Design and Implementation of Konkani Text to Speech Generation System using OCR Technique

John Colaco [*], M.E. (Electronics and Telecommunication),Sangam Borkar [**]

[*] Electronics & Telecommunication Department, Goa College of Engineering
Email: j_7685@yahoo.com

[**] Electronics & Telecommunication Department, Goa College of Engineering
Email: sangam@gec.ac.in

Abstract- India is a country of having multi spoken languages and there is a need in digitization of books and documents and conversion of this text of these books and documents into speech. This paper envisages on design and implementation of Text to Speech system for English and Konkani language using image recognition technology (Optical Character Recognition) and to develop a cost effective text image to speech conversion system using MATLAB. The main aim of this project is to recognize the text character of an image and convert this text into speech signal. To achieve this, text contained in the page is first pre-processed. The pre-processed unit prepares this text for recognition. Then this text is segmented to separate out the characters from each other. After Segmentation, letters are extracted and resized and stored them in the text file. This process is done with the help of MATLAB. This text is then converted into speech. This paper also uses Devanagri script.

Index Terms- Optical Character Recognition, Segmentation, Text processing, Speech synthesis.

I. INTRODUCTION

Optical character recognition (OCR) is the process of translating scanned images of typewritten text into machine-editable information. This technology allows a machine to automatically recognize characters through an optical mechanism. Human beings recognize many objects in this manner our eyes are the "optical mechanism." But while the brain "sees" the input, the ability to comprehend these signals varies in each person according to many factors. By reviewing these variables, we can understand the challenges faced by the technologist developing an OCR system. OCR can recognize both handwritten and printed text. But the performance of OCR is directly dependant on quality of input documents.OCR is designed to process images that consist almost entirely of text. OCRs have wide range of applications in government and business organizations, as well as individual companies and industries. Some of the major applications of OCR include:

(i) Library and office automation,
(ii) Form and bank check processing,
(iii) Document reader systems for the visually impaired.

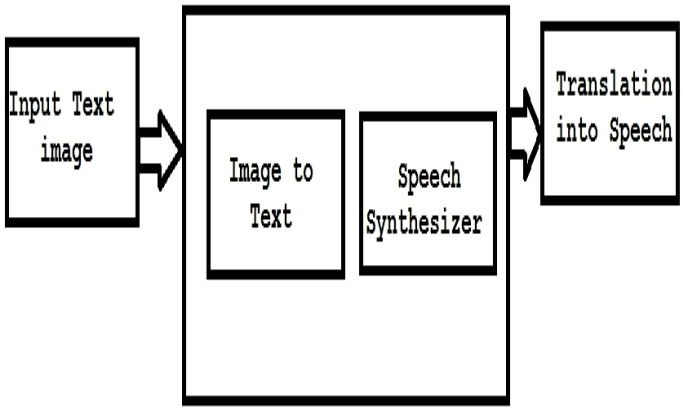


Figure 1.General Block Diagram of Text to Speech using OCR

A. Devanagri Script

Devanagari is a Brahmic script. It has been linked to various other South Asian scripts including Marathi, Gurumukhi, Bengali, and Gujarati. This devanagri script is used for over 120 spoken Indo-Aryan languages, including Hindi and Nepali. It has been also used for writing Classical Sanskrit texts.

The script is written from left to right. Letters suspend from a head stroke is normally constant throughout the length of the word. It is very helpful to become known with the traditional order of Devanagari. The below the table is read as if it were text, left to right and top to bottom.

1	अ	आ	इ	ई	उ	ऊ	ऋ
2	ए	ऐ	ओ	औ	अं	अः	
3	क	ख	ग	घ	ङ		
4	च	छ	ज	झ	ञ		
5	ट	ठ	ड	ढ	ण		
6	त	थ	द	ध	न		
7	प	फ	ब	भ	म		
8	य	र	ल	व			
9	श	ष	स				
10	ह						
11	क़	ख़	ग़	ज़	फ़	ड़	

