Image to audio, text to audio, text to speech, video to text conversion using, NLP techniques.

Sai Harshith Thanneru[1]*, *Kajal* Kumari[1], *Naresh* Kunta[1], and *Pavan Kumar* Manchalla[1]

[1] Gokaraju Rangaraju institute of engineering and technology, Hyderabad, Telangana, India.

Abstract. Often, language bias between communicators can create communication problems. This article discusses a prototype that addresses this issue by enabling users to hear the content of text images. This involves extracting the text from the image and converting it into speech in the user's preferred language. Additionally, the device can be used by people with visual impairments. Overall, this device helps users to listen to the content of images being presented. The suggested system allows the user to take a picture, which is then scanned and analysed by the application to read the English text. The information obtained is then converted into voice, enabling visually impaired individuals to understand the content of the text. The output is delivered in speech format to provide access to information present on the document. To ensure better accuracy and performance, the system uses Natural Language Processing techniques. The system is designed with a Graphical User Interface (GUI) to improve accuracy and ease of use.

1. Introduction

Eye diseases, ageing, car accidents, and other causes contribute to an annual increase in the number of persons who are blind or visually impaired. Because reading is such an important part of human life (text can be found in newspapers, commercial products, signboards, computer screens, and so on), visually impaired people encounter numerous challenges. Visually impaired people can use mobile applications to assist them in reading text. The goal of our study is to see if a visually impaired person can acquire auditory information regarding comes to taking up a lot of space and using a lot of power. Threshold inverting quantization (TIQ) comparator can be printed text, text boards, scene text, hoardings, and traffic signboard instructions.

To attain the best results and reach state-of-the-art. Accuracy for the particular challenge, many techniques have been created for picture to audio conversion for blind persons. The current system is incapable of extracting text from photos and of providing an audio output for the image to be read. It can only transform text inputs into speech, making it ineffective for usage in real-world applications to assist blind people. The current model employs CNN, which is computationally intensive. They necessitate a large training data set and a lengthy preparation period.

* corresponding author: thannerusaiharshith@gmail.com

2. Literature survey

In [1], This paper presents a method for defining the area of interest (ROI) in which objects can be segregated from messy backgrounds that is both efficient and effective. To obtain text information, this ROI extracts the text localization and recognition. In [2], A printed copy of a book is the usual means of reading it. Books came in handy because it was impossible to carry a physical copy of a book everywhere. Both printed and electronic books have an impact on the eyes. In [3], The study is based on the Android platform, as well as machine learning. For blind persons, reading text from text pictures and text boards is a difficult effort. In [4], According to surveys, many find living in today's world to be extremely difficult. The writers have created a new device called the Blind Book Reader System to assist them. When blind people are in need, it's an easy remedy to get them to read books and newspapers. In [5], In this study, we describe a photograph (I2AD) task for creating audio descriptions from photographs, which can be used to enhance the visual experience of blind and visually impaired individuals. In [6], Text-to-speech (TTS) systems convert written text into speech signals. The conversion of Devanagari text to speech for Marathi printed text is done in this study. Two techniques are used to obtain the needed output: Optical Character Recognition and Pattern Recognition (OCR). In [7], Automation Audio Closed caption is described by the authors as a cross-modal job that provides descriptions of the sounds in audio samples using natural language. However, there is a lack of studies that focus on establishing the actual auditory events in the provided video based on the subtitle. As part of this effort, we've compiled a Stereo sound dataset 1 to illustrate the relationship between the audio events and the corresponding captions in Audio caps. In [8], The proposed technology is cost-effective and assists visually challenged individuals in hearing text.

The fundamental idea behind this project is to employ optical character recognition to convert text letters into audio signals. In [9], A model for converting natural Bengali language to text is presented in this research. The suggested model necessitates the use of the open-source Sphinx 4 framework, which is built in Java and provides the necessary procedural coding tools for developing an acoustic model for a bespoke language such as Bengali. Our key goal was to make sure that the system had received appropriate word-by- word training from a variety of people so that it could recognize new speakers fluently. In [10], As cloud computing adds new capabilities and services, it also brings new obstacles, such as cost, complexity, and integration issues. Traditional cloud architecture has been employed by several academics for their applications, such as text-to- speech (TTS).

