International Journal of Research Publication and Reviews

Journal homepage: www.ijrpr.com ISSN 2582-7421

Text To Speech Reading By System

[1]Rahul Kulkarni, [2]Sneha, [3]Triveni, [4]Vaishnavi, [5]Omkar

Assistant Professor, Guru Nanak Dev Engineering College Bidar Students Dept of CSE GNDEC, Bidar Karnataka 585401 India

ABSTRACT

This software project reads an SMS file to the user using Windows-based software. This reads an SMS file from its temporary database along with any accompanying pronunciations. The user is then given a full word to read. The rest of the words, including the complicated ones, are calculated and read in accordance with the articles' and fundamental terms' pronunciations. In order to save the user from always having to stare at the computer and read the entire content, the bot can be used to help read text documents for the user. A recent software product called test to speech converter enables even visually impaired people to read and comprehend various documents. A text-to-speech system's goal is to turn any given text into the corresponding spoken waveform. A text to speech system's two major parts are text processing and speech production. The text processing component's goal is to process the provided input text and generate the proper order of phonemic units. The voice generation component creates these phonemic units either by synthesis from parameters or by choosing a unit from a large speech corpus. The text processing component must create an appropriate sequence of phonemic units corresponding to an arbitrary input text in order for the speech synthesis to sound natural. It is crucial that the text processing component generate the right sequence of phonemic units corresponding to an arbitrary input text in order to construct a voice synthesis system that produces natural-sounding speech.

Keywords: Text, speech, data, system

1. INTRODUCTION

Modern research and applications in speech communication rely heavily on digital speech processing. Speech is mostly used for communication, which refers to message transfer between humans and machines. Using a speech synthesizer, text is transformed into voice by a text-to-speech system (TTS). It is the creation of human speech artificially. A speech synthesizer is a type of computer system used for this. A text-to-speech system's goal is to turn any given text into the corresponding spoken waveform. A text to speech system's two major parts are text processing and speech production. The text processing component's goal is to process the provided input text and generate the proper order of phonemic units. The voice generation component creates these phonemic units either by synthesis from parameters or by choosing a unit from a large speech corpus. The text processing component must create an appropriate sequence of phonemic units corresponding to an arbitrary input text in order for the speech synthesis to sound natural. Why we don't have TTS systems for all or many of the 23 official Indian languages is one of the frequent queries from end users. Which complexities are present: Is this because there aren't any voice databases available for Indian languages or because the synthesis technology isn't yet developed enough to allow for building for any language? Unfortunately, because of unit selection procedures, the primary speech production technology, which involves producing speech from a phonemic sequence, has been mostly automated for the past ten years? The development of statistical parametric speech synthesis techniques has made it much simpler to create a voice in a language with a smaller speech corpus and fewer phrases. It can be challenging to persuade a user that the text input for a TTS system is not a phonemic sequence but rather the raw text found in news websites, blogs, documents, etc., which contains the necessary text in font-encodings, native scripts, and non-standard words like addresses, numbers, currency, etc. The handling of real-world text is where most of the problems in developing a TTS for a new language are found.

2. SYSTEM ANALYSIS

The goal of text-to-speech (TTS) synthesis is to create understandable, realistic-sounding speech from any arbitrary input text. Digital signal processing and natural language processing are the two major components of the TTS system. Figure 1 displays the TTS system's general block diagram. There are three processes in the processing of natural language. These three types of analysis are prosodic, phonetic, and textual. Segmentation, text normalization, and part of speech (POS) tagger are all components of the text analysis. Each word is given a phonetic transcription using phonetic conversion. In phonetic there are two methods.

Both of them rely on dictionaries and rules. If a word is unknown, rule-based analysis is employed; if a term is recognized, dictionary-based analysis is applied. Prosodic analysis is used to represent speech's intonation, loudness, and duration. It expresses the speaker's feeling

Existing System

The current technology does not enable text to speech conversion, and even if it did, it lacked effective tools for accurately reading all of the content. People with visual impairments will have a lot of difficulties reading the text file that has been emailed to them. The problem faced by those who cannot read can be greatly reduced by developing an application that converts text to speech.

Disadvantages

The words are not appropriately spoken by the current method.

The current system has many issues, like delayed processing and becoming stuck in between words, among others. There isn't a lot of software available to read text files for those who are blind.

Proposed System

Making an application speak to the user wasn't even a possibility not too long ago; it was just a pipe dream. Utilizing the sound system found in every modern PC, apps can now speak to you as easily as food thanks to modern technology and the power it provides. It is really simple to use Text to Speech (or TTS, as it is more often known) in.NET, and you may enable all the functionality you require by simply adding one reference to your application. In this paper, we'll walk through the development of a straightforward Windows program that lets you write in some text, specify some conditions for that text, and then hear that phrase said back to you..

Objective and Scope

A text-to-speech system's goal is to turn any given text into the equivalent spoken waveform. A text to speech system's two major parts are text processing and speech production. The text processing component's goal is to process the provided input text and generate the proper order of phonemic units. The voice generation component realizes these phonemic units either by synthesis from parameters or by choosing a unit from a sizable speech corpus [4]. The text processing component must create an appropriate sequence of phonemic units corresponding to an arbitrary input text in order for the speech synthesis to sound natural.

Problem Statement

This project identified the problem of persons having visually challenged and those who have problem of eye strain.

3. LITERATURE SURVEY

[1]. Voice quality variations consist of a variety of voicing sound source adjustments, from normal to breathy phonation to laryngealized phonation. There are several potential acoustic cues to this type of voice quality variation, according to analysis of repeated imitations of two sentences by ten female and six male talkers. These cues include: (1) increases in the relative amplitude of the fundamental frequency component as open quotient increases; (2) increases in the amount of aspiration noise that replaces higher frequency harmonics as the arytenoids become more separated; (3) increases in lower formant bandwidths; and (4) introductio Using a novel voicing source model for the synthesis of more realistic male and female voices, it has been possible to achieve perceptual confirmation of the relative importance of these cues for signaling a breathy voice quality.