Figure 2.The traditional order of Devanagari

1	अ	आ	इ	ई	उ	ऊ	ऋ
	a	ā	i	ī	u	ū	ṛ
2	ए	ऐ	ओ	औ	अं	अः	
	e	ai	o	au	aṃ	aḥ	
3	क	ख	ग	घ	ङ		
	ka	kha	ga	gha	ṅa		
4	च	छ	ज	झ	ञ		
	ca	cha	ja	jha	ña		
5	ट	ठ	ड	ढ	ण		
	ṭa	ṭha	ḍa	ḍha	ṇa		
6	त	थ	द	ध	न		
	ta	tha	da	dha	na		
7	प	फ	ब	भ	म		
	pa	pha	ba	bha	ma		
8	य	र	ल	व			
	ya	ra	la	va			
9	श	ष	स				
	śa	ṣa	sa				
10	ह	क्ष	त्र	ज्ञ	श्र		
	ha	kṣa	tra	jña/gya	śra		
11	क़	ख़	ग़	ज़	फ़	ड़	ढ़
	qa	**kha**	ġa	za	fa	ṛa	ṛha

Figure 3.Devangari characters and their transliteration

1) VOWELS

Devanagari script has 18 vowels out of which 11 are mostly used Vowels are transcribed in two distinct forms one is the independent form and other is the dependent (matra) form. The independent form is used when the vowel letter appears single form, at the start of a word, or following another vowel letter. When the vowel follows a consonant, Matras are used.

Independent form	Modifier or *Matras*	Independent form	Modifier or *Matras*
अ	*None*	ए	े
आ	ा	ऐ	ै
इ	ि	ऍ	ॅ
ई	ी	ओ	ो
उ	ु	औ	ौ
ऊ	ू	ऑ	ॉ
ऋ	ृ	ऌ	ॢ
ॠ	ॄ	ॡ	ॣ

Figure 4.Vowels in Devanagri

2) CONSONANTS

Devanagari script has around 33 consonants which are arranged phonetically. The first sets of 25 consonants are called occlusive consonant, and remaining 8 are called non occlusive consonant. The occlusive consonants are further divided into five groups: labials, gutturals, palatals**,** cerebral or retroflex and dentals. The last four consonants in these groups are further divided in two groups: plosive and voiced plosive and the first consonant is the nasal consonant. The plosive and voiced plosive are again divided into aspirated and unaspirated version both having one character each

Semivowels	य	र	ल	व
Sibilants	श	ष	स	
Aspirate	ह			

Figure 5.Non Occlusive Consonants.

	Plosive		Voiced Plosive		Nasal
	Unaspirated	Aspirated	Unaspirated	Aspirated	
Gutturals	क	ख	ग	घ	ङ
Palatals	च	छ	ज	झ	ञ
Cerebals	ट	ठ	ड	ढ	ण
Dentals	त	थ	द	ध	न
Labials	प	फ	ब	भ	म

Figure 6.Occlusive consonant.

3) VOWELS COMBINED WITH CONSONANTS

Vowels in combination with consonants (mātrās) always appear into view with one of the consonants. These examples are shown below:

Table 1. Combination of Vowels with consonants

अ	***a* after a consonant is not shown by any symbol. If there is no vowel sign after a consonant then it is generally pronounced with an *a*,**	क ***ka***
आ	***ā* after a consonant is: ा**	का ***kā***
इ	***i* after a consonant is: ि Note it is pronounced after the consonant but it is written before the consonant**	कि ***ki***
ई	***ī* after a consonant is: ी**	की ***kī***
उ	***u* after a consonant is: ु**	कु ***ku***
ऊ	***ū* after a consonant is: ू**	कू ***kū***

B. Speech synthesis

Speech synthesis is the non-natural creation of human speech. For this purpose a computer system used is known as speech synthesizer, and this synthesizer can be implemented in hardware as well as in software. An ordinary language text into speech has been converted by text-to-speech (TTS) system and other systems make symbolic linguistic demonstration like phonetic transcriptions into speech. Synthesized speech is normally created by concatenating pieces of recorded speech which are earlier stored in a database. A synthesizer can integrate a representation of the vocal region and other human voice characteristics to produce a "synthetic" voice output. The quality of a speech synthesizer is judged by its comparison to the human voice, and by its ability to be understood. This text-to-speech program allows people with visual impairments or interpretation disabilities to listen to written works on a home computer.

