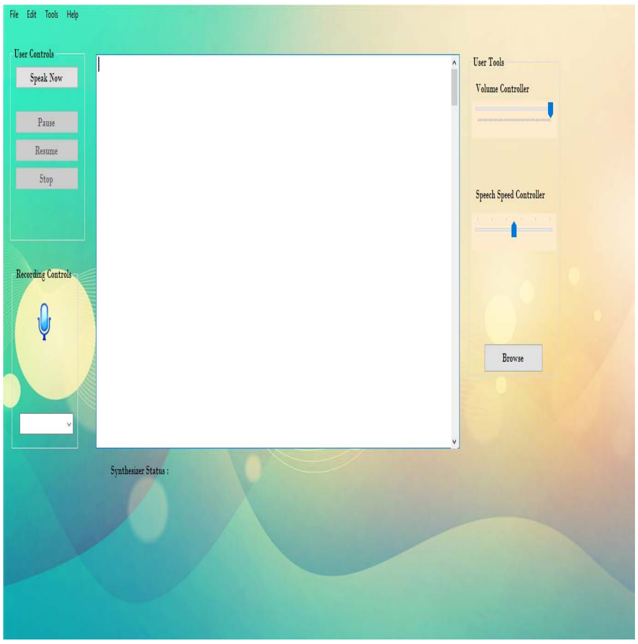
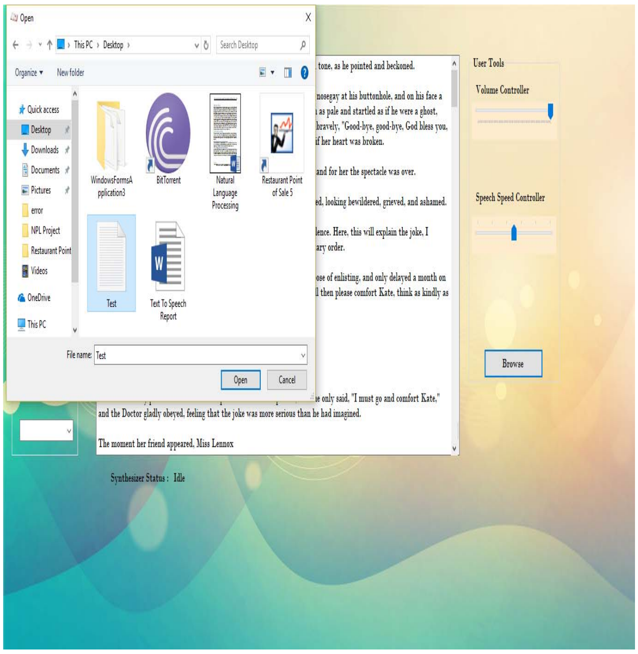


Fig. 3. Screenshot of the TTS Gramaty Interface

This is the default view of the application. This screen appears in the full screen mode when the application is launched. As we can see, there are several options and buttons present in the application window, each having different functions. The with text area in the middle is there, where we can enter our text, (if we want to manually enter some random data over there as input. If not, then we can use the BROWSE button to import some text file content. In either case, we must click on the SPEAK NOW button to start the reading process.

Let's see how the import occurs using browse button:



File name: Test

tone, as he pointed and beckoned.

nosegay at his buttonhole, and on his face a
as pale and startled as if he were a ghost,
bravely, "Good-bye, good-bye, God bless you,
if her heart was broken.

and for her the spectacle was over.

ed, looking bewildered, grieved, and ashamed.

lence. Here, this will explain the joke, I
ary order.

ose of enlisting, and only delayed a month on
l then please comfort Kate, think as kindly as

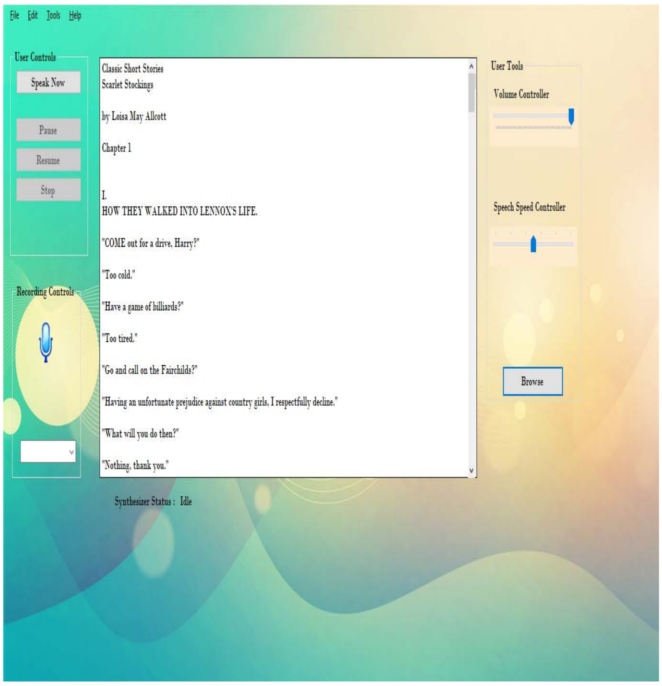
e only said, "I must go and comfort Kate,"
and the Doctor gladly obeyed, feeling that the joke was more serious than he had imagined.

