International Journal of Scientific Research and Engineering Development-– Volume 5 Issue 2, Mar-Apr 2022

Available at www.ijsred.com

RESEARCH ARTICLE OPEN ACCESS

DIGITAL AUDIOBOOK

Mrs. M.Vijaya Lakshmi[1](M. Tech), Puchakayala Ajay[2] , Dalli Bhagya Sree[3], Pothu Bhanu Prakash[4], R Devi[5] , Dharani S[6]

[1](Associate Professor ECE Dept, Siddartha Institute Of Science And Technology,Puttur,A.P,India)

[2](B.Tech. IV Year Student, ECE Dept, Siddartha institute of science and Technology, Puttur, A.P, India)

[3](B.Tech. IV Year Student, ECE Dept, Siddartha institute of science and Technology, Puttur, A.P, India)

[4](B.Tech. IV Year Student, ECE Dept, Siddartha institute of science and Technology, Puttur, A.P, India)

[5](B.Tech. IV Year Student, ECE Dept, Siddartha institute of science and Technology, Puttur, A.P, India)

[6](B.Tech. IV Year Student, ECE Dept, Siddartha institute of science and Technology, Puttur, A.P, India)

(Email: bp90120@gmail.com)

--************************----------------------------------

Abstract:

The idea of this paper is to build an automated virtual reader. In this modern era, there is an urge for an automated reader which is cost-effective, accurate, and also portable at the same time. The major project involves the complete design of the Digital Audio Book.

In this digital audiobook, we aimed to develop a convertor and a translator of an eBook. An eBook is a digital representation of a printed book. In this, we can convert an eBook into an audio source. Web Speech moduleis used to convert the text to audio format and vice versa.Here we are with the Digital audiobook. It has features like text to speech, speech to text, and a translator. During the pandemic period, everything went through online itself.Especially in Educational Institutions and they provide notes to students in the form of text.It's very difficult to study by looking and learning on a screen. So, with the help of Digital audiobooks, we can overcome these problems.

Keywords **—Web Speech API , JavaScript, HTML, TTS, SST, Translator**

--************************----------------------------------

I. INTRODUCTION

For many years, scientists and engineers have had the speech production process with the goal of building a system that can start with text and produce speech automatically. Sometimes we prefer listening to the content instead of reading we can multitask while listening to the critical file data.

Whenever you try to translate a word or a sentence from a certain language to another language. Web Speech provides many APIs to convert text to speech, speech to text, and translator. It is very easy to use the tools and provides many built-in functions which are used to save the text file an mp3 file.

The text to speech synthesis is Synthesizing speech from the text (TTS). TTS systems have based on the complex pipeline. A speech synthesizer is one of the major applications of NLP. The conversation of text to speech involved three important stages: text analysis, text processing, and waveform generation, i.e., speech formation.

A speech synthesizer is an application that converts written text into speech, the user enters a text and gets output as a sound. In this survey, Speech Synthesis is becoming one of the most important steps toward improving the human interface to the computer. The objective of text to speech is to convert arbitrary text into spoken waveform. The quality of a speech synthesizer is based on two factors naturalness and intelligibility.

II. LITERATURE SURVEY

In 2018 a paper was published by Sneha.C , S.B. Gundre about "OCR Based Image Text toSpeech Conversion Using MATLAB" in this project There are millions of blind people inthe world who are visually impaired. The disability to read has a large impact on the life ofvisuallyimpairedpeople.TheProposedsystemiscost-efficientandhelpsthevisuallyimpaired person to hear the text. The main idea of this project is optical Character recognitionwhich is used toconverttextcharactersinto the audiosignal.

In 2017 a paper was published by H Rithika, B Nithya Santhoshi "Image text to speech conversion in the desired language by translation with Raspberry Pi"in this project The mainproblemincommunicationislanguagebiasbetweenthecommunicators.Thisdevicebasically can be used by people who do not know English and want it to be translated into their native language

In 2019 a paper was published by Jing-Xuan Zhang, Zhen-Hua Ling, Yuan Jiang "Improving sequence-to-sequence voice conversion by adding text supervision".Voice conversion (VC) is a task that alters the voice of a person to suit different styles while conserving linguistic content. Previous state-of-the-art technology used in VC was based on the sequence-to-sequence (seq2seq) model, which could lose linguistic information.

In 2021 a paper published by D. C. Tran, H. S. Ha, and A. Khalyasmaa, "A Question Detection Algorithm for Text Analysis," In this paper, an FPT.AI-based text-to-speech (TTS) application is developed that converts Vietnamese text into spoken words. The application is developed based on Django for Python and in the form of an interactive web page which is connected to an FPT.