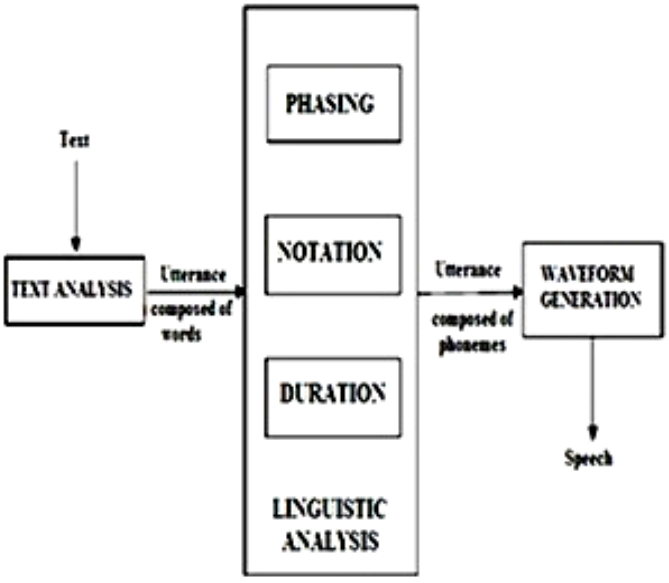


Figure 7.Text Processing

A text-to-speech system is composed of two parts known as front-end and back-end. The front-end consist of two tasks. First task is that, it converts raw text having symbols like abbreviations and numbers into the equivalent of written-out words. This process is called pre-processing text normalization, or tokenization. The front-end then allots phonetic transcriptions to each word, and breaks up and marks the text into prosodic units, like sentences, clauses, and phrases. The process of assigning phonetic transcriptions to words is known as text-to-phoneme or grapheme-to-phoneme conversion. Phonetic transcriptions and Prosody information together make up the symbolic linguistic representation which is output by the front-end. The synthesizer is often known to as the back-end, and then converts the symbolic linguistic representation into necessary sound.

II. LITERATURE SURVEY

George Nagy, Prateek Sarkar 2004 proposed four methods of converting paper documents to computer readable form which are then compared with hypothetical labor cost such as keyboarding, omnifont OCR, style adaptive OCR and style specific OCR. The best choice was to determined by i) the reject rates of various OCR systems, ii) the fraction of the material and iii) the cost of partitioning the material with respect to style. This approach works the best in guidance of fellow researchers. In this the authors continuously receives or asks inputs from their fellows. It enriches the information pool of your paper with expert comments or up gradations. And the researcher feels confident about their work and takes a jump to start the paper writing.

G.Vamvakas, B.Gatos, N. Stamatopoulos, and S.J.Perantonis, 2008 proposed a Complete Optical Character Recognition Methodology for Historical Documents either handwritten or printed without any idea of the font. In this they have produced a database for training for a set of documents and then recognition of new document images. A clustering scheme is adapted for grouping the characters of alike shape.

Raghuraj Singh, C. S. Yadav, Prabhat Verma, Vibhash Yadav 2010 proposed Optical Character Recognition (OCR) for Printed Devnagari Script Using Artificial Neural Network technique. In this paper, they have proposed a technique for OCR System of various fonts and sizes of printed Devnagari script. The

recognition rate of the proposed OCR system has been found high.

Neeraj Pratap and Dr. Shwetank Arya 2012 proposed a Devnagari Character Recognition from Past to Future. In this Paper the drawbacks of different methodologies is defined on On-line or off-line and hand written text or machine printed text.

Ravina Mithe, Supriya Indalkar, Nilam Divekar 2013 proposed mobile application of Optical Character Recognition. Using android phones they have developed user friendly application which converts image into speech.

Nisha Goyal1, Er. Shilpa Jain 2015 proposed Optimized Hindi Script Recognition using OCR Feature Extraction Technique. In this paper they have used technique that extracts the feature of individual Hindi character. But one issue they have found this OCR does not recognize true form of consonants or half characters.

Sangramsing Kayte, Monica Mundada, Dr. Charansing Kayte 2015 proposed a Marathi Text-To-Speech Synthesis by using Natural Language Processing technique. In this paper they have developed a rule based text- to- speech synthesis system which is a combination of formant and concatenation techniques with good naturalness. For the good quality of their output speech, a subjective test was performed.

