OPTICAL CHARACTER RECOGNITION BASED TEXT TO SPEECH SYNTHESIS

[1]Yamini D. Patil, [2] Nandkishor C. Patil
[1]Student, [2] Professor
[1]Electronics & Telecommunication Department
[1]P.S.G.V.P Mandal's D.N.Patel College of Engineering, Shahada, India

***Abstract*-Text to speech, there are many systems which convert normal language text in to speech. This thesis aims to study on speech synthesis technology using image recognition technology (Optical Character Recognition) to develop a cost effective user friendly image to speech conversion system using MATLAB for blind person. In this work we tried to make a system by which we can get the text through image and then speech through that text using MATLAB. The primary motivations are to provide users with a friendly vocal interface with the computer and to allow people with certain handicaps (such as blindness, dumbness, poor vision, visual dyslexia) to use the computer or to read any type of documents.**

***Key words* - OCR, MATLAB, speech synthesis, speech conversion, character extraction algorithm.**

INTRODUCTION

The typical support instrument for the visually impaired is the "white cane", which is used for the detection of obstacles by providing a limited degree of feedback to the user. The portable indoor navigation aid alerts the user to the layout of an entire indoor area upon entrance so as to assist in collision avoidance. Today in markets number of navigation aid are available for Blind persons.

Our main intension is to provide a system through which blind person can get the knowledge of words they cant see using Optical character recognition & speech synthesis technique.

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[2]. 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.

First, if we read a page in a language other than our own, we may recognize the various characters, but be unable to recognize words. However, on the same page, we are usually able to interpret numerical statements - the symbols for numbers are universally used. This explains why many OCR systems recognize numbers only, while relatively few understand the full alphanumeric character range.

Second, there is similarity between many numerical and alphabetical symbol shapes. For example, while examining a string of characters combining letters and numbers, there is very little visible difference between a capital letter "O" and the numeral "0." As humans, we can re-read the sentence or entire paragraph to help us determine the accurate meaning.

This procedure, however, is much more difficult for a machine. Third, we rely on contrast to help us recognize characters. We may find it very difficult to read text which appears against a very dark background, or is printed over other words or graphics. Again, programming a system to interpret only the relevant data and disregard the rest is a difficult task for OCR engineers.

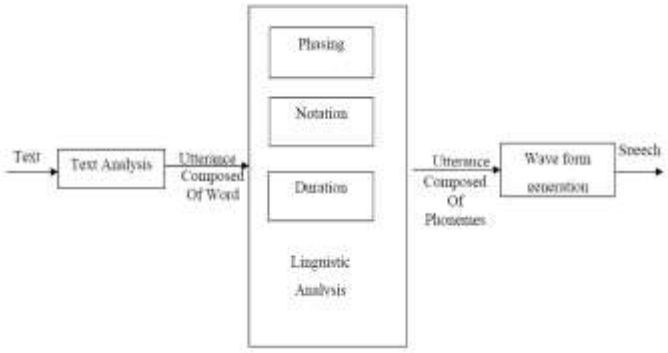


Fig 1: Text processing

II. RELATED WORK

For Text to Speech

Speech synthesis is the artificial production of human speech. A computer system used for this purpose is called a speech synthesizer, and can be implemented in software or hardware. A text-to-speech (TTS) system converts normal language text into speech; other systems render symbolic linguistic representations like phonetic transcriptions into speech. Soumyajit Dey et.al 2007 proposed architectural optimizations for text to speech synthesis in embedded system. The increasing processing power of embedded devices has created the scope for certain applications that could previously be executed in desktop environments only, to migrate into handheld platforms. An important feature of the computing systems of modern times is their support for applications that interact with the user by synthesizing natural speech output. In their work, the performance of a Text to Speech Synthesis application is evaluated on embedded processor architectures and medications in the underlying hardware platform are proposed for real time performance improvement of the concerned application [8].

For OCR

Shunji Mori, Ching Y. Suen et.al 1992 considered research and development of OCR systems from U historical point of view. They divided their work into two parts: the research and development of OCR systems, and the historical development of commercial OCR's. The R&D part was further divided into two approaches: template matching and structure analysis. It had been shown that both approaches are coming closer and closer to each other and it seemed they tend to merge into one big stream open problems are also raised in their work [4-7].

Jianli Liu, Nugent et.al 1993 developed a new AI-based OCR post processing technique, implemented as an intelligent OCR Editor (IOCRED), which could enable the automation of OCR post processing procedure and, therefore could result in the increase of throughput, the decrease of error rate and the reduction of cost per page of an OCR system. IOCRED was a novel AI approach to automating OCR post-processing procedures. The IOCRED concept was based on the premise that different OCR algorithms have distinct error characteristics and such distinction enable a cognitive device to automate its error detection and correction process. IOCRED system should be able to achieve a high throughput, low error rate and low cost OCR conversion. The utilization of the IOCRED technique could result in the removal or the reduction of the current OCR post processing techniques, error rate and cost per page [8].