Many systems have been developed for the image to audio conversion for blind people to get maximum results and achieve state-of-the-art accuracy for the given problem. The existing methods have achieved remarkable results for the text to audio, but it is still difficult to obtain speech with high accuracy. The existing system cannot detect the text from images and can neither provide an audio output for the image to be read. It can only convert the text inputs into speech and thus, isn't very efficient to be used in real-world applications for helping blind people. In the existing model, CNN is used which is computationally expensive. They require a huge training data set, and Its preprocessing time is quite high. The existing system often needs manual initialization, is not running in real-time, or does not work for images with many manuscripts, while also being sensitive to other factors.

3. Proposed work

We propose a model for the image to audio conversion for blind people using machine learning techniques. Input will be given in the form of an image, which will be processed so that the texts in the image would get converted to audio, to be received as the output of the system. The proposed system has shown excellent performance with a high accuracy rate and a much higher speed up rate as compared to the previously used state-of-the-art methods. The proposed framework is not only much faster than the previous work but also maintains competitive accuracy with the state-of-the-art human detection system. It has very precise measurements and permits for high deployment and authentication. Use of our framework is not limited to a single field of application and is useful for many more real- world applications. Our proposed model for the image to audio conversion is beneficial to the world for advanced applications in helping visually impaired people.

3.1 Feasibility study

In this stage, the various kinds of feasibility are analyzed, and business proposals with a high-level outline and some cost estimates are presented. The proposed system's viability should be examined during the system analysis phase. This is to guarantee that the planned system does not pose a burden on the business. Understanding the system's primary requirements is crucial for conducting a feasible analysis. The feasibility analysis must consider the following two factors. The financial impact on the business is evaluated to determine the system's economic feasibility. There is a cap on the company's ability to spend on the system's research and development. Any money spent must have a good reason behind it. As a result, the produced system is both functional and cost-effective, with the latter being made possible by the widespread availability of the technologies employed. It was required to buy only the personalized items.

3.2 The possibility of implementation technology

Examining the system's technical needs is an important step in determining its technical feasibility. Any system used shouldn't place an excessive burden on the existing infrastructure. Because of this, there will be a strain on the available technological infrastructure. Thus, the client will be subjected to excessive pressure as a result. The established system needs to have minimum needs, as adopting it will call for few if any adjustments.

The goal of evaluating social practicality is to determine how well the system will be received by the target audience. This includes the method of instructing the user to operate the system efficiently. The user must accept the system as a necessary evil rather than a threat. How well a system is received by its intended audience is entirely dependent on the means used to introduce it to and familiarize the user with it. Since he is the system's end user, his level of assurance must be increased so that he can also offer critical feedback, which is always appreciated.

4. Methodology

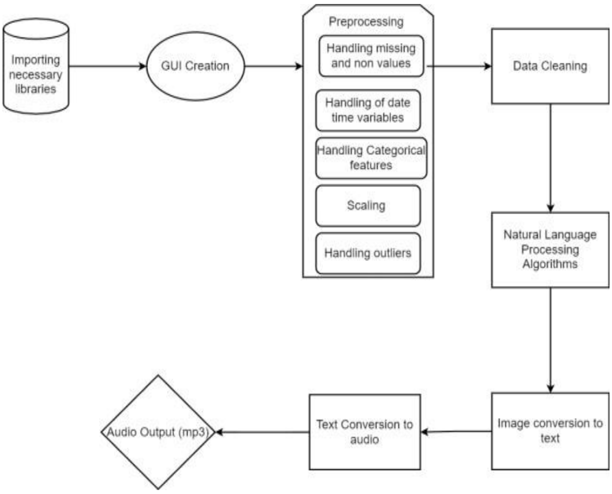

Importing necessary libraries
GUI Creation
Preprocessing
Handling missing and non values
Handling of date time variables
Handling Categorical features
Scaling
Handling outliers
Data Cleaning
Natural Language Processing Algorithms
Image conversion to text
Text Conversion to audio
Audio Output (mp3)

Fig. 1. Architecture diagram.

The diagram represents of all the entities that have been included into the system with relation between each of them and involves a sequence of decision-making processes and steps.