N.Swetha and K.Auradha 2013 proposed a Conversion of text to Speech. In this they have created a database required for character to voice conversion and it is recorded in the form of wave files (.wav). Then in converting text to speech they have created a text file. After creating text file, it is opened and read in matlab. In matlab all the data is stored in the form of a matrix. For every element read, corresponding wave file is played to output the sound of that character.

III. METHOLODOGY

In this work there is conversion of the image into text and then text into speech using MATLAB. MATLAB is a data analysis and visualization tool designed to make matrix manipulation as simple as possible. In addition, it has powerful graphics capabilities and its own programming language. MATLAB supports reading all of the common image formats.

A. Algorithm

Step 1:- Read the image by using imread command.

Step 2:-. In this step convert the image into gray scale by using rgb2gray command. This command converts RGB image into grayscale image by eliminating saturation and hue information while retaining the luminance. Then convert this gray image into black and white image. This is done by counting the threshold in gray image then according to that threshold converts it into black and white image. Then remove all the objects from less than 30 pixels by using bwareaopen command. This will morphologically open binary image (remove small objects).

Step 3:- In this step convert the black and white image into word matrix for further calculation.

Step 4:- In this step open the text.txt as file for write by using fopen command.

Step 5:- In this step load the model templates for matching the letters with the templates.

Step 6:- In this step compute the number of letters in template file by using the loop. In this loop firstly lines are separated from the text.

Step 7:- In this step label and count connected components by using bwlabel command.

Step 8:- In this step extract letter and resize letter by using imresize command. Resize the letters according to templates size.

Step 9:-In this step write in the text file and concatenate the letters by using the word matrix and use the fprintf command. fprintf write formatted data to file.

Step 10:- In this step analysis the text first, and install Microsoft Win 32 SAPI.

Step 12:- In this step, Upload Text file and convert this text file into .wav file.

Step 13:- In this step for converting the text into speech initialize the wave player.

Step 14:- Finally the speech for given image is generated.

B. Design of Proposed OCR System

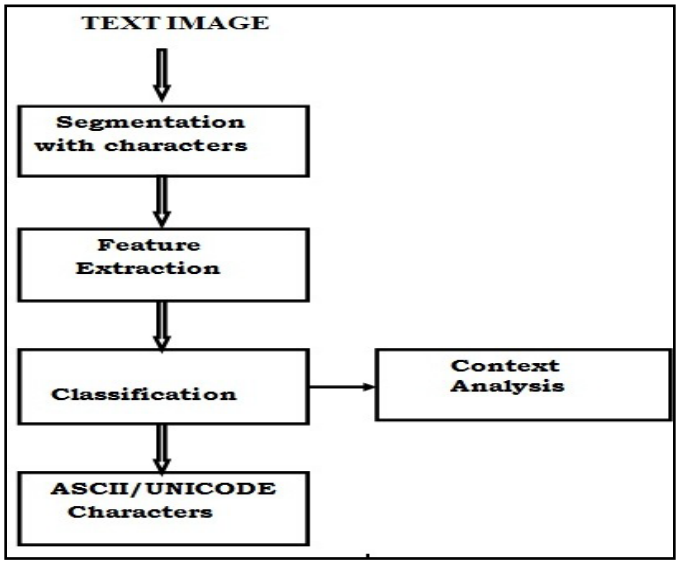


Figure 7. Stages in the design of OCR.

Following steps have been used in the design of proposed OCR system:

1) Preprocessing
2) Segmentation
3) Feature Extraction
4) Classification.

1) Preprocessing

In the proposed OCR system, digitization of text is done by using an optical scanner. The digitized images are typically in gray tone, and for a clear document, an easy histogram based threshold approach is used for converting them into two tone images. The histogram of gray values of the pixels shows two prominent peaks, and a middle gray value located between the peaks is a good choice for threshold. For salt and pepper noise median filter is used. Median filter puts back the value of a pixel by the median of gray levels in the neighborhood of that pixel. Median filters provide excellent noise reduction capabilities, with less blurring than linear smoothing filters of similar size.

