Kannada Text to Speech Conversion System

Pawan S Nadig, Pooja G, Kavya D, R Chaithra, Radhika A D

Abstract*: The following paper describes the design of a system which does text to speech generation for one of the regional language's Kannada. The printed document of Kannada text is given as input to the system, the system then converts the document to an image format. Pre-processing is done to stabilize the intensity of the images and clear the artifacts. This process boosts the precision and interpretability of an image. Optical Character Recognition (OCR) is used to unsheathe the segmented characters from a particular image and are matched with the characters that have been stored in the dataset. Once the matched characters are extracted it is stored in a suitable format and then the TTS engine is deployed to convert the saved Kannada characters to a speech format. The obtained speech output corresponds to the characters which are collected after processing the input text.

Keywords: Image, Kannada, OCR, Pre-processing, Text-to-Speech

I. INTRODUCTION

The main goal of Text-to-Speech synthesis is to generate the speech output for the given input of text. The 2 main components of Text to Speech System are Text processing and speech generation Many TTS systems have been developed for European languages like English, German, Greek using a wide variety of approaches like concatenation, digital signal processing, articulatory synthesis, Neural networks, etc. When it comes to regional languages it is classified into Indo-Aryan languages (Hindi, Marathi, Punjabi, Gujrati, etc) and Dravidian languages (Kannada, Tamil, Telugu, etc). Each regional language has a wide number of characters when compared to English, It becomes difficult to process these characters as each of these have a different font style. Identifying the different font styles and mapping these characters with the dataset is one of the major challenges. Due to this problem, a very less number of TTS systems have been designed for a few regional languages. In this paper, we are describing how the TTS system is implemented for the Kannada Language.

The Kannada language is spoken predominantly by the people of Karnataka. The letters of Kannada are classified into three categories: Svara(vowels), vyanjana (consonants), and yogavahaka (semiconsonants). Each letter has its own form and sound, providing the visible and audible representations, respectively.

Revised Manuscript Received on July 29, 2020.

Pawan S Nadig, Department of Computer Science and Engineering Vidyavardhaka College of Engineering Mysuru, India.
Pooja G, Department of Computer Science and Engineering Vidyavardhaka College of Engineering Mysuru, India.
Kavya D, Department of Computer Science and Engineering Vidyavardhaka College of Engineering Mysuru, India.
R Chaithra, Department of Computer Science and Engineering Vidyavardhaka College of Engineering Mysuru, India.
Radhika A D, Department of Computer Science and Engineering Vidyavardhaka College of Engineering Mysuru, India.

As the number of characters in Kannada is very huge and also a variety of font styles, converting all these characters to its corresponding speech output is challenging work. Optical character recognition technology is used to extract the given Kannada input text. Optical character recognition or optical character reader (OCR) can be regarded as the electronic or mechanical alteration of images that are basically typed, handwritten, or printed text into machine-encoded text. Kannada Printed text recognition becomes a complex task as it contains a diversification in Font style and vowel modifier. Different OCR tools have been developed for printed documents. The proposed system helps to reduce the storage work of Printed documents as it converts text images into electronic records. It is one of the most accurate techniques to extract the characters of the given input text and convert them to speech output using the TTS Engine [1][9].

II. RELATED WORK

A Literature review of various techniques of TTS is done before the implementation of our project. This gave us an idea about various approaches and techniques that can be used to develop our own TTS application.

A. Literature Review of TTS in other languages

1) "Text Image to Speech Conversion using Matlab and Microsoft SAPI" by Malti Bansal, Shivam Sonkar [2]

This paper makes use of Matlab, Microsoft SAPI, and OCR program was developed on Matlab for text extraction. The first step deals with the extraction of text from a sample image using the OCR program which is programmed and developed on Matlab. Then the selected image will be converted into a grayscale. This is done by eliminating the hue and saturation data while retaining the luminance. Then it will be converted into a black and white image and then to word matrix. After preprocessing of an image is done they load the template to match the letter with the templates then they resize the letter according to template size. Next, the templates are compared with the extracted characters and write the output into a text file and run the visual basic program. The text file acts as an input to a visual basic code that reads the text line by line until it gets a blank line. Finally, it uses a Microsoft SAPI to convert text into speech.