5. Use Case Diagram

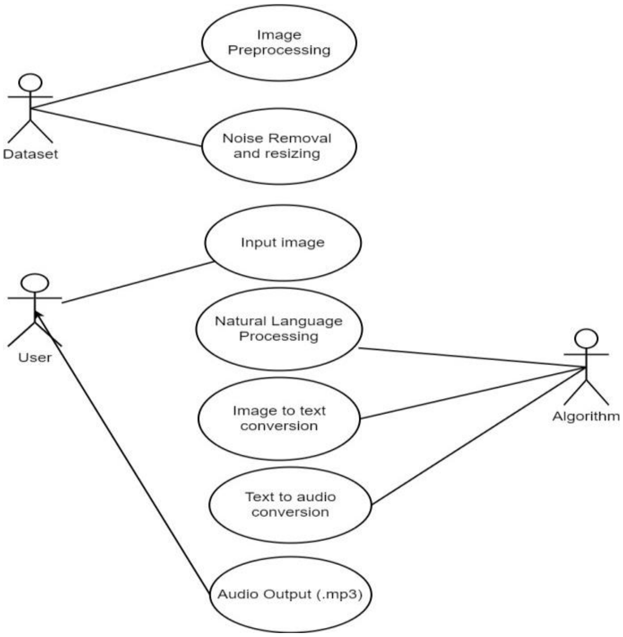


Fig. 2. Use case diagram.

6. Implementation

When a project reaches its implementation phase, the conceptual designs are transformed into a fully functional system. Therefore, it is the most crucial step for realizing a successful new system, empowering the user, and inspiring trust that the system will function as intended and provide maximum efficiency. The implementation phase entails thorough preparation, research into the issues of the current system and its limits on implementation, the development of strategies to bring about a radical shift, and an assessment of thenew approaches taken. Throughout the training phase of our proposed system, the classification models were evaluated withthe help of the confusion matrix. Results are compared with what was anticipated using the confusion matrix. True PositiveRate (TPR), False Positive Rate (FPR), Precision, Accuracy (AC), F1 score, and Misclassification rate were also used as measures of effectiveness alongside proper classification rate or accuracy. The accuracy coefficient (AC) measures how often aforecast turns out to be correct. By dividing the sum of false positives and true negatives by the sum of erroneous positive predictions, we obtain the false-positive rate of our model.

7. Input and Output design

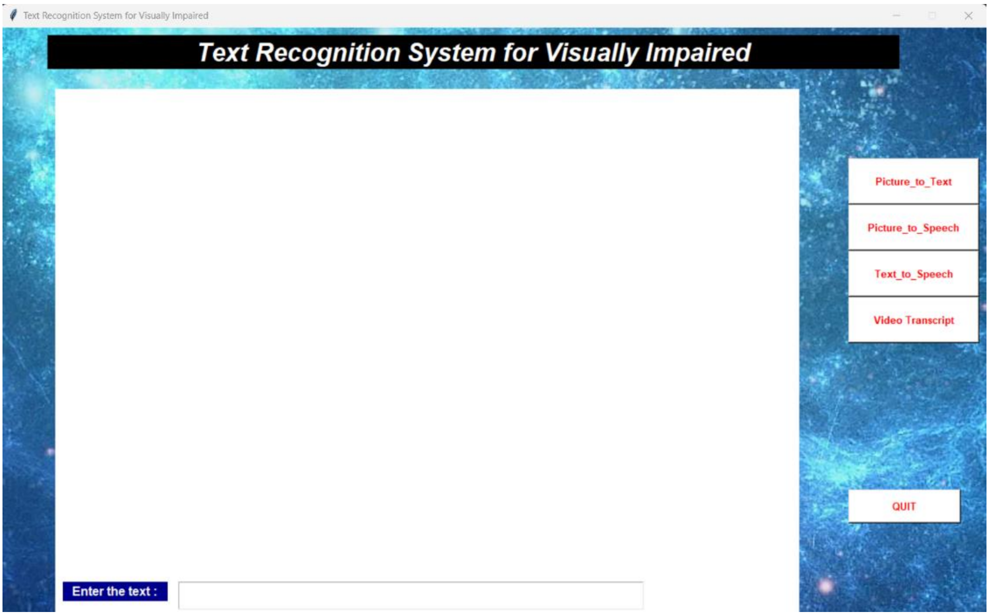
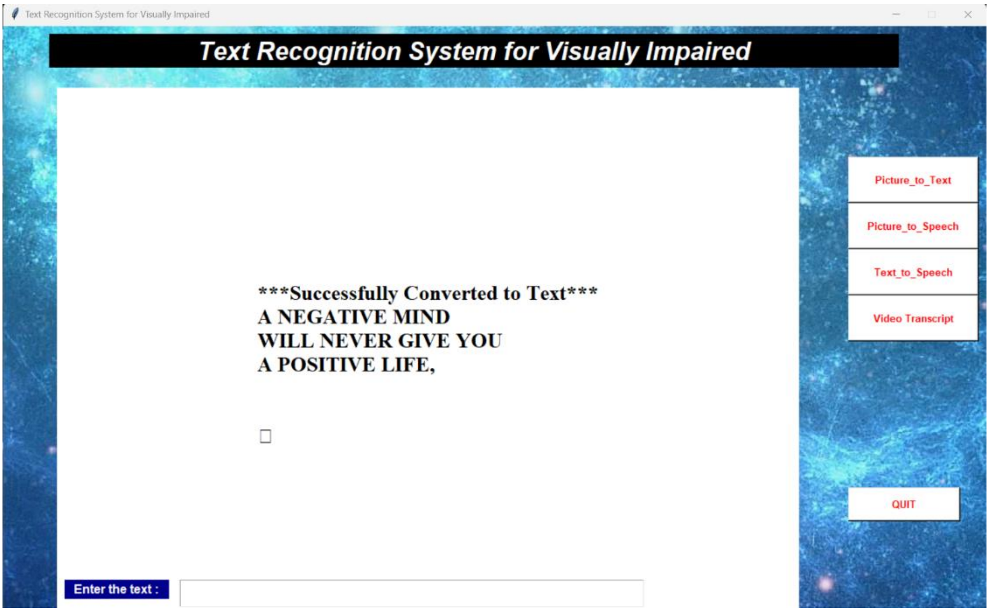