[2]. This study used the modified rhyme test (MRT) to assess the segmental comprehensibility of synthetic speech produced automatically by rules, and the findings are presented in this paper. Ten text-to-speech systems' created artificial speech was examined and contrasted with real speech. To investigate the impact of response set size on perceptual confusions, a modified version of the standard MRT was employed. The segmental intelligibility scores, according to the results, formed a continuum. A few systems demonstrated extremely high levels of performance that were on par with or somewhat better than scores achieved using natural speech, while other systems demonstrated noticeably subpar performance. The best system, DECtalk—Paul, performed the best overall, matching the data acquired from genuine speech for consonants in syllable start position. The results of this study are presented in terms of the usage of a collection of standardized techniques for evaluating the understandability of synthetic speech in a lab setting. Recent research has looked at how synthetic speech is perceived in more trying circumstances where the listener's cognitive power is put under additional stress.

[3]. Currently, a number of laboratory systems and commercial devices accomplish the automatic conversion of English text to synthetic speech quite well. Advancements in this field have Improvements in linguistic theory, acoustic-phonetic characterization of English sound patterns, perceptual psychology, mathematical modeling of speech production, structured programming, and computer hardware design have all contributed to this development. The early work on the creation of voice synthesizers, the identification of minimal acoustic cues for phonetic contrasts, the development of phonemic rule programs, the addition of prosodic rules, and the design of text analysis techniques are all covered in this overview. The state of the art is illustrated with numerous examples of regulations. Many of the examples are from Khattak, the author's text-to-speech program. There are a lot of scientific issues that are brought up that prohibit the existing systems from producing entirely human-sounding speech. Although rule programs that operate a form-ant synthesizer are the main focus, alternatives like articulatory synthesis and waveform concatenation are also discussed. An large bibliography has been put together to demonstrate the range of synthesis activity

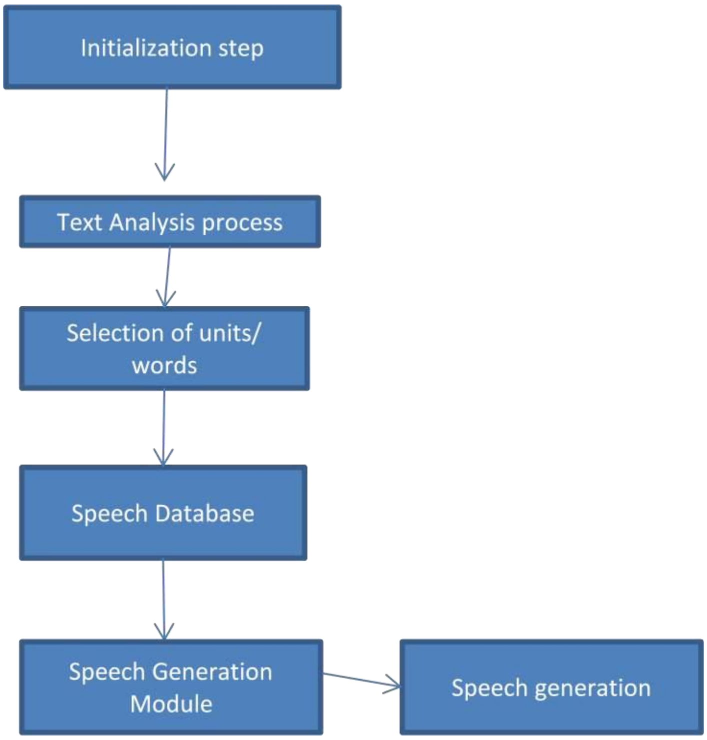
and the wealth of phenomena that the finest of these programs' rules can encompass.

4. SYSTEM DESIGN

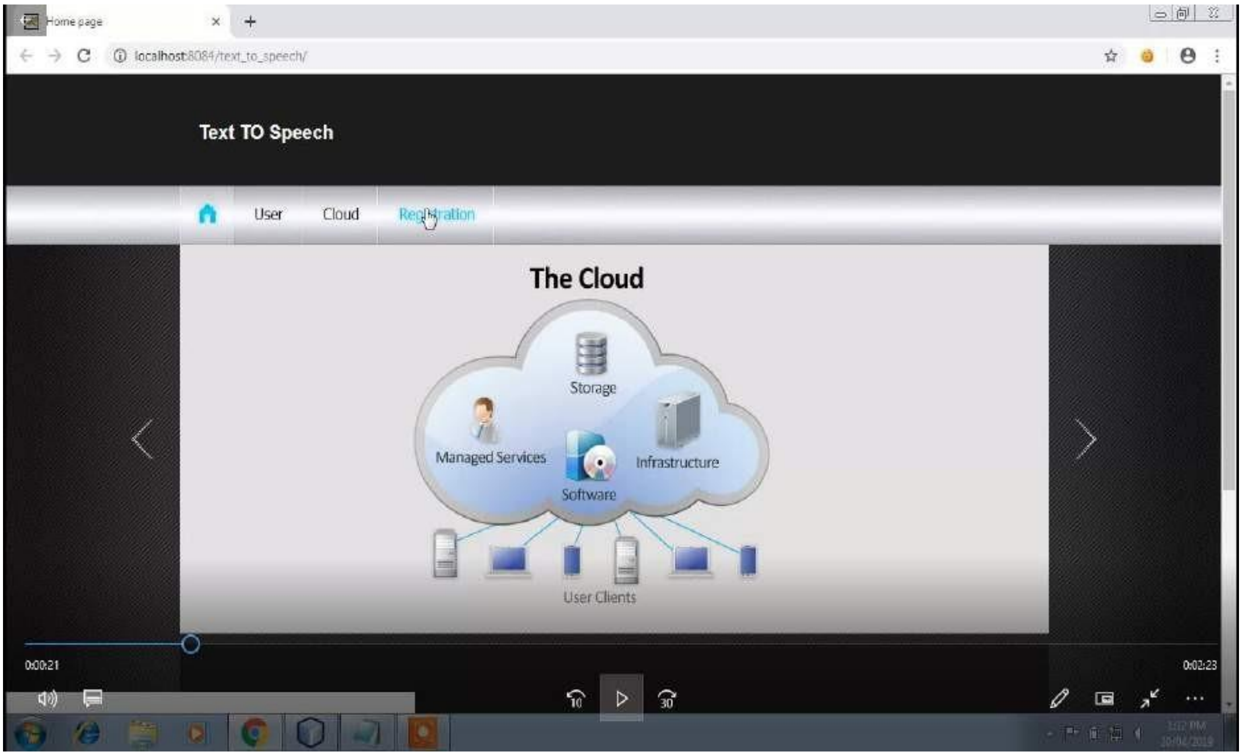
The desired system's overall architecture is chosen at the following development stage known as System Design. The system is set up as a collection of connected subsystems. The analyst considers both the specifications found in system analysis and what the end user expects from the new system as they design the system as a collection of interdependent subsystems.