2)"Implementation of Text to Speech Conversion" by Chaw Su Thu Thu, Theingi Zin [3]

This paper is mainly concerned with the implementation of the OCR system for the English capital character A to Z and number 0 to 9 identification. A character is identified at one time and then this is saved in a notepad file in text format. The program is basically divided into 2 portions.

The first part gives a text output based on the image and converts it into speech, while the second one directly converts the e-text in the computer to a speech.

3)"Real-Time Implementation of Optical Character Recognition Based TTS System Using Raspberry pi" by Sagar G.K, Shreekanth T [4]

The above-mentioned paper follows a pioneering yet well-organized real-time cost-beneficial technique. This approach allows users to actually hear the image contents and discards the traditional method of reading through them. The concepts of Optical Character Recognition(OCR) and Text to Speech Synthesizer(TTS) are combined in the Raspberry pi. This system is optimal for virtually impaired people. They can easily interact with computers through a vocal interface. There are 2 modules in this device. The first module does the work of processing the image while the second one does the work of processing the voice. A processer with 900MHz processor speed was developed on a Raspberry Pi.

4)"Image to Speech Conversion for Visually Impaired" by Asha G. Hagargund, Shasha Vanria Thota, Mitadru Bera, Eram Fatima Shaik [5]

The main concern of this particular paper is to define information about the device that captures the image's text and converts it into speech. A framework is created which is an embedded system. It takes a picture and unsheathes only the area of concern. The above process is done using Raspberry Pi and Raspberry Pi Camera. The image taken is made to undergo a sequence of pre-processing steps. This is to trace only the writing in the image and remove the background. There are mainly 2 tools that are being used for this. The first one being OCR (Optical Character Recognition) software and the second one being TTS (Text-to-Speech) engines. The audio output can be obtained by the Raspberry pi's audio jack which can be directly connected to the speaker or earphones.

5)"Text to Speech Conversion System Using OCR" by Jisha Gopinath, Arvind S, Pooja Chandran, Saranya S [6]

This paper uses open-source OCR software called Tesseract is used as a basis for the implementation of a text reading system for visually disabled in the Android platform. Tesseract is considered the most accurate free OCR engine in existence. Users can choose an image already stowed on the Android device or make use of the device's camera to capture a new image. When the OCR process is whole as it produces a return a string of text which is displayed on the user interface screen, where the user is also allowed to edit the text then using the TTS API enables our Android device to speak the text of different languages. The TTS engine that ships with the Android platform support several languages: English, French, German, Italian, and Spanish. Also, American and British accents for English are both supported. The TTS engine needs to know which language to speak. So the voice and dictionary are language-specific resources that need to be loaded before the engine can start to speak.

6) "Image Text to Speech Conversion Using OCR Technique in Raspberry PI" by Jarlin James, Michael Aldo, Adellina Andrew [7]

This particular paper mainly takes the following steps. Firstly, the device is hovered over the page that is printed and the text on it is captured. Letters are extracted and then it will be converted into a digital format. Three steps are required to be followed in this method. Them being, Skew Correction, Linearization, and Noise Removal. The picture taken is checked for skewing. The determination for skew recognition finds for a position of direction and if noticed then a simple image rotation is carried out till the lines match with the horizontal axis, which produced a skew corrected image. The noise that is being introduced during a capture of an image or due to the low quality of the page is cleared before processing it further. Next, the course is moved on to check for interline spaces. After the interline spaces are detected, the image is further segmented into sets of paragraphs. This is done across the interline gap. The lines are scanned vertically for vertical space intersection and horizontally for intersection with the background. Histograms are deployed to detect the width of words. In the next step, words are decomposed into character width composition. To convert the image file to text file, they have used Tesseract software. This is done by collecting the texts from the image and stowing it in the file with.txt extension. A Text to Speech Google Text to Speech API is used to convert the text the user inputs into speech.

