Text Embedded Image to Speech Conversion

COMMUNITY SERVICE PROJECT

Submitted by

NELAVALLI BASAVA ADARSH (99210041251)
NELAKURTHI SUBHASH (99210041796)
NARALA VINAY KUMAR REDDY (9921004936)
NAKKA MOHITH(9921004485)

in partial fulfillment for the award of the degree of

BACHELOR OF TECHNOLOGY
in

COMPUTER SCIENCE AND ENGINEERING

SCHOOL OF COMPUTING COMPUTER

SCIENCE AND ENGINEERING

KALASALINGAM ACADEMY OF RESEARCH

AND EDUCATION

KRISHNANKOIL 626 126

Academic Year 2023-2024

DECLARATION

We affirm that the project work titled **"Text Embedded Image to Speech Conversion"** being submitted in partial fulfillment for the award of the degree of **Bachelor of Technology in Computer Science and Engineering** is the original work carried out by us. It has not formed the part of any other project work submitted for award of any degree or diploma, either in this or any other University.

NELAVALLI BASAVA ADARSH

(99210041251)

NELAKURTHI SUBHASH

(99210041796)

NARALA VINAY KUMAR REDDY

(9921004936)

NAKKA MOHITH

(9921004485)

KALASALINGAM
ACADEMY OF RESEARCH & EDUCATION
(DEEMED TO BE UNIVERSITY)
Under sec. 3 of UGC Act 1956. Accredited by NAAC with "A" Grade

BONAFIDE CERTIFICATE

Certified that this project report **"Text Embedded Image to Speech conversion"** is the bonafide work of "**NELAVALLI BASAVA ADARSH, NELAKURTHI SUBHASH, NARALA VINAY KUMAR REDDY, NAKKA MOHITH"** who carried out the project work under my supervision.

SUPERVISOR

Mrs. S. RESHNI
Assistant Professor
Computer Science and Engineering
Kalasalingam Academy of Research and Education
Krishnankoil 626126

HEAD OF THE DEPARTMENT

Dr. N. Suresh Kumar
Professor & Head
Computer Science and Engineering
Kalasalingam Academy of Research and Education
Krishnankoil 626126

Submitted for the Project Viva-voce examination held on......................................

Supervisor **Faculty Advisor** **External Examiner (s)**

ACKNOWLEDGEMENT

First and foremost, I wish to thank the **Almighty God** for his grace and benediction to complete this Project work successfully. I would like to convey my special thanks from the bottom of my heart to my dear **Parents** and affectionate **Family members** for their honest support for the completion of this Project work.

I express deep sense of gratitude to “Kalvivallal” Thiru. **T. Kalasalingam** B.com., Founder Chairman, “Ilayavallal” **Dr. K. Sridharan,** Ph.D., Chancellor, **Dr. S. Shasi Anand**, Ph.D., Vice President (Academic), **Mr. S. Arjun Kalasalingam** M.S., Vice President (Administration), **Dr. S. Narayanan**, ViceChancellor, **Dr. V. Vasudevan,** Ph.D., Registrar, **Dr. P. Deepalakshmi,** Ph.D., Dean (School of Computing), **Dr. N. Suresh Kumar**, Professor & Head, Department of CSE, Kalasalingam Academy of Research and Education for granting the permission and providing necessary facilities to carry out Project work.

I would like to express my special appreciation and profound thanks to my enthusiastic Project Supervisor **Mrs. S. RESHNI**, Assistant Professor/CSE of Kalasalingam Academy of Research and Education (KARE) for his inspiring guidance, constant encouragement with my work during all stages. I am extremely glad that I had a chance to do my Project under my Guide, who truly practices and appreciates deep thinking.

I will be forever indebted to my Faculty Advisor ‘**Mrs.P.Kalaiarasi’** for all the time. She gave me the moral support and the freedom I needed to move on.

Besides my Project guide, I would like to thank the committee members, all faculty members and non-teaching staff for their insightful comments and encouragement. Finally, but by no means least, thanks go to all my school and college teachers, well wishers, friends for almost unbelievable support.