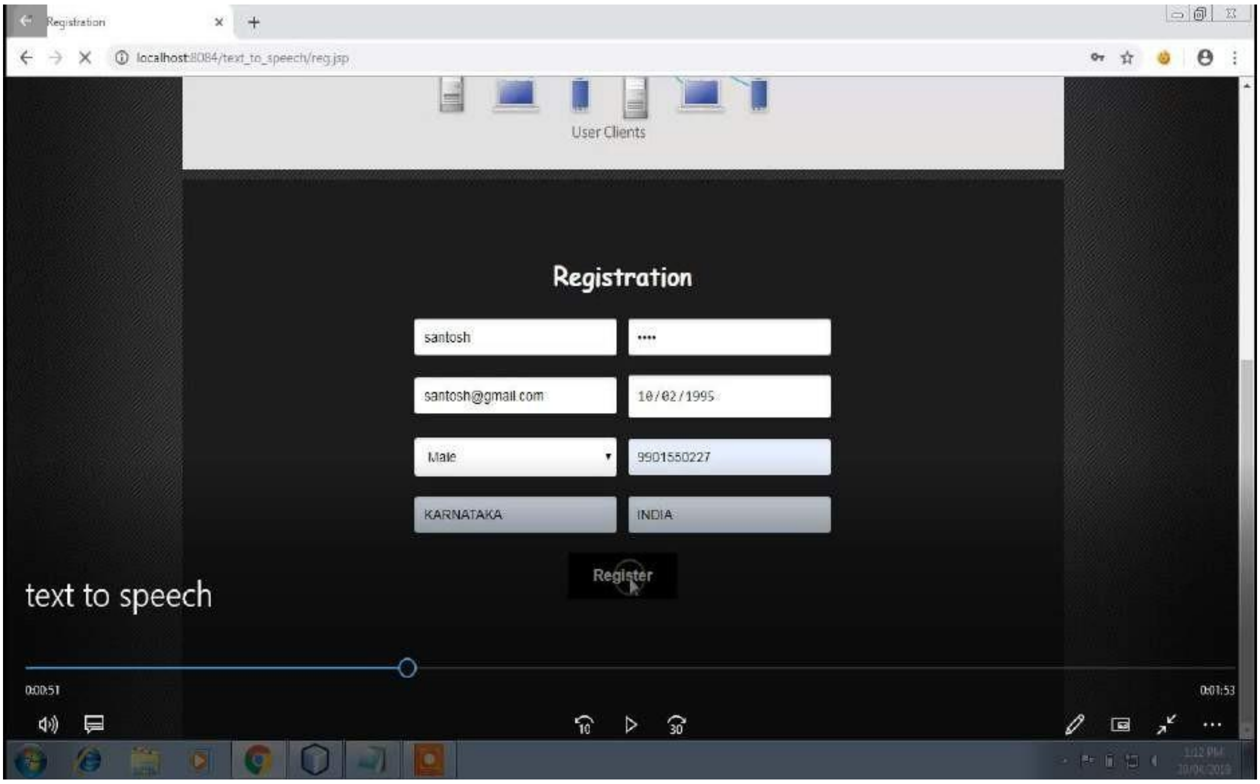
Since the fundamental tenet of the object-oriented approach to system analysis is to view the system as a collection of interdependent objects, a larger system may also be viewed as a collection of interdependent smaller subsystems, each of which is made up of a collection of interdependent objects. As opposed to the traditional Waterfall Model, where the processes are the most crucial component of the system, the emphasis while creating the system is on the objects that make up the system.

