Visual Assistant Using TTS

[1]Prachi Bandekar, [2]Shubham Jadhav, [3]. Omkar Shirgaonkar, [4]. Radiya Budye [5]. Dr. Sharada Chougule

Department of Electronics & Telecommunication,
Finolex Academy of Management & Technology, Maharashtra, India

Abstract: TTS which is the Text to Speech Converter which is based on application that reads a text file or document to the user. Nowadays looking at screen constantly can affect the user eyes or this software can be great use to them as this will help them to read the text document for them. This is the software project that challenge the visual world to read and understand various file and documents. This software will be great help for blind people as they cannot read the document or file so this can be assistant to them. Even this can be useful for the people who can't speak as they can type whatever they want to say and software will give voice to them. And also, it will be appreciable for the people who are uneducated as they can't read it so it will be assistant to them too. So, this software will be a considerable achievement and growth towards the future scope for the people who can't see, speak or who are uneducated. By taking this into consideration, this project "Visual assistant using TTS" will be a new hope for blind and uneducated people. The main concept is to help people as much as possible by giving an visual assistant to them.

Keyword-text to speech- linguistic analysis; natural language processing; prosodic prediction

INTRODUCTION

This work importance is on developing and creating of software application using TTS. The project has been done by considering the need for blind and uneducated people which will be helpful for them as assistant. Many people in India as problem because of disability as they can't see, can't speak and who are not educated. This software application is made by considering the problem that can affect user eyes like when a person can't continuously see the screen so this can help them in reading document. A Text to speech converter system consist of two main parts a front end and back end. The front end consists two major tasks, work done by front end is to convert raw text containing symbols like numbers and short form into identical of writtenout words. This process is known as normalization, pre-processing, or tokenization. Further front end assigns phonetic transcriptions to every word and divides and marks the text into prosodic units, as clauses, sentences and phrases. The process in which phonetic transcriptions assigning to words take place is called as text-to-phoneme or grapheme-to-phoneme conversion. Further step is together made up of phonetic transcriptions and prosody information which symbolic linguistic represent that is the output by front end. Then the back end which is referred to synthesizer then it is converts symbolic linguistic representation into sound. In some systems, this part includes computation of target prosody which is then imposed on output speech. To make this application more useful it has OCR technique in it. As OCR technique allows TTS system to read the text aloud from image. E.g., If someone click a photo and the image have words on it then it turned into audio. This design software allows the feature giving option of converting image text into speech.

LITERATURE SURVEY

When the idea of this system is come for this work, searched for the different methods that can be implemented to read and understand the various documents as well as assisting to the visually impaired and those who cannot speak. So started searching for various text and speech conversion related journals to find any relevant information. Some of them are stated below:

• Deepshikha Mohanta, Bidisha Sharma, P Sarmah and SRM Prasanna [1] system stated that An American English TTS, which is highly understandable for American users may be not easy for Indian listener. There are so many people in India speak in non-native English. So, this paper is developing an Indian accent English Text to Speech which is important for the people of India. To modifying the existing Carnegie Mellon University Pronunciation dictionary by adding the Indian Accent English pronunciation.

• Selvaraj Chithra & N. Bhalaji [2] system stated that This research is highly developed to help the visually impaired. A scanner which is handheld device is used to scan the text which is to be read. The scanned image from the handheld scanner is sent to a smartphone device via Bluetooth. Then OCR is used to extract scanned text from that image issnd convert them into the speech. This OCR is trained with handwritten database. This paper is support two regional languages in addition to the English, i.e Hindi, Bengali

• Shivangi Nagdewani & Ashika Jain [3] system stated that the objective of this research is to draw the structure for the development of an interactive voice response-based mailing system that enables users to manage their e-mails by using audio commands only. For this system they are using Speech to text (STT), Text to Speech (TTS), Hidden Markov Model (HMM), Interactive Voice Response (IVR), Artificial Neural Network. The system can implement for different languages like English, Hindi, Punjabi etc. depending upon the user requirements.