The moment her friend appeared, Miss Lennox

Synthesizer Status : Idle

Fig. 4. Import of file using TTS Gramaty

Now open the .txt file



Classic Short Stories
Scarlet Stockings

by Loisa May Allcott

Chapter 1

I.
HOW THEY WALKED INTO LENNOX'S LIFE.

"COME out for a drive, Harry?"

"Too cold."

"Have a game of billiards?"

"Too tired."

"Go and call on the Fairchilds?"

"Having an unfortunate prejudice against country girls, I respectfully decline."

"What will you do then?"

"Nothing, thank you."

Synthesizer Status : Idle

Fig. 5. Opening of file using TTS Gramaty

When opened, the contents of the .txt gets automatically loaded into the text area of the application.

Now, before we get any further, let us get some knowledge about the controls.

As we can see, there is a volume controller to control the audio output volume, a Speech Speed Controller, to control the speech rate. Speak now, Pause, Resume and stop buttons with their respective purpose. Below that we have a microphone icon that allows us to generate the audio copy of the total text document. Below that we have the dropdown list, from where we can select the quality of the output audio file. At last we have the Synthesizer Status that displays the status of the synthesizer, is it idle, paused or running.

V. CONCLUSION

TTS synthesis is a flexible robust dynamic growing aspect of modern computer era and it is increasingly playing a more significance role in the way we interact with the system and interfaces which is based on platform independent concept. 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 conversion of one language to other make text to speech technology more accessible to a wider range. Accuracy of the software is excellent in the context of its ability to work in real-life environment. We also have plans to make it a web based real-time synthesis system, so that its uses can get more expanded.

References

[1] J. Kanisha, G. Balakrishanan, “Speech Transaction for Blinds Using Speech-Text-Speech Conversions”, Communications in Computer and Information Science book series (CCIS), vol 131, part I, pp. 43-48, 2011

[2] C. S. T. Thu, T. Zin, “Implementation of Text to Speech Conversion”, International Journal of Engineering Research & Technology, vol. 3(3), pp. 911-915, 2014

[3] P. S. Shetake, S. A. Patil, P. M. Jadhav, “Review of Text To Speech Conversion Methods”, International Journal of Industrial Electronics and Electrical Engineering, vol. 2(8), pp. 29-35, 2014

[4] P. Khilari, V. P. Bhope, “A Review on Speech To Text Conversion Methods”, International Journal of Advanced Research in Computer Engineering & Technology, vol. 4(7), pp. 3067-3072, 2015

[5] A. Joshi, D. Chabbi, M. Suman, S. Kulkarni, “Text To Speech System for Kannada Language”, International Conference on Communications and Signal Processing, 2015

[6] H. M. Htun, T. Zin, H. M. Tun, “Text To Speech Conversion using Different Speech Synthesis”, International Journal of Scientific & Technology Research, vol. 4(7), pp. 104-108, 2015

[7] S. Patil, M. Phonde, S. Prajapati, S. Rane, A. Lahane, “Multilingual Speech and Text Recognition and Translation using Image”, International Journal of Engineering Research & Technology, vol. 5(4), pp. 85-87, 2016

[8] D. B. K. Kamesh, S. Nazma, J. K. R. Sastry, S. Venkateswarlu, “Camera based Text to Speech Conversion, Obstacle and Currency Detection for Blind Persons”, Indian Journal of Science and Technology, vol 9(30), pp. 1-5, 2016

[9] B. Sanjana, J. R. Parvin, “Voice Assisted Text Reading System for Visually Impaired Persons Using TTS Method”, IOSR Journal of VLSI and Signal Processing, vol. 6(3), pp. 15-23, 2016