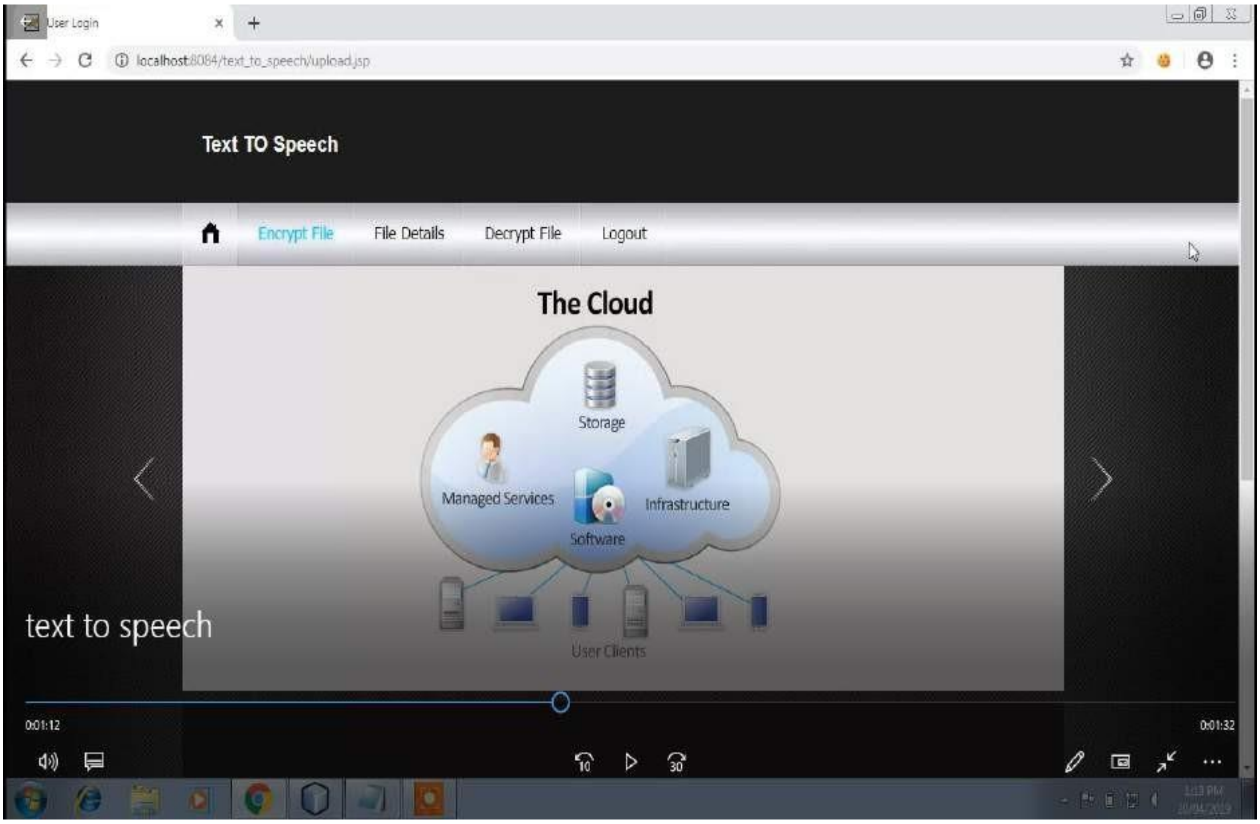
Data Flow Diagram

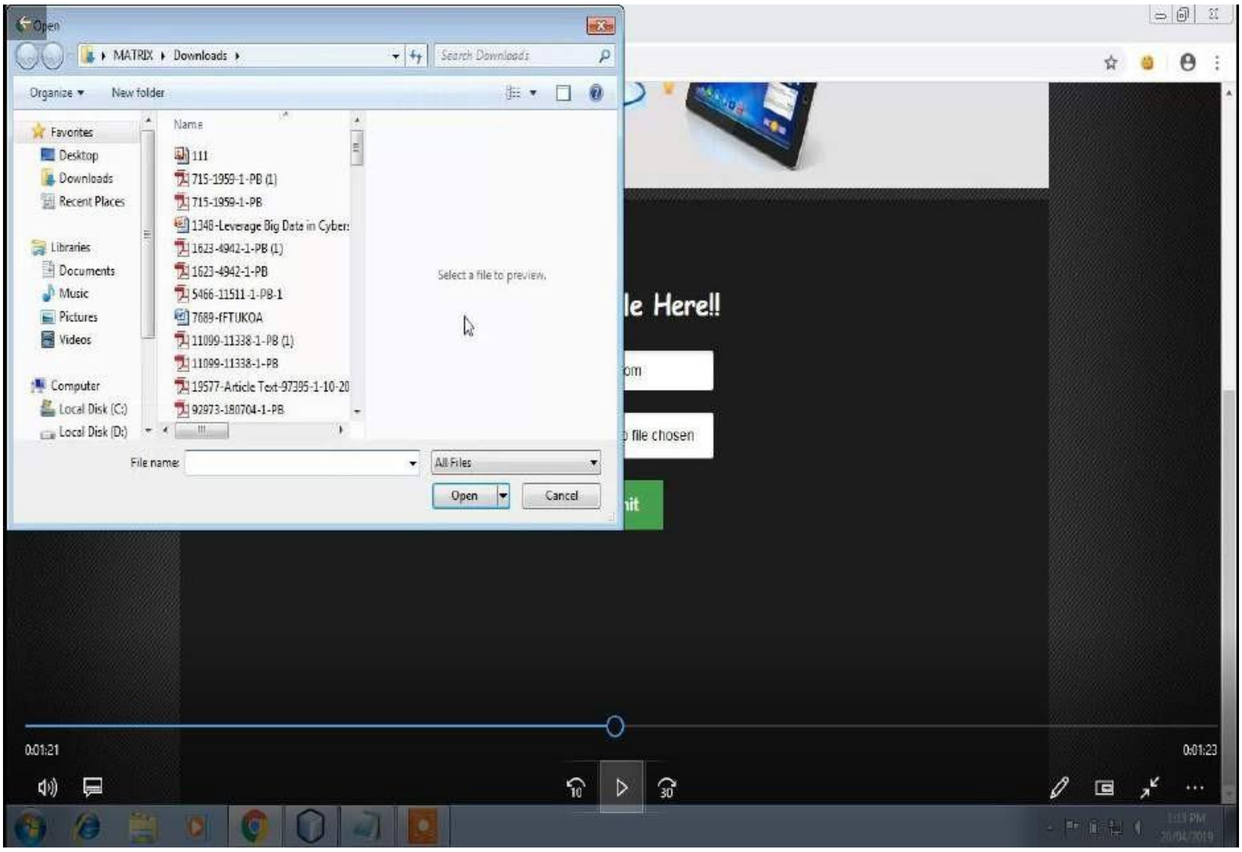


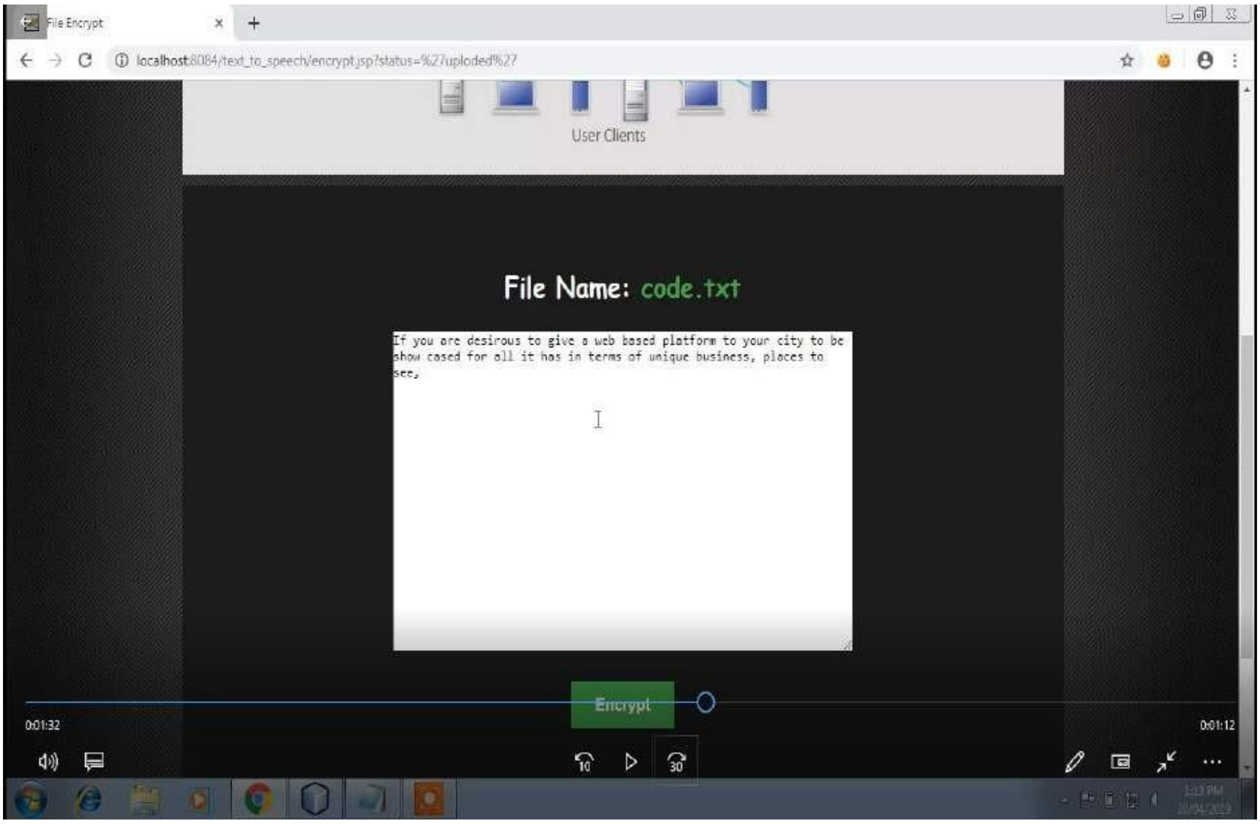
5. OUTPUT SCREEN

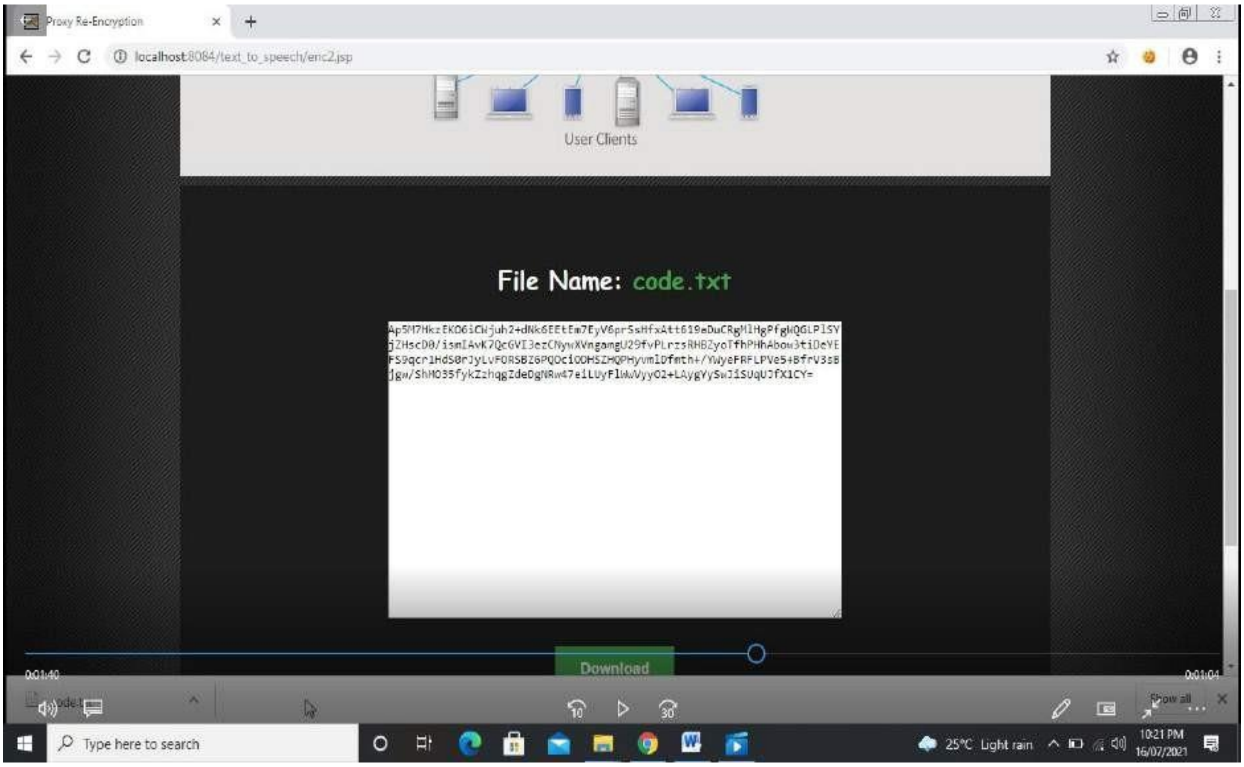