Text Recognition System for Visually Impaired
Text Recognition System for Visually Impaired
Picture_to_Text
Picture_to_Speech
Text_to_Speech
Video Transcript
QUIT
Enter the text :

Fig. 3. Input Design.


Text Recognition System for Visually Impaired
Text Recognition System for Visually Impaired
Picture_to_Text
Picture_to_Speech
Text_to_Speech
Video Transcript
QUIT
Enter the text :

Successfully Converted to Text
A NEGATIVE MIND
WILL NEVER GIVE YOU
A POSITIVE LIFE,

□

Fig. 4. Picture to text

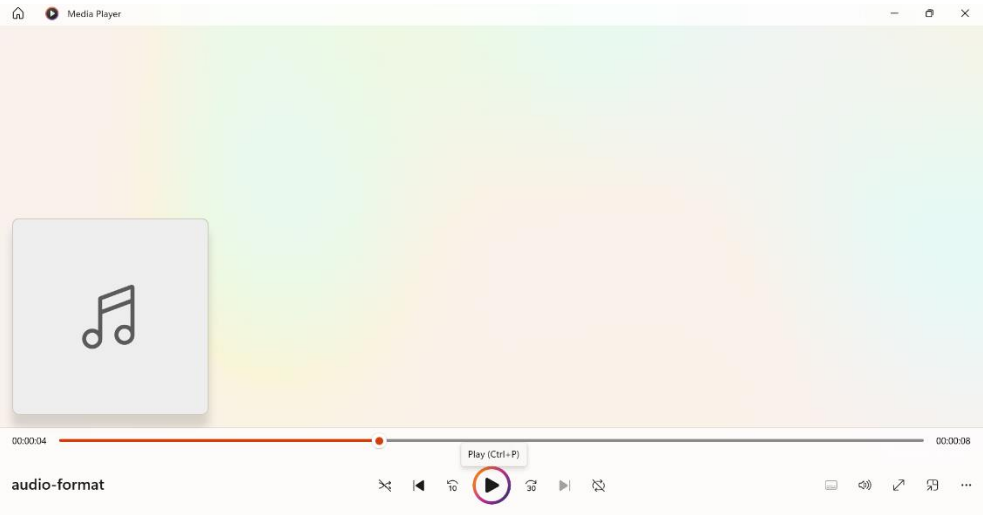
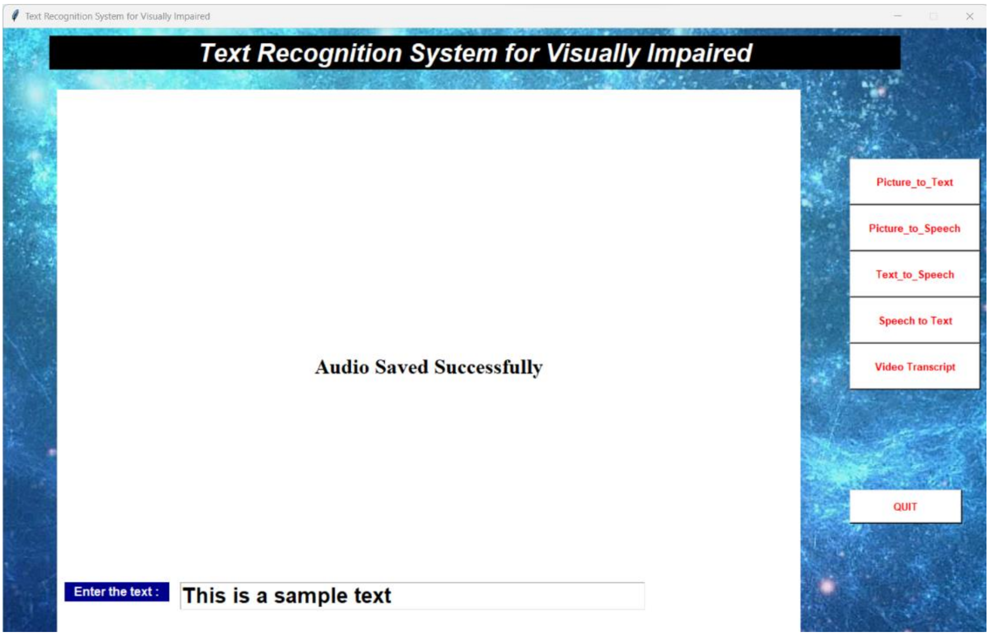