B. TTS using a different modeling approach

1) Using the Lab VIEW

There are number of graphical languages used in most of the applications. LabVIEW is one among those graphical programming languages. In Lab VIEW icons are used to create the applications rather than a text. Dataflow programming language is the main approach used in this. The data flows through the different nodes involved. The execution order is determined by the VIs and functions. VI's are used to emulate physical instruments. The user can build their own user interface by using different tools and resources that are available in LabVIEW. Front end created by the users is known as user interface. To control different objects in the front user can add their code using different functions and referring graphical representations. G code or block diagram code refers to the graphical source code. A LabVIEW programming basically acts like a text related program which uses different functions and methodologies. LabVIEW acts as a virtual instrument in appearance. LabVIEW accepts a input and process it to produce output which in turns behaves as a real instrument.

2) Using Android

- Android one of the most important technology is mainly used as a Linux-based operating system in most of the appliances including mobile phones and laptops. Android is developed by the Open Handset Alliance led by Google. Open source OCR software called Tesseract is basically used for the development of text reading system for visually disabled in Android platform.

Currently, Tesseract is considered the most accurate free OCR engine in existence. Here the user can choose the image which is previously saved on the Android device or the users can capture the image using their device camera. Rectification Algorithm is used to process the and further the input image is passed to the Tesseract service.

When the process is done completely it produces a string of text which is projected on the user interface screen, where the user can edit the text further using the TTS API enables the device to produce speech output of different languages.

Android software is embedded with the TTS engine which helps us to support a variety of languages: English, German, Italian, Spanish and French. Before making the TTS engine to speak all the required language-specific resources must be loaded.

III. METHODOLOGY

The TTS Application for the Kannada language is designed using MATLAB programming platform. It is an easy-to-use environment where a different problems and solutions can be represented in familiar mathematical notation.

The system design includes the following phases:

A. Optical Character Recognition (OCR)

Currently, in the running world, there is a huge demand for the conversion of printed documents into electronic forms to maintain the data securely. Researches have proved that the barcode and character/text recognition can be done through Optical Character Recognition in a much reliable way. The main role of OCR is to convert the electrical or mechanical images, handwritten or printed text into machine-encoded text. It involves various stages of functioning to the process of the text file. Each stage has its own functionalities which help us to extract the characters as per the requirement. It is used in a wide variety of applications because of its feature which extracts the given text [8].

Here we use the image of Kannada printed text for the conversion. The process of OCR is composed of different phases. These phases are as follows:

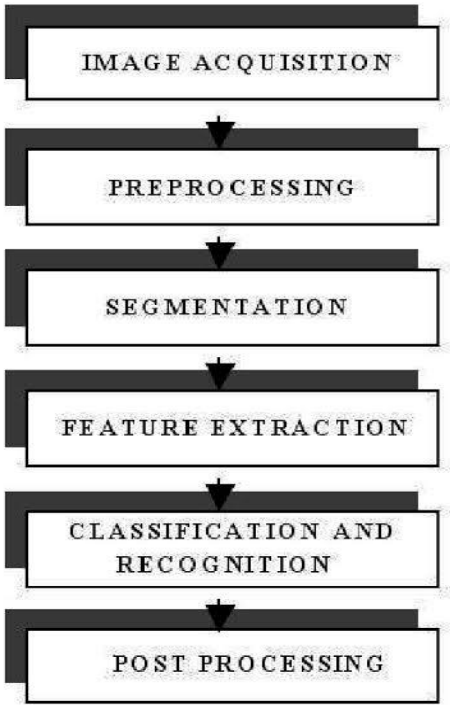

Fig 1: Phases of OCR

1) Image Acquisition

Image Acquisition is the initial step of OCR to capture the image from an external source. The obtained digital image is then converted to a suitable format that can be easily processed by the system in further steps. This step involves quantization and also a compression of the image. A special case sometimes includes binarization of an image which represents the image in a binary format.

2) Pre-processing

Once the image has been obtained, different pre-processing steps are involved to improve the quality of the image. These steps include removal of noise like blur effects and other unwanted characters/ artifacts from the image. It sets the threshold value of an image based on our requirements. Finally, a black and white image without any noise with a particular threshold value is obtained after pre-processing.

3) Segmentation

In this step, the image can be segmented into characters, words, or sentences according to our requirements. This phase helps to provide the contextual information of the image. Each character segmented represents a bounded box. This gives the measurement of the character in a matrix form. The segmented characters act as a by-product for the further phases.