KALASALINGAM
ACADEMY OF RESEARCH & EDUCATION
(DEEMED TO BE UNIVERSITY)
Under sec. 3 of UGC Act 1956. Accredited by NAAC with "A" Grade

School of Computing

Department of Computer Science and Engineering

Project Summary

Project Title	Text Embedded Image To Speech
Project Team Members (Name with Register No)	1. NELAVALLI BASAVA ADARSH (99210041251) 2. NELAKURTHI SUBHASH (99210041796) 3. NARALA VINAY KUMAR REDDY (9921004936) 4. NAKKA MOHITH (9921004485)
Guide Name/Designation	Mrs. S. RESHNI, Associate Professor, Department of Computer Science and Engineering
Program Concentration Area	To Serve Blind People
Technical Requirements	Easyocr module

Engineering standards and realistic constraints in these areas: (Refer Appendix in page 4 of this doc.)

Area	**Codes & Standards / Realistic Constraints**	**Tick** ✓
Economic		
Environmental		
Social	This project is mainly used for recognising text present in image and it is helpful to blind people.	
Ethical		✓
Health and Safety	To promote the user well being and accessibility, speech quality.	

Manufacturability		
Sustainability		

Realistic Constraints:

Ethics:

In our neighborhood we find many people who leave their pets in to the streets because of reasons like behaviour of their pet, health and hygene, kids at home, lost love for their pets or health of the pet. They abandon their pets for any of the above reasons. These pets then become stray dogs and street animals. It also becomes difficult for the pets that were fed at homes with love to live in the streets without any care and affection that they had before. Some pets may be harmed by street dogs. These pets may not find food. These pets also become stray dogs and street animals.

On the other hand some people in the same society love to have pets. They spend thousands and even lakhs of amounts on pet shops to buy a pet. Our intention is to provide a platform where pet lovers can find their pets for free from those who are willing to donate or ready to abandon their pets. This prevents the pets from being homeless. We provide pets a new home with love and affection from their owners.

ABSTRACT

Visual impairment is one of humanity's most significant limitations, especially in the modern era, when text messages—both electronic and printed—are frequently used to convey information instead of spoken word. Our recommended device is intended to help people who are blind or visually impaired. We developed a tool for this project that speaks text that is contained in images. The core architecture consists of an embedded system that reads an image, recognises the region of interest that contains the text, and speaks the text alone into voice. They are optical character recognition (OCR) software and text-to-speech (TTS) engines. This study article's main objective is to investigate the most recent cutting-edge methods used in text-to-image to image-to-speech conversion and to emphasise.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF ABBREVIATIONS

Abbreviation	Full form
OpenCV	Open Source Computer Vision
OCR	Optical Character Recognition
TTS	Text-to-speech
CSS	Cascading Style Sheets
HTML	Hyper Text Markup Language

CHAPTER-1

INTRODUCTION

Worldwide around 285 million persons are visually impaired and there are nearly 39 million blind people. This has a significant effect on the lives of those who are visually impaired. While numerous attempts have been made to assist visually impaired people in seeing objects by alternative means like sound and touch, the development of textreading devices is still in its early stages. A technology known as "Image to Speech Conversion" makes it possible to transform images into audible speech in the English language. There are a lot of uses for this technology, especially in the accessibility space. This technology can be very helpful to those who are visually impaired since it enables them to understand the information contained in images that they would not otherwise have access to. Thanks to developments in artificial intelligence and deep learning, image-to-speech conversion has changed Frequently in recent years. Text in a variety of fonts, sizes, and styles can be recognized by this technology and turned into voice. Processing textual data from various image formats, including printed documents, web pages, and social media posts, is one area in which the technology excels. Ensuring that everyone, including those who are visually impaired, has equal access to information is the primary goal of image-to-speech conversion.

The output is fed to an output device depending on the user’s choice. Output can be heard through headphones connected to audio jack of raspberry pi.

1.1 Aim and Objective

The aim of the project was to convert an image to speech. An image is processed and segmented to identify the text in the image. Then the characters are combined to form words and save it as a text file. This text file is converted to speech. We use two tools for the completion of image to text to speech conversion.

- The code is to convert the image to speech.
- An image is processed and segmented to identify the characters in the image. Then the characters are combined to form words and save it as a text file.
- We have divided the project into four sub parts : image is pre-processed, segmented to extract the images of characters.
- Then characters are recognized and combined , then the text is translated then converted into speech.