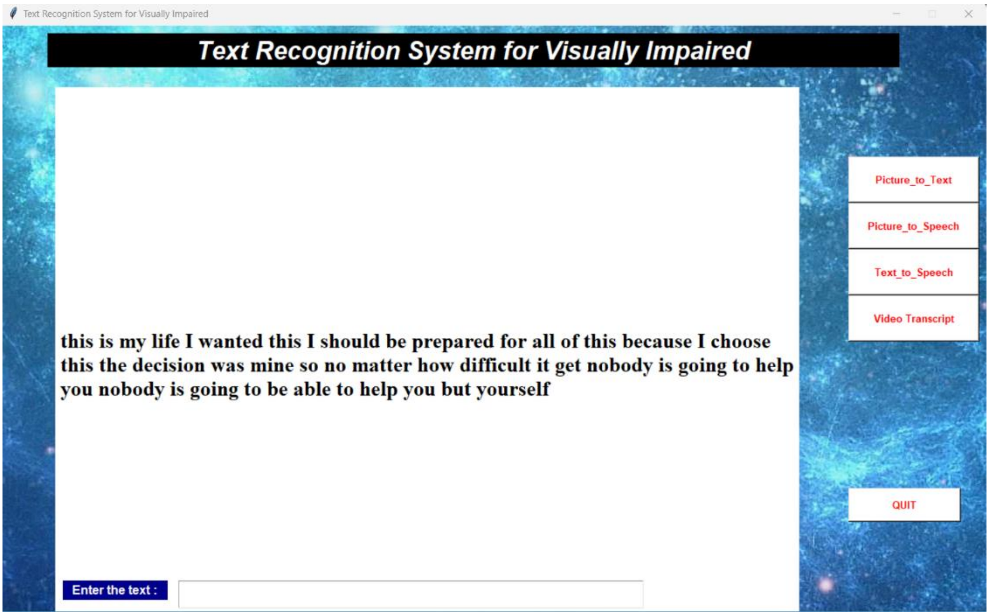
Fig. 5. Picture to speech.



Audio Saved Successfully

This is a sample text

Fig. 6. Text to speech.


Text Recognition System for Visually Impaired
Text Recognition System for Visually Impaired
Picture_to_Text
Picture_to_Speech
Text_to_Speech
Video Transcript
QUIT
Enter the text :

this is my life I wanted this I should be prepared for all of this because I choose this the decision was mine so no matter how difficult it get nobody is going to help you nobody is going to be able to help you but yourself

Fig. 7. Video transcript.

8. Results and Discussion

Efficiency of the Proposed System:

- The proposed system is done in real-time and also theaccuracy is high in the proposed system.
- The proposed system loading speed and execution speed is really fast when compares with the existing system.
- The proposed system is highly efficient and scalable and is also further improved for complex use cases.