Home page
localhost:8084/text_to_speech/
Text TO Speech
User
Cloud
Registration
The Cloud
Storage
Managed Services
Infrastructure
Software
User Clients
0:00:21
0:02:23

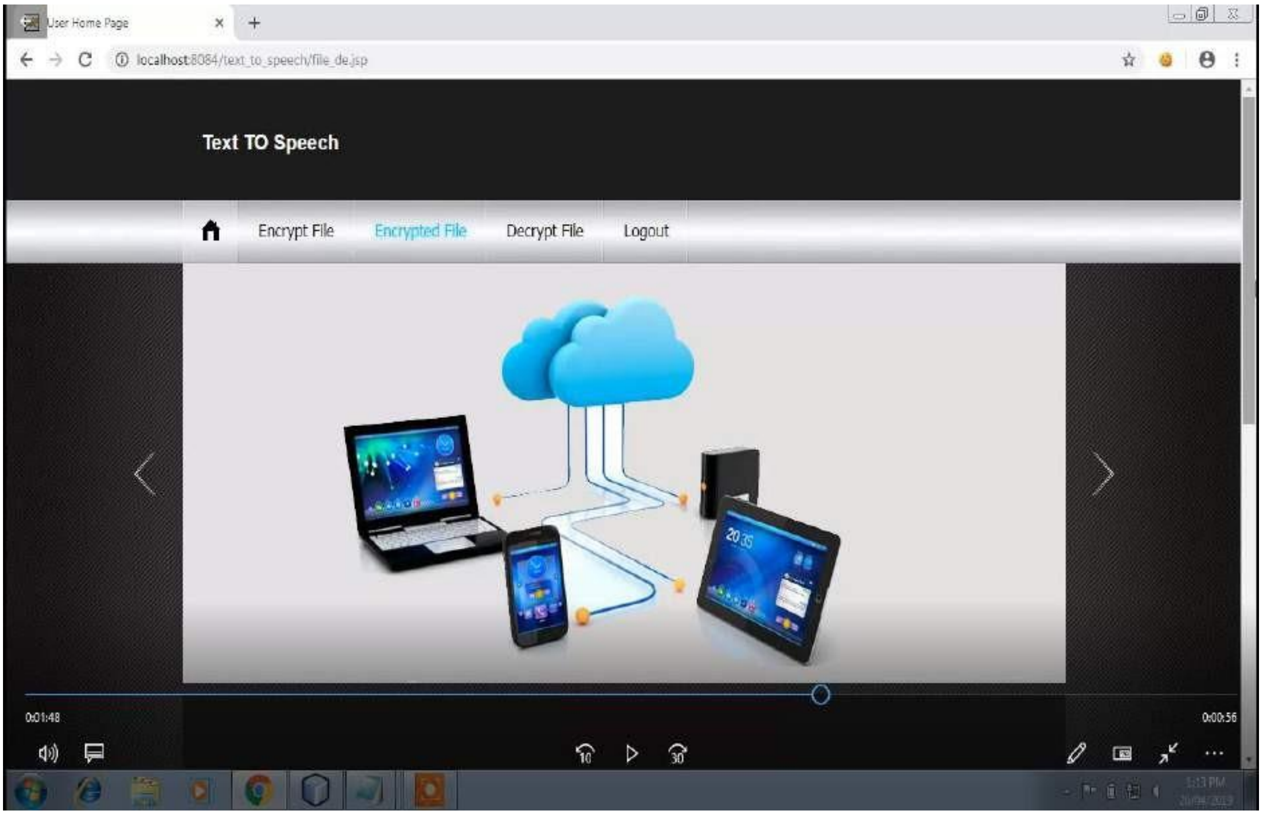

Registration
localhost:8084/text_to_speech/reg.jsp
User Clients
Registration
santosh
santosh@gmail.com
10/02/1995
Male
9901550227
KARNATAKA
INDIA
Register
text to speech
0:00:51
0:01:53