1.2 Problem Statement

The blind people and illiterates are facing difficulty in understanding the content they have. Now a days this leads to manipulations and scams.

CHAPTER-2

LITERATURE REVIEW

As we see in recent years have seen a considerable increase in interest in this technology because of its potential to enhance the lives of those who are visually impaired. We will look at some of the most recent findings in the field of image-to-speech conversion in this review of the literature. Many methods, such as object recognition-based, deep learning-based, and OCR-based approaches, have been proposed for converting images to speech. OCR-based methods, which use optical character recognition (OCR) algorithms to extract text from images, are among the most popular approaches. However this method has limitations: it works only with text-based images, and it might not be able to identify complex images or handwritten text. Computer vision algorithms are used in object recognition-based techniques algorithms to detect and recognize objects in images. This approach has shown promising results. There has been a growing interest in developing mobile applications for image to speech conversion, which would enable visually impaired individuals to use the technology on their smartphones. Several such applications have been developed, including Aipoly, Envision AI, and Be My Eyes. These applications use a combination of OCR-based, object recognitionbased, and deep learning-based approaches to convert images to speech and provide other features such as object recognition and text-to-speech conversion.One of the most widely used approaches for text embedded in image to speech conversion is OCR (Optical Character Recognition). OCR algorithms are capable of detecting and recognizing text within an image and converting it into a digital text format. The digital text can then be converted into speech using text-to-speech (TTS) technology. However, OCR-based approaches have limitations, particularly when it comes to recognizing handwritten text or text embedded in complex images. However, it may not be effective in recognizing text or other forms ofinformation.Deep learning-based approaches, particularly convolutional neural networks (CNNs) and recurrent neural networks (RNNs), have shown great potential in image to speech conversion. These approaches involve training a neural network on a large dataset of images and their corresponding audio output. The network learns to recognize patterns in images and generate corresponding speech output. Several studies have shown that deep learning-based approaches can achieve high accuracy in image to speech conversion

CHAPTER-3

METHODOLOGY

The proposed method will involve a combination of OCR-based, object recognition-based, and deep learning-based approaches to achieve high accuracy in text recognition and speech output. he suggested methodology for text-embedded image-to-speech conversion takes a methodical approach to the task of converting text-embedded images into audio descriptions that sound appropriate and natural. Data collection and preprocessing are the first steps in ensuring that a diverse dataset of images with corresponding text is cleaned and prepared for analysis. After that, text recognition models are used to extract and transcribe the embedded text while simultaneously developing a language model to understand its semantics. Speech synthesis techniques generate high-quality spoken content; a fusion mechanism is presented to synchronize the synthesized speech with the visual elements of the image. Continuous assessment and optimization serve as a roadmap for system enhancement, with integration and deployment constituting the final stage.

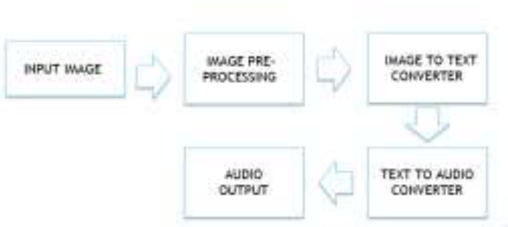
Text Detection: OCR-based algorithms will be used to detect and extract text from the pre-processed images. The OCR algorithm will be trained on a large dataset of images and corresponding text to achieve high accuracy in text recognition.

Text Recognition: Deep learning-based approaches, particularly convolutional neural networks (CNNs) and recurrent neural networks (RNNs), will be used to recognize the extracted text from the OCR algorithm. The neural network will be trained on a large dataset of images and corresponding text to achieve high accuracy in text recognition.

Text-to-Speech Conversion: The recognized text will be converted to speech output using text-tospeech (TTS) technology. The TTS algorithm will be trained on a large dataset of text and corresponding speech to achieve high accuracy in speech output.

Object Recognition: Object recognition-based algorithms will be used to detect and recognize objects in the images, including text. The recognized text will be extracted and passed through the deep learning-based approach to achieve high accuracy in text recognition and speech output.