III. EXISTING SYSTEM

Previously so many members used the tts module in order to convert the text to speech and some members used Optical Character Recognition to convert the text to speech. So many of the authors using the tts module in python for thisproject. Some authors give features like text to speech and speech to text. But tts only reads the typed text it does not read the text. They encouraged todo this text to speech for impaired persons

IV. PROPOSED SYSTEM

In this proposed method system, we are overcome to solve the problems faced by the students during the learning process. Here we are with the Digital audio book. It has features like text to speech, speech to text, and a translator. During the pandemic period, everything went through online itself. Especially in Educational Institutions and they provide notes to students in the form of text.It's very difficult to study by looking and learning on a screen. So, with the help of Digital audio books, we can overcome these problems

A. BLOCK DIAGRAM

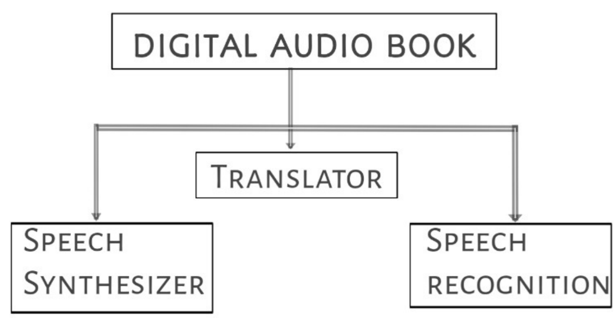

Fig1:*Digital Audio Book*

Digital Audio Book consists of a speech synthesizer, speech recognition, and translator. we are going to do this with the help of a web speech API in JavaScript.

B. SPEECH TO TEXT

The speech is converted into text by using web speech API.

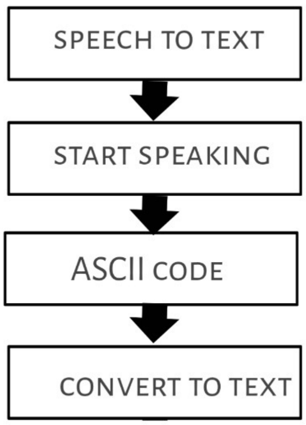

Fig 2: *speech to text*

C. TEXT TO SPEECH

The text is converted into speech by using web speech API.

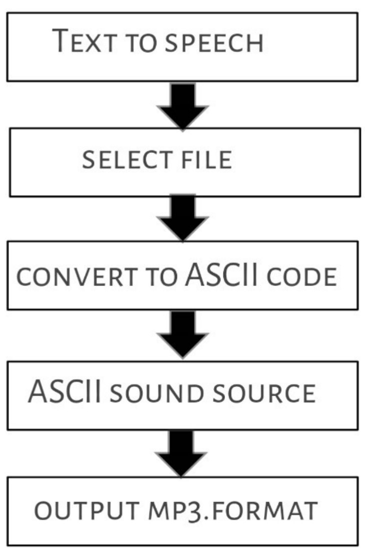

Fig 3: *text to speech*

D. TRANSLATOR

The translator can translate the information provided by the user

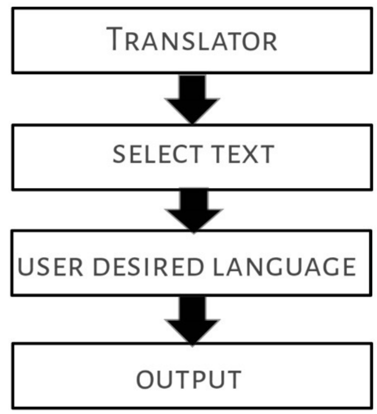

Fig4: *Translator*

V. SOFTWARE USED WEBSPEECH API

The Web Speech API enables you to incorporate voice data into web apps. The Web Speech API has two parts: Speech Synthesis (Text-to-Speech),and speech Recognition (Asynchronous Speech Recognition).

A. SPEECH SYNTHESIS

Let'screateaninstanceoftheSpeech Synthesis Utterance class.We'llconfigurethisinstance withvarious properties.

The Speech Synthesis Interface of the Web Speech API is the controllerinterfacefor thespeechservice;thiscanbeusedtoretrieveinformationaboutthesynthesisvoicesavailableonthedevice,start andpause speech,andothercommandsbesides.

i. PROPERTIES

Speech Synthesis alsoinheritspropertiesfromitsparentinterface, Event Target.