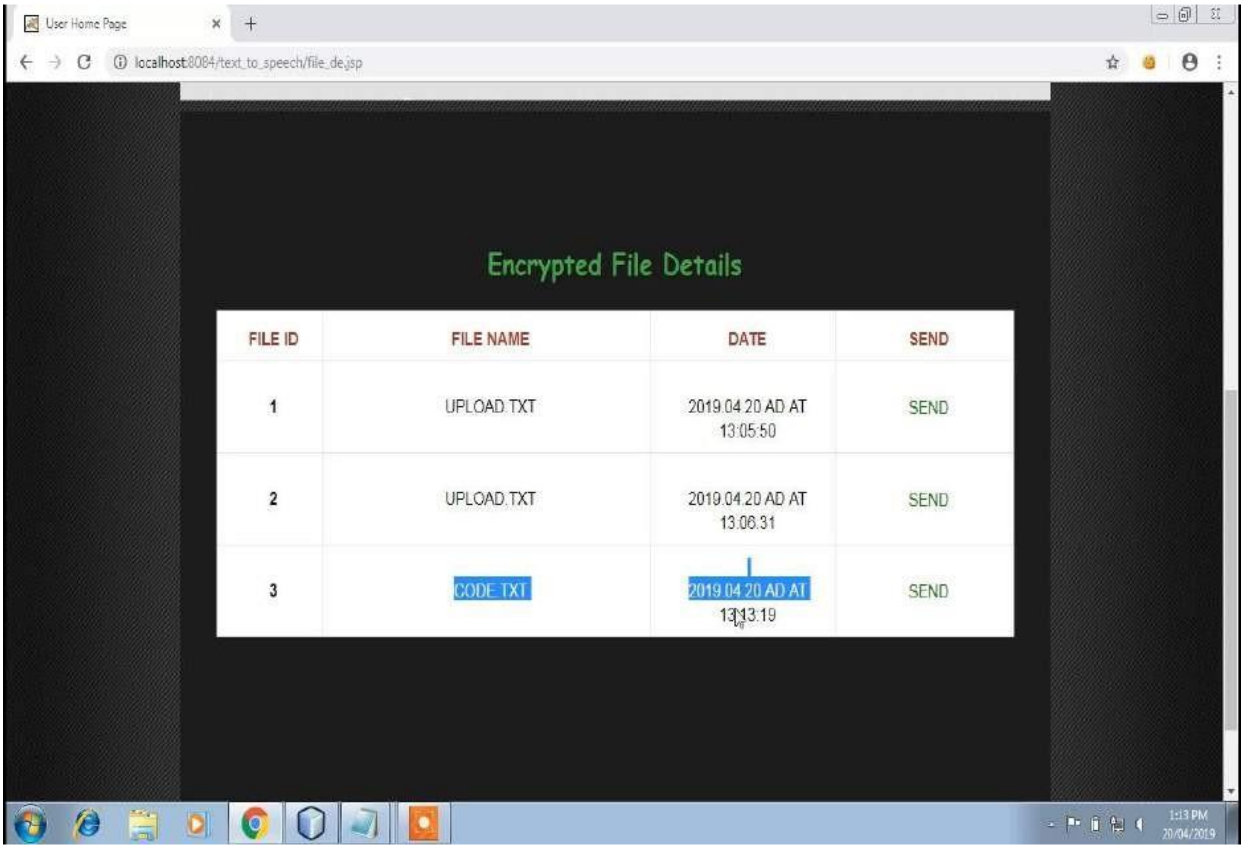

User Login
localhost:8084/text_to_speech/upload.jsp
Text TO Speech
Encrypt File
File Details
Decrypt File
Logout
The Cloud
Storage
Managed Services
Infrastructure
Software
User Clients
text to speech
0:01:12
0:01:32