Figure 1: Work Flow



3.1. User Interface

Our website homepage has an attractive interface with extract button which display the text in the form sound. The User Interface is as follows:

Figure 2: User Interface

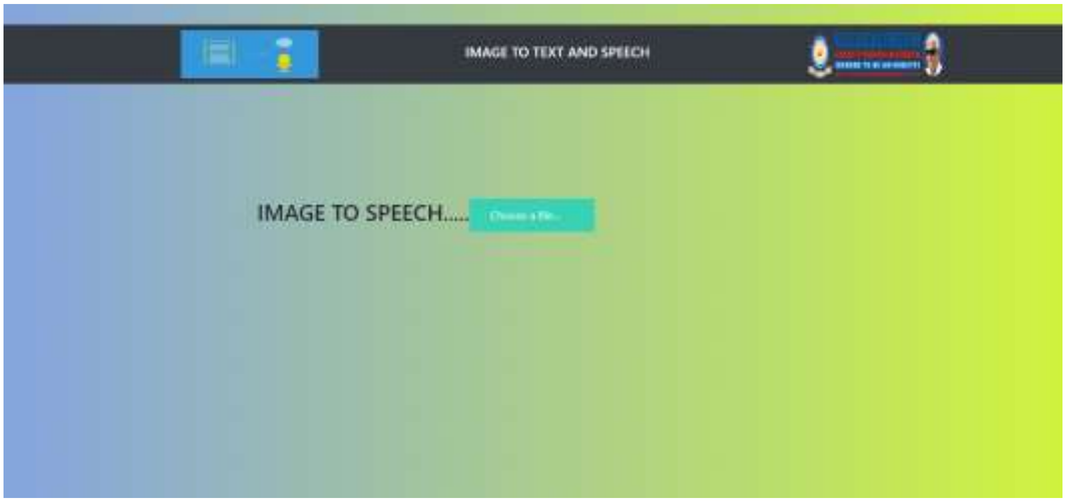


Figure 3: Sample Output 1

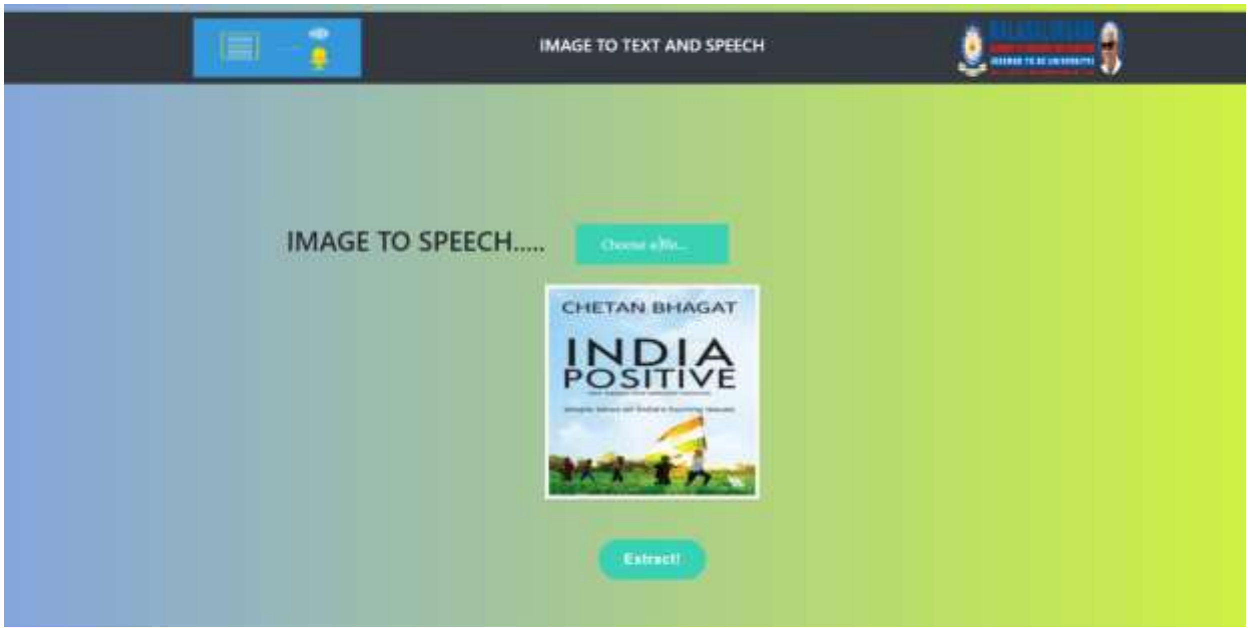
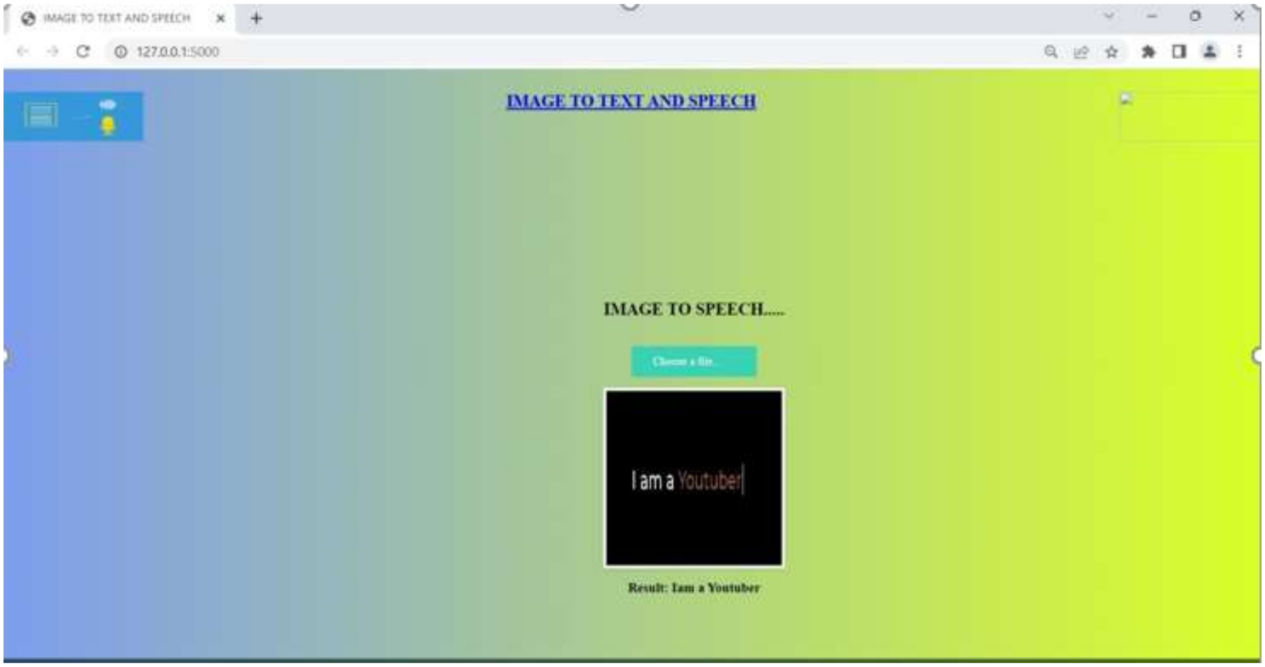


Figure 4: Sample Output 2



Result: Iam a Youtuber

REQUIREMENTS

4.1. Hardware Requirements

- 1 GHz CPU
- 1 GB RAM
- 500 MB disk space

4.2. Software Requirements

- Visual Studio Code
- Browser

Any of the following Operating Systems:

- Windows
- Mac
- Linux
- Unix

CHAPTER-4

MODULES DESCRIPTION

Pickle:

The main uses of pickle in Python are for serializing and deserializing object structures. To put it another way, it's the process of translating a Python object into a byte stream so that you can transfer data over a network, store it in a file or database, or maintain a programme state between sessions. By unpickling the pickled byte stream, the original object hierarchy can be recreated. The entire process is comparable to Java or.Net object serialization.

Easyocr:

The OCR process begins with scanning a document into a digital image. Once the document is digitized, the OCR software analyzes the image and identifies each character or symbol, such as letters, numbers, and punctuation marks. This process is achieved using machine learning algorithms and pattern recognition technology. Once the characters are recognized, the OCR software uses various algorithms to convert the image into text. This process involves recognizing the structure of the text and its relationships to other elements on the page, such as lines, paragraphs, and columns.