2) Segmentation

One of the most important phases of OCR system is segmentation. Good segmentation technique increases the performance of OCR. Segmentation divides an image into its constituent objects or regions. Basically in segmentation, there is an extraction of basic constituents of the script, which are certainly characters. Consider an example as shown in figure.8 Devanagari script text word may be partitioned into three zones. The upper zone represents the portion above the headline, the middle zone covers the portion of basic and compound characters below the headline, and

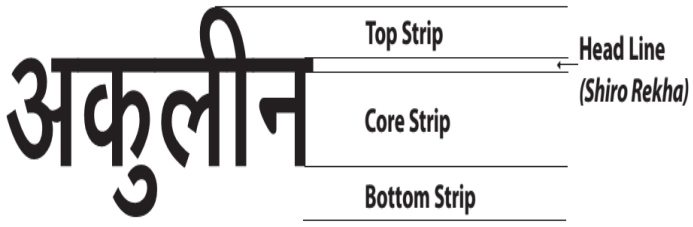


Figure 8. Three Part of Devanagri Word

the lower zone contain where some consonant and vowel modifiers can reside. For a long number of characters of both basic as well as compound there exists a horizontal line at the upper part called “shirorekha” or headline in Hindi. Then imaginary line separating the middle and lower zone may be called the base line.

Line, Word and Character Segmentation: Once the text blocks are detected, the OCR system automatically finds individual text lines, segments the words, and then separates the characters accurately.

Segmentation of Line: Text lines are detected by horizontal scanning. For segmentation of line, scanned document page is scanned horizontally from the top first. Then the last row is containing all white pixels, before a black pixel is found. Then the first row containing entire white pixel just after the end of black pixels is found. This process is repeated on entire page.

Segmentation of Words: After finding a particular line individual words are separated. This is done by vertical scanning.

Segmentation of Individual Characters: Once the words are segmented then individual characters are segmented. Before segmenting words to individual characters, the head line is located. This is done by finding the rows having maximum number of black pixels in a word. After locating head line, convert it in white pixels. After removing head line, word is divided into three horizontal parts known as upper zone, middle zone and lower zone. Individual characters are separated from each zone by applying vertical scanning.

3) Feature Extraction

Selection and Feature extraction can be defined as extracting the most diplomat information from the raw data, which reduces the within class pattern variability while enhancing between class pattern variability. For this reason, a set of features are extracted for each. Various feature extraction methods are classified in three major groups:

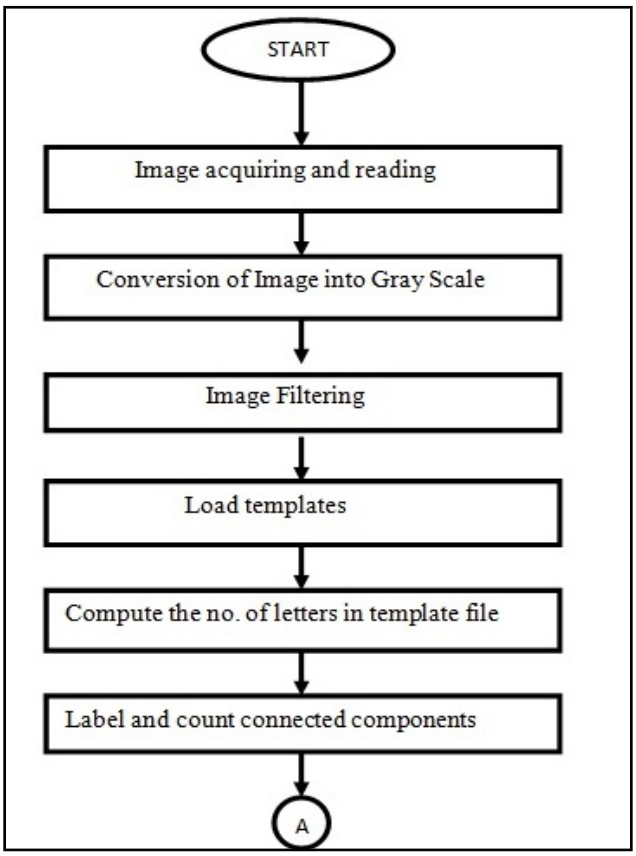
i. Global Transformation and Series Expansion.
ii. Statistical Features.
iii. Geometrical and Topological Features