Shaolei Feng and R. Manmatha 2006 have proposed a hierarchical, HMM- based automatic evaluation of OCR accuracy for digital library of books. They proposed a Hidden Markov Model (HMM) based hierarchical alignment algorithm to align OCR output and the ground truth for books. They believed this was the first work to automatically align a whole book without using any book structure information. The alignment process worked by breaking up the problem of aligning two long sequences into the problem of aligning many smaller subsequences. This can be rapidly and effectively done. Experimental results showed that their hierarchical alignment approach works very well even if OCR output has a high recognition error rate. Finally, they evaluate the performance of a commercial OCR engine over a large dataset of books based on the alignment results [1]

III. TECHNIQUE

OCR Methodology

Figure 2 shows the framework of OCR. Most of the designs in OCR follow a modification of this architecture. Given a page for recognition, first it is preprocessed. The aim of the preprocessing module is to prepare the image for recognition. Preprocessing involves binarization, skew correction and normalization. It undergoes some image enhancements such as filtering out noise and increasing the contrast. Then, the image is segmented to separate the characters from each other. Segmentation occurs at two levels.On the first level, text, graphics and other parts are separated. On the second level, text lines, words and characters in the image are located. Information from connected component analysis and projection analysis can be used to assist text segmentation. Segmentation is followed by feature extraction, which is concerned with the representation of the object.

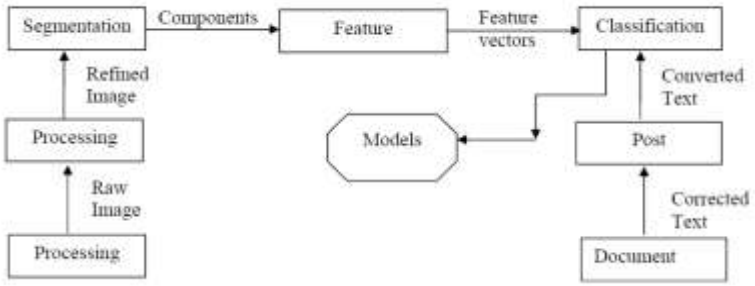
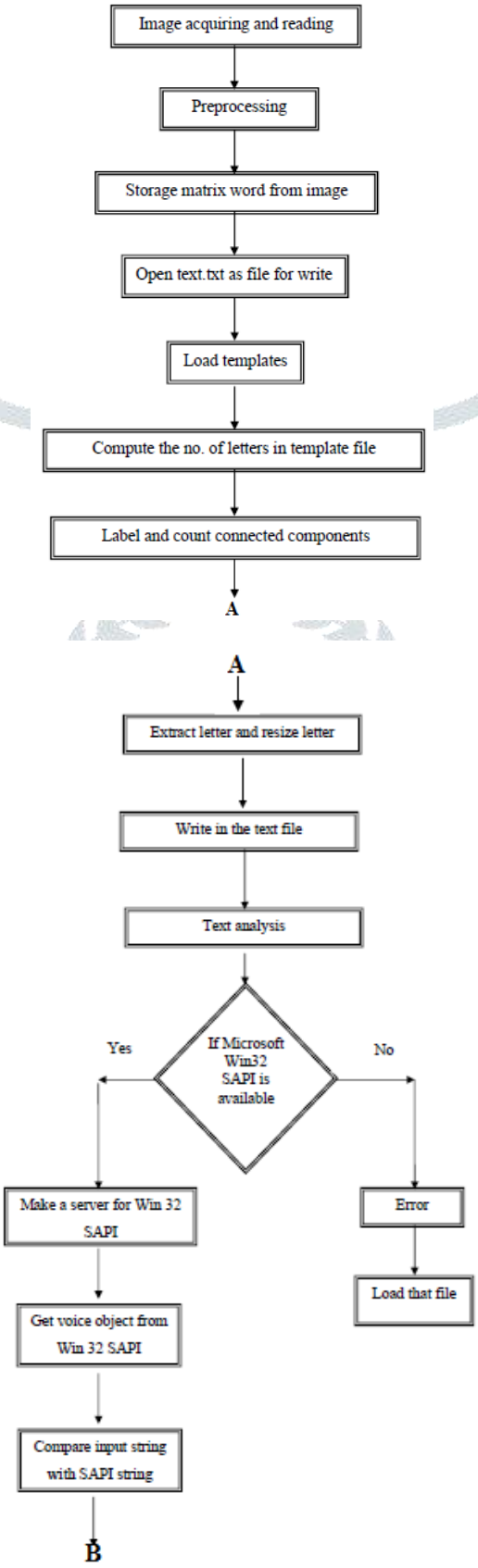


Fig. 2: Overview of OCR design.