4) Feature Extraction

It is one of the most important phases in OCR. Here the segmented characters are processed to extract different features. The features can be of different types such as image itself, geometrical and statistical features. Finally, various techniques can be used to reduce the dimensionality of the image.

5) Classification, Recognition, and Post-preprocessing:

This phase includes the following stages in the processing of text:

a) Template Matching

After comparing the characters with the predefined data stored in the system there will be an exact unique match of the character. This match helps us to identify the character's name. The original character is extracted in this process which is stored afterward. Template matching is done using the predefined dataset i.e. a collection of characters that is used to match the stored characters.

b) Storing the data

After template matching the characters obtained can be stored in any suitable format. Later using this stored data the characters are converted to speech format.

B. TTS API

API is a widely used computing interface that is used to interact between multiple software intermediaries. Microsoft Windows 10 Speech API is used in our project to convert the characters to speech format.

Steps to enable Speech API in Windows 10:

1. Type in “Control Panel” in the search bar.
2. Click on the Ease of Access Selection.
3. Click on Speech Recognition Selection.
4. To the left, you will see the Text to Speech Selection, click that.
5. Adjust your Settings and apply from here

C. Speech Output

Finally, the speech output is obtained concerning the given input format by combining the results of both the phases such as OCR and Speech API. The speech is mainly dependent on the API used in the system design since the accent varies for different APIs. Fig: shows the characters obtained after template matching.

The processing of input text through OCR is fast. Large quantities can be fed as a input to the system which then processes the text removing all noises and gives the required formatted text for further conversion. The OCR technique provides the accurate information of extracted text this helps in easy conversion of text to speech format.

The Fig.2 below shows the flow diagram of project implementation:

The first stage is to give the input image of the text in .JPG in MATLAB framework. Further, the image is pre-processed followed by segmentation and feature extraction using the OCR approach. The matched characters after template matching are stored and Speech API is used to generate output for the given input. Here each block in the diagram performs its own tasks. The output obtained from one phase serves as an input to the other. Finally, all information is collected and stored in a suitable format to convert the given input text into a speech format using APIs.

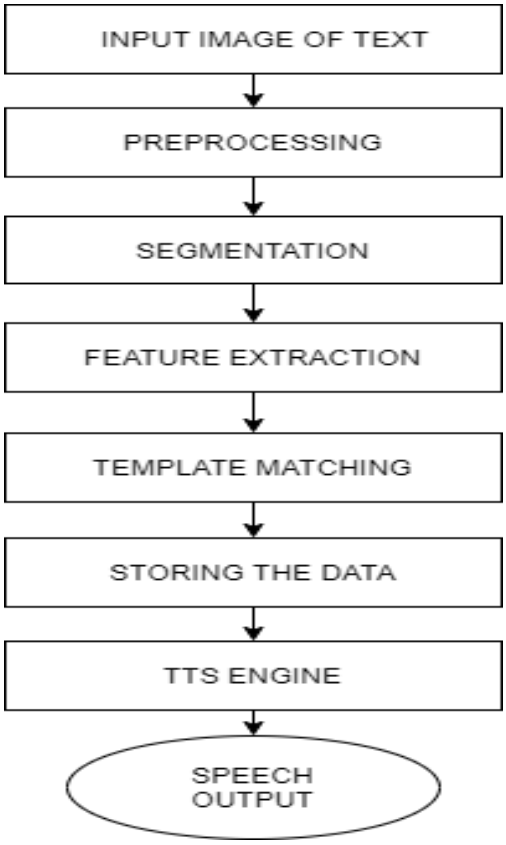

Fig. 2: Flow Diagram of TTS

IV. RESULT

Using Optical Character Recognition technique Kannada text images are effectively converted into its respective speech. The speech output produced is driven from text-to-speech API. The result accuracy will be dependent on the training data. Also, use input language as English (India) to get a proper accent. Second text to speech conversion using MATLAB. Like other TTS system such as Google TTS internet connectivity is not needed, it is one of the main advantages. Here, for example we have considered the input image as ಶುಭ ಸಂಜೆ and converted that into English using the OCR technique and produced speech output for SHUBHA SAANJE. Like this different Kannada text image can be fed to the system to produce speech output.