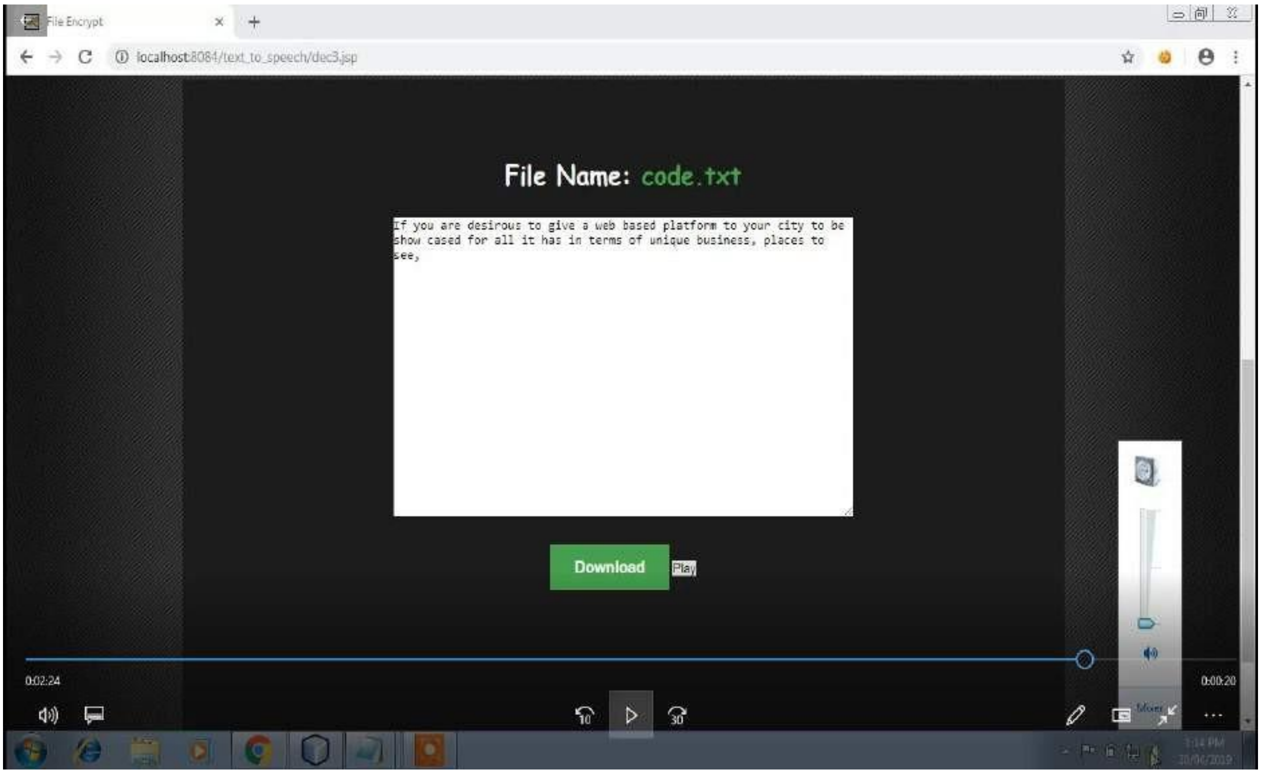

Open
MATRIX
Downloads
Search Downloads
Organize
New folder
Favorites
Desktop
Downloads
Recent Places
Libraries
Documents
Music
Pictures
Videos
Computer
Local Disk (C:)
Local Disk (D:)
Name
111
715-1959-1-PB (1)
715-1959-1-PB
1623-4942-1-PB (1)
1623-4942-1-PB
5466-11511-1-PB-1
7689-fFTUKOA
11099-11338-1-PB (1)
11099-11338-1-PB
92973-180704-1-PB
Select a file to preview.
File name:
All Files
Open
Cancel
le Here!!
file chosen
0:01:21
0:01:23


File Encrypt
localhost:8084/text_to_speech/encrypt.jsp?status=%27uploded%27
User Clients
File Name: code.txt
If you are desirous to give a web based platform to your city to be show cased for all it has in terms of unique business, places to see,
Encrypt


Proxy Re-Encryption
localhost:8084/text_to_speech/enc2.jsp
User Clients
File Name: code.txt
Download
Type here to search


User Home Page
localhost:8084/text_to_speech/file_de.jsp
Text TO Speech
Encrypt File
Encrypted File
Decrypt File
Logout


User Home Page
localhost:8084/text_to_speech/file_de.jsp
Encrypted File Details
FILE ID
FILE NAME
DATE
SEND
1
UPLOAD.TXT
2019.04.20 AD AT 13:05:50
SEND
2
UPLOAD.TXT
2019.04.20 AD AT 13:06:31
SEND
3
CODE.TXT
2019.04.20 AD AT 13:13:19
SEND



ADVANTAGES

- People who are visually impaired or have learning impairments can benefit from the system.
- Prevents eye strain, allowing the user to sit comfortably and listen.
- Saves time, particularly when driving or working out.
- Simple to use.
- Assistance with reading, writing, and spelling skills.

APPLICATION

This application can be very useful for blind persons who are unable to see the messages, and it can also be useful for those who are unable to read the messages.

- One can utilize the program while driving if they can't see the screen to read it.
- The program can also be used for training purposes, allowing for the preparation and placement of slides for potential trainees.

MODULES

Natural Language Processing (Nlp) Module

It creates a prosodic and phonetic transcription of the material read. Module for digital signal processing (DSP): It converts the NLP-provided symbolic data into audible and understandable speech. The following are the main functions of the NLP module

Text Analysis

To begin, the text is divided into tokens. The token's orthographic form is produced by the token-to-word conversion. The token "Mr" is expanded to the orthographic form "Mister," given the number "12," and the year "1997" is changed to "nineteen ninety-seven." .

Application of Pronunciation Rules

Pronunciation rules might be used after the text analysis is finished. Because correspondence is not always parallel, letters cannot be converted into phonemes 1:1. In some contexts, a single letter can represent either no phoneme (like the letter "h" in the word "catch") or many phonemes (like the letter "m" in the word "Maximum").

CONCLUSION

This paper develops a text-to-speech system for words, sentences, and numbers. The output speech for numbers (one or more digits) is natural and enjoyable to hear. When speech waveforms are concatenated, the speech delay must be eliminated. Consequently, domain-specific synthesis can effectively create speech. However, the output speech of words consists of gaps between phoneme transitions. Syllabification is preferable for words with two or more syllables. The audio quality is understandable. Thus, compared to other methods that require the use of numerous complex algorithms, domain-specific and phoneme-based synthesis is very simple to execute and quite effective. However, unit selection synthesis is more difficult to apply than these other two techniques. There are few errors in the speech output sentence. The speech output, however, outperforms phoneme-based synthesis.

REFERENCES

[1]. D.Sasirekhaand E.Chandra, ―Text to speech: A simple tutorial‖, ISSN: 2231-2307, Volume-2, Issue-1, March 2012.

[2]. Prof. Dr. Hilal M. Yousif, Dr. Mouyad A. Fadhiland Yahya M. Hadi, ―Text-to-Speech Synthesis State-Of- Art‖, 2004.

[3]. Sumeer Mittal , MrNavdeep Singh Sethi and Sanjeev KumarSharma,―Partof Speech Tagging of Punjabi Language using N Gram Model‖,International Journal of Computer Applications (0975 – 8887) ,Volume 100– No.19, August 2014.

[4]. Ing. Milan Legat, ―Concatenation Cost in Unit Selection Speech synthesis‖,Faculty of Applied Science, University of West Bohemia in Pilsen ,2012.

[5]. N.Swetha and K. Anuradha, ―Text-to-speech conversion‖ ,International Journal of Advanced Trends in Computer Science and Engineering, Vol.2, No.6, Pages: 269-278(2013)