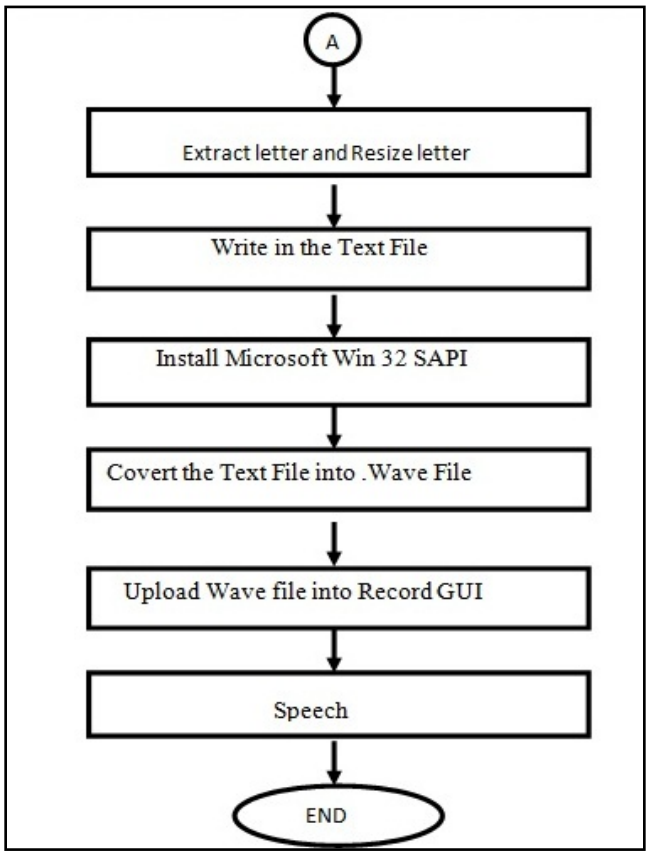
4) Classification

The classification is the process of identifying each character and assigning it to the correct character class. Classification performed is based on the extracted features. For classification and recognition, Artificial Neural Network (ANN) approach has been used. It is a computational model used in a condition where the problem is tough and data may subject to statistical variation. The structural design of a neural network determines how a neural network transfers its input into output. This transfer can be viewed as a computation.

C. Flowchart

Below flowchart illustrates the conversion of text image into text file and then converting it into speech.





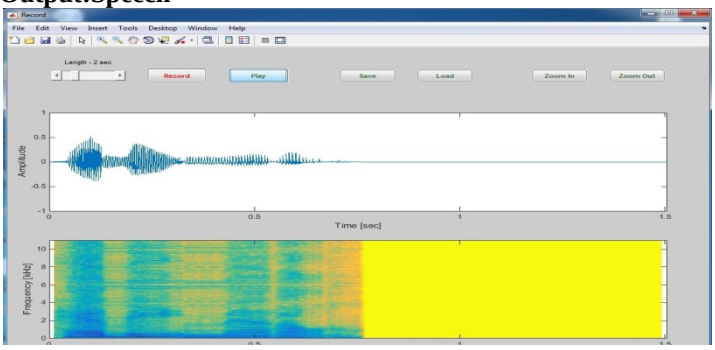
IV. RESULT

A. Analysis of different images

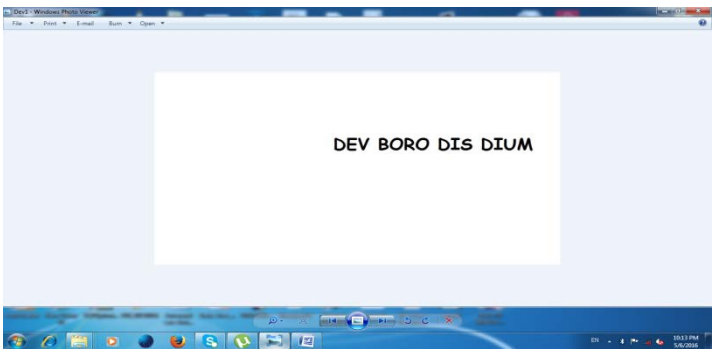
Input image 1(JPEG)