• Pobasara Jayawardhana, Achalo Aponso, Naomi Krishnarajah [4] system stated that the quality of Text-tospeech, basically quality of speech synthesizer is decided by its similarity to the human voice and by its ability to understand clearly. This research is investigated Text-to-speech algorithm based on deep voice which is an attention based, fully convolutionally method. In this method human vocal production mechanism which is eligible of taking text and converts that as speech.

• Pawan Pooja G. Kavya D, R. Chaithra, Radhika A.D [5] system stated that Text-to-speech conversion has wide variety of applications. Neural Network is circuit of neurons composed of artificial neurons and nodes. These artificial networks are used for Predictive modeling, adaptive control, and applications where they can be trained via dataset. In this paper we have explained Neural Networks approach using different types of Neural Networks like DNN, RNN, and CNN, etc.TTS conversion using neural networks approach is more accurate when compared to all other approaches. Since Neural Networks are more scalable and proper computations of the results can be achieved.

• Visamya V & Darvin Reynald J [6] system stated that Natural Language Processing (NLP) is an area of application and research that explores how computers can be used to understand and manipulate natural language speech or text to do useful things. This research aims to collect knowledge on how human beings use and develop natural languages to perform specific tasks so that appropriate tools and techniques can be developed.

METHODOLOGY

Text-to-Speech conversion is a method that scans and reads English alphabets that are in the image using OCR technique and changing it to voices. Text to speech is program that lets you type in English text and then play it is an audio. Instantly convert Desire takes to audio. Converting files into speech format. File can be pdf, text or image. The method has introduced beneficial and effective technique that enables user to hear the content of text image instead of reading through them. It is useful for the vision impaired people as they cannot read but understand the message by hearing. It is also useful for the people who understand a language but they cannot speak fluently or write that language. Text to speech converting technique consist mainly two modules. The image processing module and Voice processing module. For image processing OCR technique is used. OCR convert JPG to text form and for Vice processing TTS techniques are used. while processing model changes the text into sound and process it with the specific characteristics so that sound can be understood.
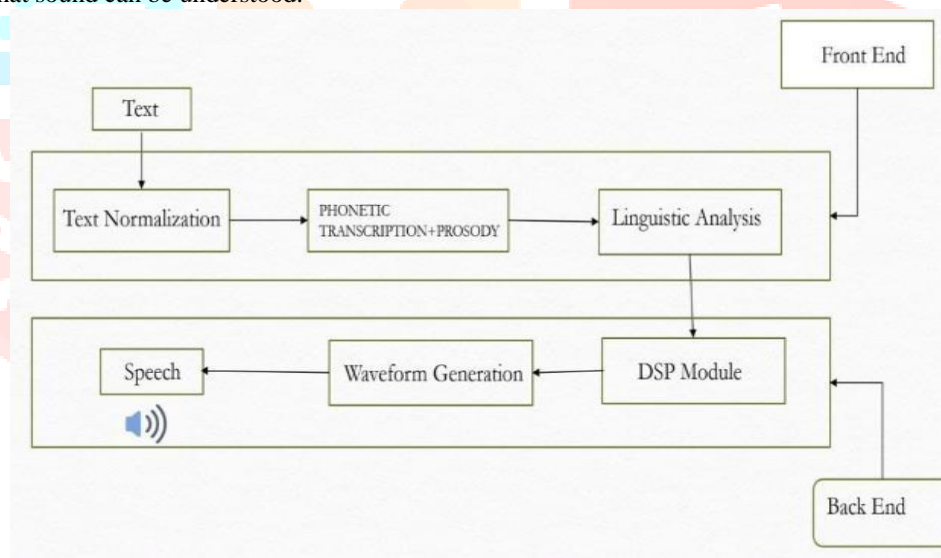