Pyttsx3:

Pyttsx3 is a text-to-speech conversion library in Python. Unlike alternative libraries, it works offline and is compatible with both Python 2 and 3. An application invokes the pyttsx3.init() factory function to get a reference to a pyttsx3. Engine instance. it is a very easy-to-use tool which converts the entered text into speech. The pyttsx3 module supports two voices first is female and the second is male which is provided by “sapi5” for Windows Flask:

Flask is a web application framework written in Python. Armin Ronacher, who leads an international group of Python enthusiasts named Pocco, develops it. Flask is based on the Werkzeug WSGI toolkit and Jinja2 template engine. Both are Pocco projects.

CHAPTER-5
EXPERIMENTAL RESULTS

Most of the images have text regions that can be identified by the suggested methodology, and it extracts text from those regions with high accuracy. We found that the suggested method can reliably identify text regions from images with varying text sizes, styles, and colors based on our experimental analysis. Our method still has trouble working on images with very small text regions and blurry text regions, even though it solves the majority of the problems other algorithms encounter. the result we get in this model will be in audio format and the text which is extracted from the image is converted to an audio file and we can download the audio file to understand the text in audio.

CONCLUSION

Nowadays, there is an increasing demand for text information extraction from images. As a result, numerous extraction methods have been created to retrieve pertinent data. Furthermore, it takes time to extract text from a color image, which annoys users. In this paper, we have proposed an improved method for accurately extracting text from images. With our approach, information can be extracted quickly. We developed it to make the model more accurate. Although our connected component-based approach to text extraction from color images has several advantages over the current method, it loses effectiveness when the text is too small, the text region is not visible, or the text's color is not visible. In the future, this work could be expanded to automatically document text in Word Pad or any other editable format for later use, as well as detect text from video or real-time analysis.

6.1. Future Scope

. In future, we can extend this work to detect the text from video or real time .

CHAPTER-6

REFERENCES

[1] Archana A. Shinde, D.G. Chougule “Text Preprocessing and Text Segmentation for OCR” IJCSET, January 2012.

[2] Benjamin Z. Yao, Xiong Yang, Liang Lin, MunWai Lee and Song-Chun Zhu, “I2T: Image Parsing to Text Description”.

[3] Bernard Gosselin Faculté Polytechnique de Mons, Laboratoire de Th´eorie des Circuits et Traitement du Signal, “From Picture to Speech: An Innovative Application for Embedded Environment”.

[4] Huizhong Chen1, Sam S. Tsai1, Georg Schroth, David M. Chen, Radek Grzeszczuk and Bernd Girod, “Robust text detection in natural images with edge-enhanced maximally stable extremal regions”, International Conference on Image Processing • September 2011

[5] Jisha Gopinath, Aravind S, Pooja Chandran, Saranya S S, “Text to Speech Conversion System using OCR”, International Journal of Emerging Technology and Advanced Engineering, January 2015.

[6] Itunuoluwasewon, JeliliOyelade, Olufunke Oladipupo, “Design and Implementation of Text-toSpeech Conversion for Visually Impaired People”, International Journal of Applied Information Systems (IJAIS, 2014).