The Fig.3 & 4 below show the input of Kannada image text and the output obtained respectively.

Fig. 3. Input image of printed text

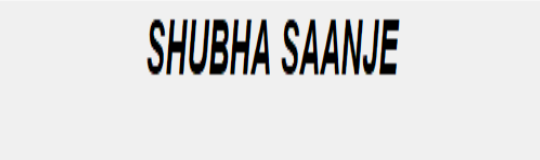

Fig. 4. Speech output obtained

V. CONCLUSION AND FUTURE WORK

In this paper, we have developed an application that converts the given Kannada image of text into speech in MATLAB platform. This system consists of two sessions: In initial stage OCR approach is used and in the second phase is speech is synthesised using APIs. In OCR the given Kannada image is scanned and converted into a gray image and further to black and white image and then it is converted into text by MATLAB and then the text is segmented and recognised using segmentation and template matching methods respectively. Finally, an API is used to read the stored characters. Recognition of extracted text is completed through the OCR approach with better accuracy and the text, which is obtained as the output after OCR process is sent to the TTS engine which then produces the speech output. This application helps blind people to read textual characters. As a further extension, we can develop an application that includes handwritten and complex font text. And also we can develop an application that performs text extraction from videos instead of images.

ACKNOWLEDGMENT

We would like to express gratitude to our project guide Ms. Radhika A D, for providing her guidance throughout our entire project work, we were able to face the challenges that encountered at all stages and were to meet the standards that was set for the project. The project would not have been a great success without her support. We would also like to extend our gratitude to our Head of the Department, Computer Science and Engineering and also the principal of Vidyavardhaka College of Engineering, for providing us with all the facilities that were required.

REFERENCES

1. B.M. Sagar, Dr. Shobha G, And Dr. Ramakanth Kumar P, OCR for printed Kannada text to Machine editable format using Database approach, 9th WSEAS International Conference on AUTOMATION and INFORMATION (ICAI'08) Bucharest, Romania, June 24-26, 2008.
2. Malti Bansal, Shivam Sonkar, *Text Image to Speech Conversion using Matlab and Microsoft SAPI*, IJEECS ISSN 2348-117X, vol. 6, Issue 11 *November 2017*.
3. Chaw Su Thu Thu and Theingi Zin. "Implementation of Text to Speech Conversion," International Journal of Engineering Research & Technology (IJERT), ISSN: 2278-0181, vol. 3, Issue 3, March, 2014.B. Smith, "An approach to graphs of linear forms (Unpublished work style)," unpublished.
4. Sagar G.K and Shreekanth T, *Real Time Implementation of Optical Character Recognition Based TTS System using Raspberry pi*, International Journals of Advanced Research in Computer Science and Software Engineering, July 2017.
5. Asha G. Hagargund, Sharsha Vanria Thota, Mitadru Bera, Eram Fatima and Shaik, *Image to Speech Conversion for Visually Impaired*, International Journal of Latest Research in Engineering and Technology (IJLRET), ISSN: 2454-5031, vol 03, June 2017, pp. 09-15C. J. Kaufman, Rocky Mountain Research Lab., Boulder, CO, private communication, May 1995.
6. Jisha Gopinath, S Aravind , Pooja Chandran, and S S Saranya, Text to Speech Conversion System using OCR, International Journal of Emerging Technology and Advanced Engineering, January 2015.
7. Jarlin James, Michael Aldo, Adellina Andrew, *Image Text to Speech Conversion using OCR Technique in Raspberry PI*, International Journal of Scientific & Engineering Research vol. 9, Issue 3, *March-2018*.
8. Nilesh Jondhale and Dr. Sudha Gupta, *Reading text extracted from an image using OCR and android Text to Speech*, International Journal of Latest Engineering and Management Research (IJLEMR), ISSN: 2455-4847, vol. 03, April 2018, pp. 64-67.
9. K Nirmala Kumari and Meghana Reddy J, Image Text to Speech Conversion Using OCR Technique in Raspberry Pi, International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering, vol. 5, May 2016.
10. Suhas R. Mache, Manasi R. Baheti, and C. Namrata Mahender, *Review on Text-To-Speech Synthesizer*, International Journal of Advanced Research in Computer and Communication Engineering, vol. 4, August 2015.
11. John Colaco and Sangam Borkar, *Design and Implementation of Konkani text to Speech Generation System using OCR Technique*, International Journal of Scientific and Research Publications, vol. 6, September 2016, ISSN 2250-3153.
12. M. Nagamani, S.Manoj Kumar and Uday Bhaskar, *Image to Speech Conversion System for Telugu Language*, International Journal of Engineering Science and Innovative Technology (IJESIT), volume 2, November 2013, ISSN: 2319-5967.
13. K.N. Natei, J. Viradiya, and S. Sasikumar, *Extracting Text from Image Document and Displaying Its Related Information*, K.N. Natei Journal of Engineering Research and Application, ISSN : 2248-9622,vol. 8, Issue5 (Part -V) May 2018, pp 27-33.
14. A.G. Ramakrishnan, H.R. Shivakumar, Lakshmi Chithambaran, *Text to Speech Synthesis in Indian Languages*, Apr 7, 2018.
15. Yamini D. Patil and Nandkishor C. Patil, Optical Character Recognition Based Text To Speech Synthesis, Journal of Emerging Technologies and Innovative Research (JETIR), ISSN: 2349-5162, vol. 2, July 2015.
16. Aravind Sasikumar, *Text to Speech Conversion System using OCR*, vol. 5, January 2015
17. Akshay.A, Amrrith.N P, Dwishanth.P and Rekha.V, A Survey on Text to Speech Conversion, International Journal of Trend in Research and Development, vol. 5(2), ISSN: 2394-9333, Mar - Apr 2018.
18. Anusha Joshi, Deepa Chabbi, Suman M, Suprita Kulkarni, *Text to speech system for Kannada language*, IEEE *Xplore*: 12 November 2015.