Figure 1-Proposed System

Different implementations of textual content - to speech device exist. This segment discusses a number of the ideas on which those structures are built. Generally, a textual content - to - speech system split into three parts: a linguistic, a phonetic and an acoustic part. First, a regular textual content is entered to the device. linguistic module converts this text right into a phonetic representation. From this representation, the phonetic processing module calculates the speech parameters [4]. Finally, an acoustic module makes use of those parameters to generate a artificial speech signal. Using this product consumer can concentrate his entered textual content or decided on textual content. He or she will give input file textual content which may be txt, pdf format. User can concentrate the entered textual content withinside the interactive mode. User also can concentrate to the textual content written in an image.[5] Software Requirements used are Windows XP and every other contemporary edition, Python technology.
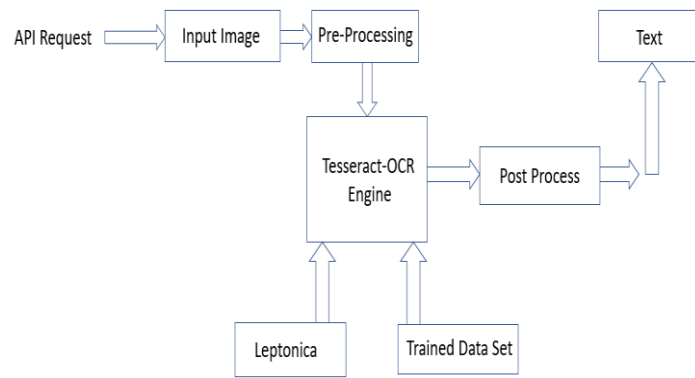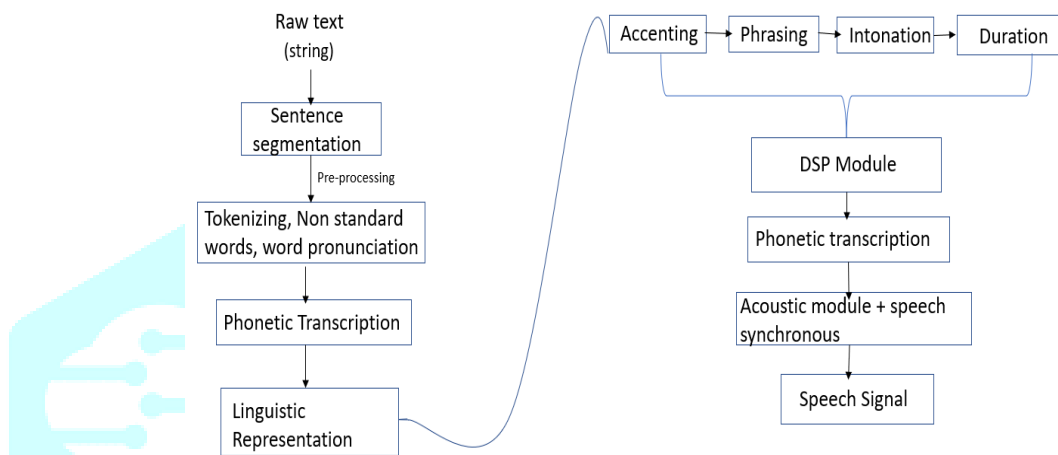
Figure 2-OCR technique

Figure 3- Flowchart

As you can see, the TTS system consists of two main parts: the first is the front of the and the second is the rear. Front end operation called as Text to phonetics and Back-end operation Text to prosody. Give the input in form of text to system. The text analysis stage is extremely difficult at this stage needs to produce all the information required for the DSP module (Digital Signal processing) for producing speech. It converts raw texts containing symbols like numbers and abbreviation into the equivalent of written-out words. It identifies numbers, abbreviation, and expand them into full text.[3]

In pre-processing there are 3 steps-
1] Tokenizing: - In its simplest form is achieved by splitting text at white spaces and at punctuation marks.
2] Non-standard words: - This are tokens to be expanded to an appropriate orthographic form before Linguistic Analysis.
3] Word pronunciation: - Once the sequence of words has been generated using text tokenizing their pronunciation determined. A morpho syntactic analyzer tags the speech with various identifies such as prefixes, roots and suffixes and organizer the sentences into syntactically related group of words, such as noun, verbs and adjective.