Output: Text

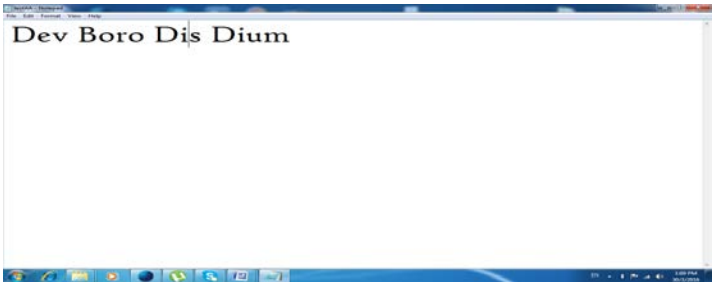
Output:Speech



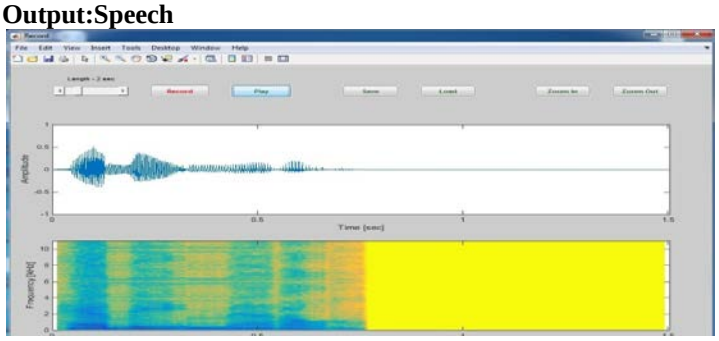
Input image 2(JPEG)



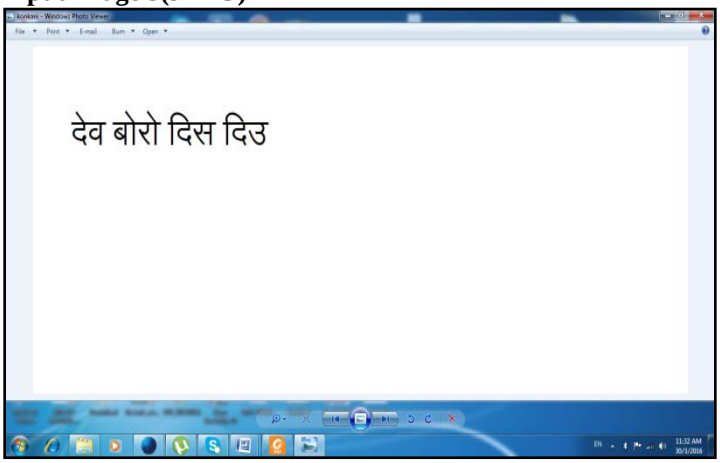
Output: Text



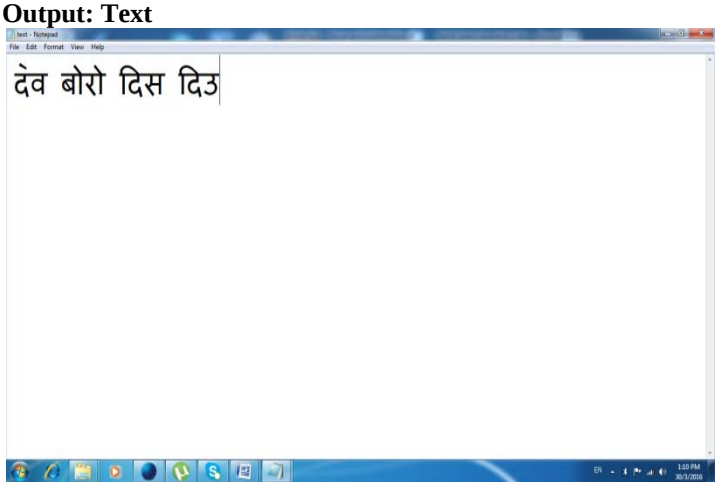
Output:Speech



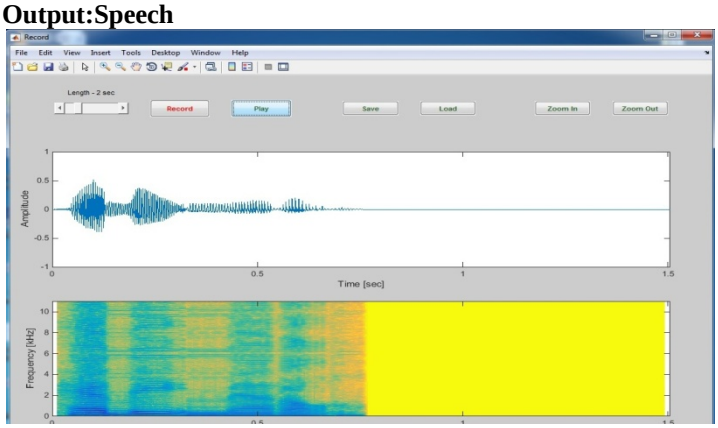
Input image 3(JPEG)