AUTHORS PROFILE

Pawan S Nadig is a final year Engineering student in Vidyavardhaka College of Engineering Mysuru, majoring in Computer Science. He is currently working as an intern in Informatica Business Solutions Pvt Ltd., Bengaluru. He has worked on projects like Leukemia detection using image processing, Smart parking using IoT and also Radar on roads which is an IoT project for replacement of tollgates in highways.

Pooja G is a final year Engineering student in Vidyavardhaka College of Engineering Mysuru, majoring in Computer Science. She has worked on Machine Learning projects like Sentimental Analysis using twitter data, Credit Card Fraud Detection and Weather Forecast. Web Technology projects such as Placement Automation System and Content Management System. She had also published a Research Paper on A Comparative Study on Different Techniques of Sentimental Analysis in SoCTA conference. Her area of research interests are Machine Learning, Web Design and Development and Big Data. She had also took part in many mini project competitions during Engineering course

Kavya D is a student. She has completed diploma in Computer Science at Vidyavardhaka polytechnic, Mysore under department of technical education and is currently pursuing B.E. in computer science and engineering at Vidyavardhaka College of Engineering, Mysore under Visveswaraya technological university. She has done a project on Brain tumor detection using Matlab and she is currently involved in project on text-to-speech system for Kannada language using OCR technique. she has done internship on Web design and development and also took part in some of the mini projects involved in Engineering course.

R Chaithra has completed Diploma in Computer Science from Vidyavardhaka polytechnic, Mysuru, in 2017 and currently pursuing BE in Computer Science and Engineering from Vidyavardhaka College of Engineering, Mysuru. Her area of research interests are Internet of Things, Android Application Development and Image Processing. She has done project on Real Time Water Quality Monitoring System Using IOT. This is her first publication.

Radhika A D is a BE graduate from Canara Engineering College Mangalore, Mtech Graduate from St. Joseph College of Engineering, Visveswaraya Technological University and is persuing PhD at BMS College of Engineering under Visveswaraya Technological University. She has 6 years of teaching experience in Engineering Education. Her areas of research include Computer Vision, Artificial Intelligence, Machine Learning, Neural Networks, Audio Processing and Data Analytics. She has done projects on Multi Objective evolutionary algorithms for disease prediction, soft computing and Human Activity Recognition. She is currently working as Assistant Professor in Dept. of Computer Science and Engineering at Vidyavardhaka College of Engineering.