SpeechSynthesis.paused Read-only

ABooleanvaluethatreturnstrue ifthe Speech Synthesis object isinapausedstate

SpeechSynthesis.pending Read-only

ABooleanvaluethatreturnstrue iftheutterancequeuecontainsas-yet-unspokenutterances.

SpeechSynthesis.speaking Read-only

ABooleanvaluethatreturnstrue ifanutteranceiscurrentlyintheprocessofbeingspoken evenif Speech Synthesis isinapausedstate.

ii. METHODS

Speech Synthesis alsoinheritsmethodsfromitsparentinterface,Event Target

SpeechSynthesis.cancel()

Removesallutterancesfromtheutterancequeue.

SpeechSynthesis.getvoices()

Returns a list of Speech Synthesis Voice objects representing all the available voices on thecurrentdevice.

SpeechSynthesis.pause()

PutstheSpeechSynthesisobject intoapausedstate.

SpeechSynthesis.resume()

PutstheSpeechSynthesisobjectinto anon-pausedstate:resumesitifitwasalreadypaused.

SpeechSynthesis.speak()

Addsanutterance totheutterancequeue;itwillbespokenwhen anyotherutterancesæ queuedbeforeithasbeenspoken.

B. SPEECH RECOGNITION

It is an interface to web speech API;it is used to convert speech to text

i. PROPERTIES

Speech Recognition alsoinheritspropertiesfromitsparentinterface,Event Target

Speech Recognition.Grammar

ReturnsandsetsacollectionofSpeech grammar objectsthatrepresentthegrammarsthat willbeunderstoodbythe currentSpeech Recognition

Speech Recognition. Lang

Returns and sets the language of the current Speech Recognition If not specified, thisdefaults to the HTML Lang attribute value, or the user agent's language setting if that isn'tseteither.

Speech Recognition. Continuous

Controlswhethercontinuousresultsarereturnedforeachrecognition,oronlyasingleresult.Defaultstosinglefalse.

Speech Recognition.intermResults

Controlswhether interimresultsshouldbereturned(true)ornot(false)Interimresultsareresultsthatarenotyetfinal(e.g.,the SpeechRecognitionResult.final propertyis false.

ii. METHODS

Speechrecognitionalsoinheritsmethodsfromitsparentinterface,Event Target

SpeechRecognition. Abort()

Stopsthespeechrecognitionservice fromlisteningtoincomingaudio,anddoesn'tattempttoreturnaSpeechRecognition. Result

SpeechRecognition. Start()

Startsthespeechrecognitionserviceby listeningtoincomingaudiowiththe intenttorecognizegrammarassociatedwiththe currentspeechrecognition.

SpeechRecognition. Stop()

Stops thespeech recognition servicefrom listening toincomingaudio,andattempts toreturnaspeechrecognition result using theaudiocaptured so far.

Speech grammar()

Createsanewspeech Grammar object.

SpeechGrammar.src

Sets and returns a string containing the grammar from within inthe speechgrammarobjectinstance.

C. TRANSLATOR

QuerySelector()

It is used to add the CSS property to the HTML element (or) query selector method returns the first element that matches a CSS property

Google Translate API

Google Translate is a free multilingual machine translation service. It can translate the Website's text content from one language to another. It offers a huge list of languages to translate and has an efficient, reliable, and easy way to translate the webpage into whatever language the user wants. It supports over 100 languages. Use this website translator to convert webpages into your choice of language