Output: Text



Output:Speech



V. CONCLUSION

This paper gives an approach for converting text image of both Devanagri and Roman Script into readable text and then converting this text into speech using optical character recognition and text to speech technology. By this approach People with poor vision or visual dyslexia or totally blindness goan people will able to read text loud. This approach will also help in reading roman script document.

VI. ACKNOWLEDGMENT

I am **Dr.H.G. Virani,** Professor, Head of department of Electronics and Telecommunication engineering department, GEC, Goa for giving me, an encouragement to work providing me his kind co-operation in enriching me in various roles and providing me all necessary facilities to work in the lab.

VII. REFERENCES

[1] Jaehwa Park, Govindaraju, V. Srihari, S.N, "*OCR in a hierarchical feature space*," Pattern Analysis and Machine Intelligence, IEEE Transactions on , vol.22, no.4, pp.400-407, Apr 2000.

[2] G.Vamvakas, B.Gatos, N. Stamatopoulos, and S.J.Perantonis, "*A Complete Optical Character Recognition Methodology for Historical Documents,*" Document Analysis System, 2008. Proceedings. The Eighth IAPR Workshop on Document Analysis Systems, 2008.

[3] Raghuraj Singh, C. S. Yadav, Prabhat Verma, Vibhash Yadav, " *Optical Character Recognition (OCR) for Printed Devnagari Script Using Artificial Neural Network*," International Journal of Computer Science & Communication, Vol. 1, No. 1, pp. 91-95, January-June 2010.

[4] Neeraj Pratap and Dr. Shwetank Arya., "*A Review of Devnagari Character Recognition from Past to Future*," International Journal of Computer Science and Telecommunications, Vol. 3, Issue 6, June 2012.

[5] Ravina Mithe, Supriya Indalkar, Nilam Divekar, "*Optical Character Recognition,*" International Journal of Recent Technology and Engineering (IJRTE), Vol.-2, Issue 1, March 2013.

[6] Nisha Goyal1, Er. Shilpa, "*Optimized Hindi Script Recognition using OCR Feature Extraction Technique*", International Journal of Advanced Research in Computer and Communication Engineering, Vol. 4, Issue 8, August 2015.

[7] Sangramsing N. Kayte1, Monica Mundada1, Dr. Charansing N. Kayte2, Dr.Bharti Gawali, "*Implementation of Text To Speech for Marathi Language Using Transcriptions Concept*". IOSR Journal of VLSI and Signal Processing (IOSR-JVSP) Volume 5, Issue 6, PP 63-67, Ver. I (Nov -Dec. 2015).

[8] Mr.S.D.Shirbahadurkar, Dr.D.S.Bormane, "*Speech Synthesizer Using Concatenative Synthesis Strategy for Marathi language (Spoken in*

Maharashtra, India)". International Journal of Recent Trends in Engineering, Vol.2, No. 4, November 2009.

[9] Mrs. S. D. Suryawanshi, Mrs. R. R. Itkarkar, Mr. D. T. Mane 2014, "*High Quality Text to Speech Synthesizer using Phonetic Integration*", International Journal of Advanced Research in Electronics and Communication Engineering (IJARECE) Vol. 3, Issue 2, February 2014.

[10] N.Swetha and K.Auradha, "*Text to Speech Conversion*", International Journal of Advanced Trends in Computer Science and Engineering, Vol.2, No.6, pp.269-278.

VII. AUTHORS

First Author – John colaco, M.E. (ECI) Goa Engineering college, Farmagudi, Goa and j_7685@yahoo.com.

Second Author – Prof. Sangam Borkar, M.E. (Electronics), Goa Engineering College, Farmagudi, Goa and sangam@gec.ac.in.