3.2 Flowchart Of Given Methodology



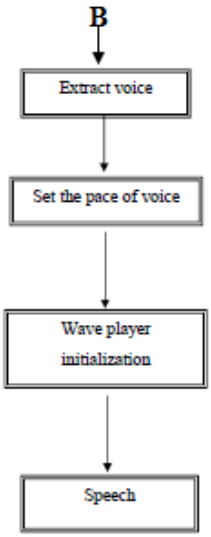


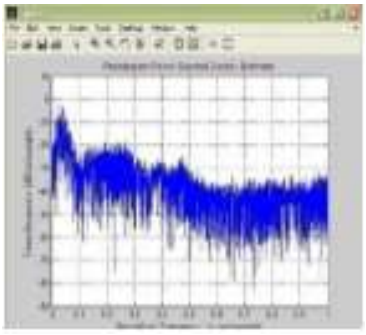
Fig. 3 Flow chart of used methodology.

IV. RESULTS & DISCUSSION

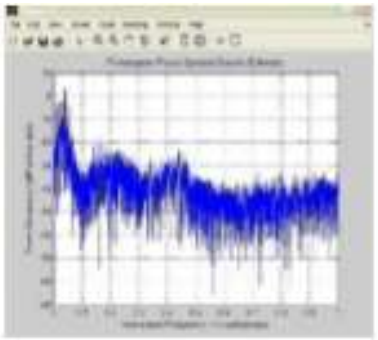
we gave the input image in which THAPAR UNIVERSITY PATIALA was written, then it is converted into text line-wise and after it the text is converted into speech, again line-wise which is shown according to text in periodogram.

Fig. 4 : Input Image

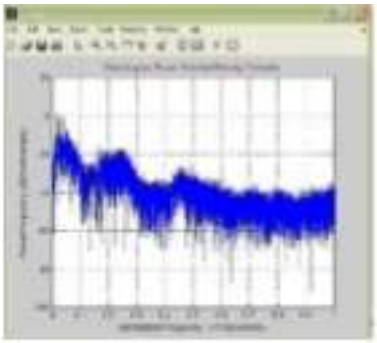
Fig.5 :Output text



THAPAR



UNIVERSITY



PATIALA

Fig. 6: Periodogram of Sound wave

V. CONCLUSION

Image into text using OCR and then that text into speech by speech synthesis is done by MATLAB. For image to text conversion firstly image is converted into gray image then black and white image and then it is converted into text by MATLAB. Microsoft Win 32 SAPI library has been used to build speech enabled applications, which retrieve the voice and audio output information available for computer. This library allows selecting the voice and audio device one would like to use. By MATLAB we can select the voices from the list and can change the pace and volume, which can be listen by installing wave player in the MATLAB. This method can save time by allowing the user to listen background materials while performing other tasks.

REFERENCES :

[1] Ainsworth, W., "A system for converting English text into speech," Audio and Electroacoustics, IEEE Transactions on , vol.21, no.3, pp. 288-290.
[2] Fushikida, Katsunobu; Mitome, Yukio; Inoue, Yuji, "A Text to Speech Synthesizerfor the Personal Computer," Consumer Electronics, IEEE Transactions on ,vol.CE-28, no.3, pp.250-256.
[3] Hertz, S., "English text to speech conversion with delta," Acoustics, Speech, and Signal Processing, IEEE International Conference on ICASSP '86.,vol.11, no., pp.2427-2430.
[4] Lynch, M.R.; Rayner, P.J., "Optical character recognition using a new connectionist model," Image Processing and its Applications, 1989,Third International Conference on, vol., no., pp.63-67.
[5] Malyan, R.R.; Sunthankar, S.; Teranchi, H.; Yeghiazarian, A., "Perception of multi author hand printed text," Character Recognition and Applications, IEE Colloquiumon , vol., no., pp.8/1-8/3.
[6] Curtis, K.M.; Race, P.; Aziz, A.A., "Parallelism and the transputer in the automatic translation of text to speech," Acoustics, Speech, and Signal Processing.ICASSP-89., 1989 International Conference on , vol., no., pp.809-811 vol.2, 23-26 .
[7] Jianli Liu; Nugent, J.H.; Bowen, D.G.; Bowen, J.E., "Intelligent OCR editor,"Electrical and Computer Engineering.