VI. RESULTS AND DISCUSSION

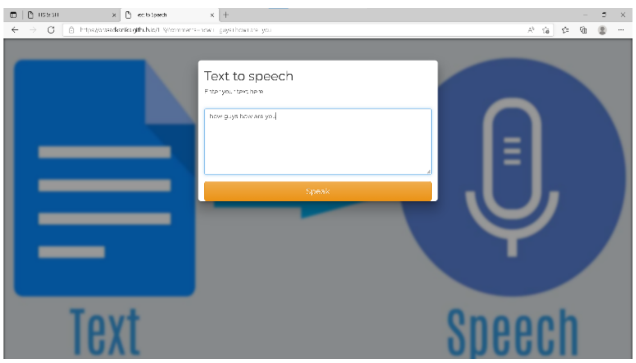

Fig 5:*text to speech output*

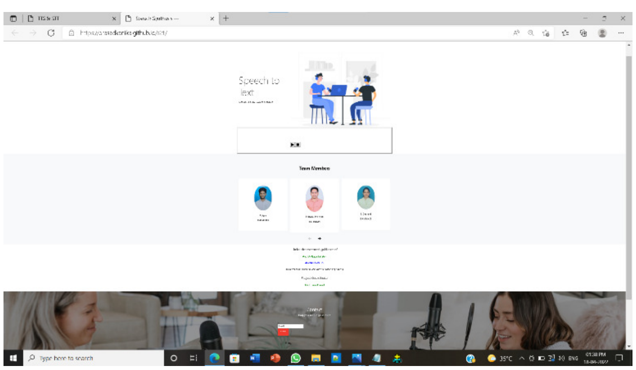

Fig 6: *Speech to text output*

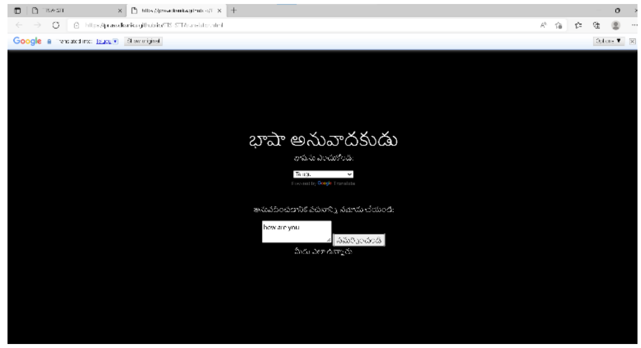

Fig 7: *Translator output*

Themainmottoof TTSis toconvertarbitrary textintowaveform.Speechgeneration is the generation of an acoustic waveform corresponding to text and each of theseunits in the sequence. This involved text analysis, text normalization, text processing,grapheme-to-phonemeconversion,andspeechsynthesis.

Text analysis which analyses the input text such as dividing the text into words andsentences.Textnormalizationisthetransformationoftextintothepronounceableform,it is the front end of TTS that assigns phonetic transcription to each and every word. Theprocess of assigningphonetictranscription toa wordis calledgraphemetophonemeconversion.

The phonetic analysis also known as word analysis focuses on the phone within the word.finallysymboliclinguisticrepresentationproducessound.

The general objective of the project is to develop a digital audiobookfor the students.The specificobjectives are:

Itenablestohelppeoplewhostrugglewithreading.

Itenablestoconvertthespeechto text

Itenables you inlearninglanguagesthatyoudonotknow.

Itenablestohelp thestudentsinlistening to ebooksore-materialduring the examination.

VII CONCLUSION

Speechsynthesishasbeendevelopedsteadily overthelastdecadesandithasbeenincorporatedintoseveral newapplications.Formostapplications,theintelligibility andcomprehensibilityofsyntheticspeechhavereachedanacceptablelevel.However,inprosodic,textpreprocessing,andpronunciationfieldsthereisstillmuch workandimprovements to be done to achieve more natural-sounding speech. Natural speech has somany dynamic changes thatperfectnaturalnessmay beimpossible toachieve.However,since the markets for speech synthesis-related applications are increasing steadily, the interestin giving more effort and funds to this research area is also increasing. Present speechsynthesis systems are so complicated that one researcher cannot handle the entire system.With good modularity, it is possible to divide the system into several individual moduleswhose developing process can be done separately if the communication between the modulesismadecarefully.

REFERENCES

[1] Fernando Alonso Martin , Maria Malfaz, Alvaro Castro-Gonzalez (2020) “ Four-Features Evaluation of Text to Speech System for Three Social Robots ”,Volume 9 ,Issue 2

[2] JakeeIndapanya,VorapratRattana,Sangsuree, SurapongChatpun, KometThongkhao, Kanthika kraisawat(2020)“Blue compass ”

[3] EmreKaragoz ,Kutankoruyan(2020)“DesignofAudioDescriptionSystemUsingCloudBasedComputerVision”,Volume 4,Issue1,74-85

[4] ThreeSocialRobots”,Volume11876
Sneha.C,S.B.Gundre(2018) “OCRBasedImageTexttoSpeechConversion UsingMATLAB ”

[5] HRithika,BNithyaSanthoshi(2017)“ Imagetextto speechconversioninthedesiredlanguagebytranslationwithRaspberryPi”,Volume9,Issue 2

[6] Y. Yasuda, et al., “Investigation of enhanced Tacotron text-to-speechsynthesis systems with self-attention for pitch accent language,” in Proc , pp . 2019.

[7] Jing-Xuan Zhang, Zhen-Hua Ling, Yuan Jiang, Li- Juan Liu, Chen Liang, and Li-Rong Dai, “Improving sequence-to-sequence voice conversion by adding textsupervision,” in ICASSP 2019. IEEE, 2019, pp. 6785–6789.