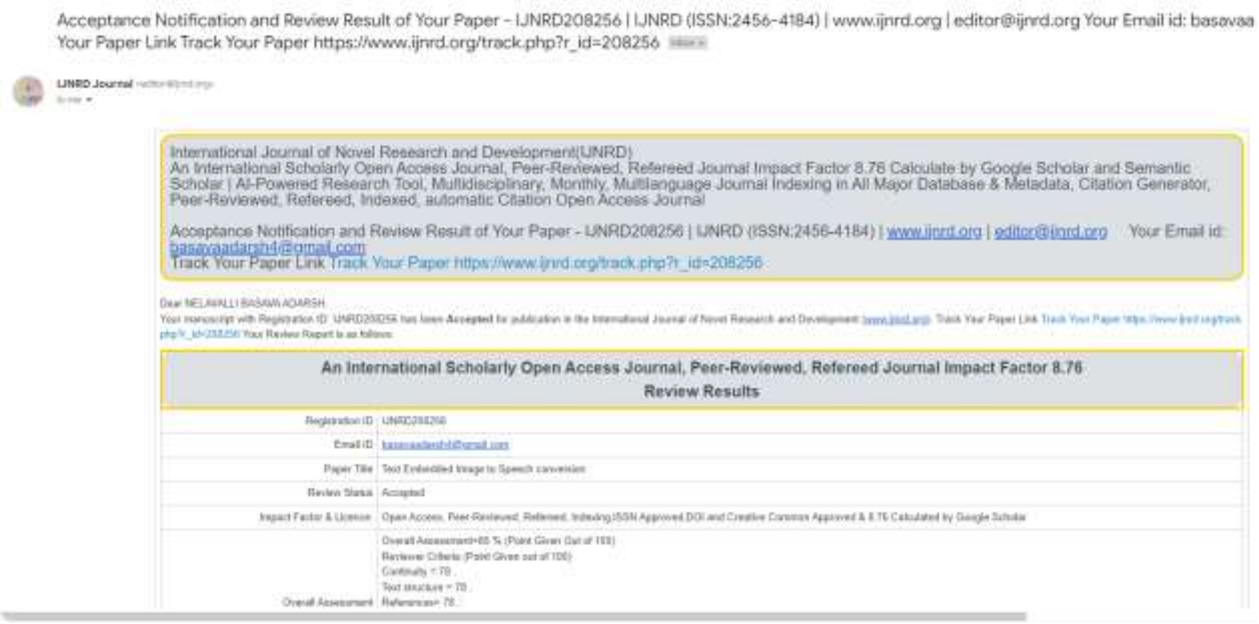
[7]Yao Li and Huchuan Lu, “Scene Text Detection via Stroke Width”, 21st International Conference on Pattern Recognition (ICPR 2012) November 1115, 2012. Tsukuba, Japan.

[8] Poonam.S.Shetake, S.A.Patil and P.M Jadhav “Review of text to speech conversion methods”, International Journal of Industrial Electronics and Electrical Engineering, Vol-2, Issue-8, August-2014.

[9] Kaveri Kamble and Ramesh Kagalkar, “Translation of text to speech conversion for the Hindi language”, “International Journal of Science and Research”, Impact Factor (2012): 3.358 Vol- 3 Issue-11, November 2014.

[10] Kumar Patra, Biplab Patra, PuspanjaliMohapatra, “Text to speech conversion with phonematic concatenation,” International Journal of Electronics Communication and Computer Technology “ (IJECCT) Vol- 2 Issue-5, September-2012, ISSN:2249-7838.

Status of Paper Publication:



Acceptance Notification and Review Result of Your Paper - IJNRD208256 | IJNRD (ISSN:2456-4184) | www.ijnrd.org | editor@ijnrd.org Your Email id: basavaa Your Paper Link Track Your Paper https://www.ijnrd.org/track.php?r_id=208256

IJNRD Journal editor@ijnrd.org
to me

International Journal of Novel Research and Development(IJNRD)
An International Scholarly Open Access Journal, Peer-Reviewed, Refereed Journal Impact Factor 8.76 Calculate by Google Scholar and Semantic Scholar | AI-Powered Research Tool, Multidisciplinary, Monthly, Multilanguage Journal Indexing in All Major Database & Metadata, Citation Generator, Peer-Reviewed, Refereed, Indexed, automatic Citation Open Access Journal

Acceptance Notification and Review Result of Your Paper - IJNRD208256 | IJNRD (ISSN:2456-4184) | www.ijnrd.org | editor@ijnrd.org Your Email id: basavaadarsh4@gmail.com
Track Your Paper Link Track Your Paper https://www.ijnrd.org/track.php?r_id=208256

Dear NELAVALLI BASAVA ADARSH
Your manuscript with Registration ID: IJNRD208256 has been **Accepted** for publication in the International Journal of Novel Research and Development (www.ijnrd.org). Track Your Paper Link Track Your Paper https://www.ijnrd.org/track.php?r_id=208256 Your Review Report is as follows:

An International Scholarly Open Access Journal, Peer-Reviewed, Refereed Journal Impact Factor 8.76
Review Results

Registration ID	IJNRD208256
Email ID	basavaadarsh4@gmail.com
Paper Title	Text Embedded Image to Speech conversion
Review Status	Accepted
Impact Factor & License	Open Access, Peer-Reviewed, Refereed, Indexing,ISSN Approved DOI and Creative Common Approved & 8.76 Calculated by Google Scholar
Overall Assessment	Overall Assessment=65 % (Point Given Out of 100) Reviewer Criteria (Point Given out of 100) Continuity = 78 , Text structure = 78 , References= 78 ,