After that assigning phonetic transcription to text is called text-to-phonetic. Phonetic transcription means visual representation of speech sound. There is another factor here is prosody. The prosody assigns to certain properties of speech signal such as audible changes in the pitch. For e.g., loudness, duration, rhythm. Naturalness of speech determined in terms of prosody. After that phonetic transcription and prosody information add and make up the symbolic linguistic representation. After that linguistic further broke down in four processes: -[3]

1] Accenting: - Accent or stress assignment is based on the concept of the word for e.g., noun, verbs, etc. This information is used for predicting the duration.
2] Phrasing: - Sentences are broken down into phrases units and phrase boundaries are allot to the text.
3] Intonation: - Intonation clarifies types and meaning of the sentences. In addition, intonation as well conveys information about the speaker characteristics (such as gender and age) and emotions. The intonation module creates a pitch contour for the sentences.
 4] Duration: - Segmental duration is an essential feature of prosody. It affects the overall rhythm of the speech, stress, emphasis.

So that is output by the front end. The back end referred as a synthesis that converts the symbolic linguistic representation into sound. With the help of digital signal processing module (DSP). The DSP module uses the phonetic transcription and prosodic information produces by linguistic analysis module to produce speech. Linguistic module converts text into a phonetic representation from this representation, phonetic processing module are, the speech parameter. And after that Acoustic module uses these parameter and work with, speech synchronous to generate speech signal. After that speech synchronous takes signal from DSP module and create a waveform and we get the output in the form of speech.
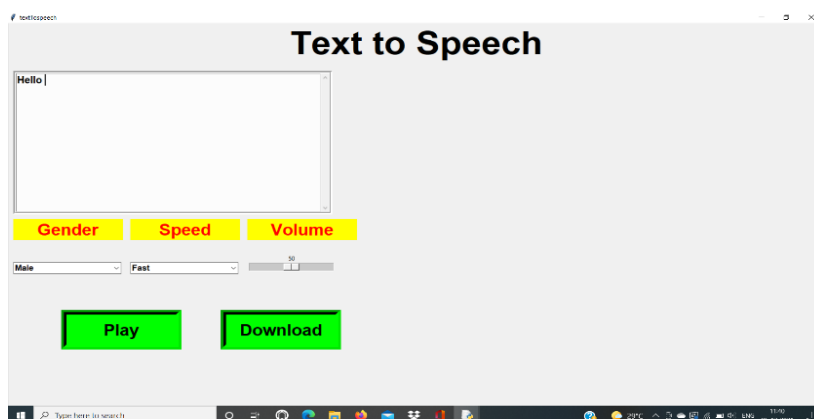
RESULT

Figure 3

CONCLUSION

After developing this project using various technology result that it can used by many peoples in there day to day life. As this can be great help for them. The user can get voice by just typing what he/she want to say which can incentive blind people to come forward and also user get voice only by selecting pdf file what he/she wanted to listen. This technology is blessing for the people who can't speak they can get voice for themselves. Even this technology is utilized by various purposes like announcement in railway station, car navigation. Thus, this will be more helpful for the growth of our society/country by developing and helping people to express.

REFERENCE

1. Deepshikha Mohanta, Bidisha Sharma, P Sarmah and SRM Prasanna "Text to Speech Synthesis System in Indian English",2016 IEEE Region 10 Conference (TENCON).
2. Selvaraj Chithra & N. Bhalaji "Review and Comparison of Enhanced portable text to speech converter for visually impaired", 2018.International Journal of Intelligent System Technology and Application.
3. Shivangi Nagdewani & Ashika Jain "Review for speech-to-text and text-to-speech conversion" May 2020.International Research Journal of Engineering and Technology (IRJET).
4. Pobasara Jayawardhana, Achalo Aponso, Naomi Krishnarajah "An Intelligent Approach of Text-To-Speech Synthesizers for English and Sinhala Languages." 19 IEEE 2nd International Conference on Information and Computer Technologies (ICICT)978-1- 72813323-2/19/©2019 IEE.
5. Pawan Pooja G, Kavya D.R. Chaithra, Radhika A.D "Survey on text-to-speech Kannada using Neural Networks",2019 International Journal pf Advance Research, Ideas and Innovation in Technology
6. Visamya V & Darvin Reynald J "Language Processing using Python ", International Journal of Scientific & Engineering