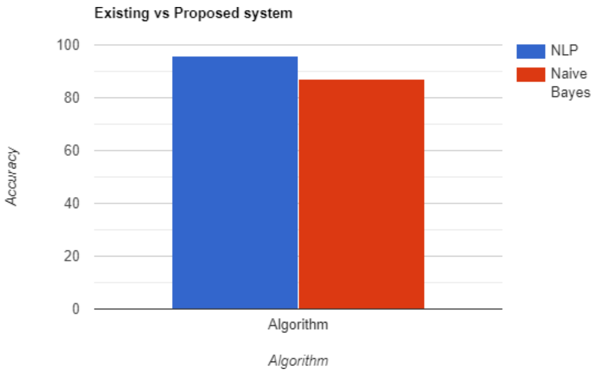


Fig. 8. Graph of NLP vs Naïve Bayes (Proposed vs Existing).

- In the Existing System, there are individual functions and are not efficient.
- In the existing system preprocessing time is very high and It requireshuge Data set. But in this proposed work, one of the efficient calledNLP technique are used.
- In the proposed work we use the effective tool called Pytesseract- OCR.

9. Conclusion and Future enhancements

Some sort of program is required to aid the visually impaired and the elderly in detecting the text so that they can read labels on medications and other printed materials. We therefore offer a model for blind persons to use machine learning approaches to convert images into sounds. The visually challenged will benefit from having the document's contents read aloud to them, which is why the output will be in the form of voice/speech. Using the camera on their phone, users of the present scheme will be able to scan any document presented to them, have the content read out to them in English, and have the text converted into speech. The system can be used with little effort and at a reasonable cost.

This project can be deployed in any cloud platform like netlify or digital ocean so everyone can access this application. Also, the accuracy can be increased by trying with many algorithms.

References

[1] K. C. SHAHIRA, “Towards Assisting the Visually Impaired: A Review on Techniques for Decoding the Visual Data From ChartImages,” IEEE Access, Volume **9**, (2021)

[2] Sai Aishwarya Edupuganti, Vijaya Durga Koganti, Cheekati Sri Lakshmi, Ravuri Naveen Kumar, “Text and Speech Recognition for Visually Impaired People using Google Vision,” 2021 2nd International Conference on Smart Electronics and Communication (ICOSEC), (2021)

[3] Asha G. Hagargund, Sharsha Vanria Thota, Mitadru Bera, Eram Fatima Shaik, “Image to speech conversion for visually impaired,”International Research Journal of Engineering and Technology (IRJET), Volume **03**, (2017)

[4] Prabhakar Manage, Veeresh Ambe, Prayag Gokhale, Vaishnavi Patil, “An Intelligent Text Reader based on Python,” 2020 3rd International Conference on Intelligent Sustainable Systems (ICISS), (2020).

[5] Samruddhi Deshpande, Revati Shriram, “Real time text detection and recognition on hand held objects to assist blind people,” 2016 International Conference on Automatic Control and Dynamic Optimization Techniques (ICACDOT), (2016).

[6] D.Velmurugan, M.S.Sonam, S.Umamaheswari, S.Partha-sarathy, K.R.Arun. A Smart Reader for Visually Impaired People Using Raspberry PI. International Journal of Engineering Science and Computing IJESC Volume **6,** Issue No. 3. (2016)

[7] K Nirmala Kumari, Meghana Reddy J. Image to Text to Speech Conversion Using OCR Technique in Raspberry Pi. International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering Vol.-**5**, Issue-5, May-(2016).

[8] Silvio Ferreira, C´eline Thillou, Bernard Gosselin, From Picture to Speech: An Innovative Application for Embedded Environment. Faculté Polytechnique de Mons, Laboratoire de Théorie des Circuits et Traitement du Signal Bˆatiment Multitel - Initialis, 1, avenue Copernic, 7000, Mons, Belgium. (2009)

[9] Nagaraja L, Nagarjun R S, Nishanth M Anand, Nithin D, Veena S Murthy Vision. based Text Recognition using Raspberry Pi. International Journal of Computer Applications (0975 – 8887) National Conference on Power Systems & Industrial Automation. (2015)

[10] Poonam S. Shetake, S. A. Patil, P. M. Jadhav Review of text to speech conversion methods.s (2014)