ISSN (ONLINE): 2454-9762
ISSN (PRINT): 2454-9762
Available online at www.ijarmate.com

International Journal of Advanced Research in Management, Architecture, Technology and Engineering (IJARMATE) Vol. 3, Special Issue 6, March 2017

OPTICAL CHARACTER BASED TTS SYNTHESIZER USING MYRIO

B.Aravind Balaji
PG Scholar
Department of Electronics And communication Engineering
Gojan School of Business and Technology
Chennai, Tamil Nadu
balajiaravi@gmail.com

R.Raj Mohan
Associate Professor
Department of Electronics And communication Engineering
Gojan School of Business and Technology
Chennai, Tamil Nadu
rajmohan.r@gojaneducation.com

***Abstract* - Knowledge extraction by just Reading books is a distinctive property. Although text can be a medium of communication but speech signal is more effective means of communication than text. In this paper work Optical Character recognition (OCR) Based Speech Synthesis System has been discussed using myRIO under LabVIEW platform as front end .Although lot of work has been done in the field of OCR and Speech Synthesis individually, but it is first OCR based Speech Synthesis System using FPGA implementation in myrio. Each page of the book to be read are been turned using automatic page turner and the text is been captured using a camera, and electronic translation of the captured image is been done by myRio using OCR technique and the recognized text is been converted to speech using ACTIVEX speech synthesizer.**

Index Terms – myRIO, OCR, LabVIEW , Text To Speech, ACTIVEX.

I. INTRODUCTION

Machine replication of human functions, like reading, is an ancient dream. However, over the last five decades, machine reading has grown from a dream to reality. Character recognition or optical character recognition (OCR), is the process of converting captured images of machine printed or handwritten text (numerals, letters, and symbols), into a computer format text (such as ASCII). Optical character recognition has become one of the most successful applications of technology in the field of pattern recognition and artificial intelligence. Many commercial systems for performing OCR exist for a variety of applications. Speech is probably the most efficient medium for communication between humans. A Text-To-Speech (TTS) synthesizer is a computer-based system that should be able to read any text aloud, whether it was directly introduced in the computer by an operator or scanned and submitted to an Optical Character Recognition (OCR) system.

II. INTRODUCTION TO MYRIO

The NI myRIO embedded device features a 667 MHz dual-core ARM Cortex-A9 programmable processor and a customizable Xilinx field programmable gate array (FPGA) that students can use to start developing systems and solve complicated design problems faster—all in a sleek and simple enclosure with a compact form factor. The NI myRIO device features the Zynq-7010 All Programmable system on a chip (SoC) to unleash the power of NI LabVIEW system design software both in a real-time (RT) application and on the FPGA level.

Figure1.1: myRIO

NI myRIO shown in figure: 1.1 is a reconfigurable and reusable teaching tool that helps students learn a wide variety of engineering concepts as well as complete design projects. The RT and FPGA capabilities along with onboard memory and built-in WiFi allow deploying applications remotely and running them "headlessly" without a remote computer connection. Forty digital I/O lines overall with support for SPI, PWM out, quadrature encoder input, UART, and I2C; eight single-ended analog inputs; two differential analog inputs; four single-ended analog outputs; and two ground-referenced analog outputs allow for connectivity to countless sensors and devices and programmatic control of systems. All of this

ISSN (ONLINE): 2454-9762
ISSN (PRINT): 2454-9762
Available online at www.ijarmate.com

International Journal of Advanced Research in Management, Architecture, Technology and Engineering (IJARMATE) Vol. 3, Special Issue 6, March 2017

functionality is built in and preconfigured in the default FPGA functionality, which eliminates the need for expansion boards or "shields" to add utility.

III. INTRODUCTION TO LABVIEW

LabVIEW, short for Laboratory Virtual Instrument Engineering Workbench, is a programming environment in which you create programs using a graphical notation (connecting functional nodes via wires through which data flows); in this regard, it differs from traditional programming languages like C, C++, or Java, in which you program with text. However, LabVIEW is much more than a programming language. It is an interactive program development and execution system designed for people, like scientists and engineers, who need to program as part of their jobs. The LabVIEW development environment works on computers running Windows, Mac OS X, or Linux. LabVIEW can create programs that run on those platforms, as well as Microsoft Pocket PC, Microsoft Windows CE, Palm OS, and a variety of embedded platforms, including Field Programmable Gate Arrays (FPGAs), Digital Signal Processors (DSPs), and microprocessors.

IV.OCR (Optical Character Recognition))

Optical character recognition, usually abbreviated to OCR, is the mechanical or electronic translation of images of handwritten, typewritten or printed text (usually captured by a scanner or camera) into machine-editable text. Optical character recognition belongs to the family of techniques performing pattern recognition using labview figure: 1 given below shows the basic block diagram.

Figure4.1: OCR Block diagram

The image acquisition is done using Vision Acquisition using IMAQ and then acquired image is been pre processed in such a way that the image are been converted from gray scale image (0-255) to (0, 1) as binary code depending upon the intensity level. and the segmentation is been done words, character wish in order to detected the exact pattern of each character to compare with the database and the detected pattern is been matched with the database which is been created during the training using the vision acquisition user block in the labview thus the patterns similar to the data base are been recognized and displayed the figure given below shows the Block diagram of labview.

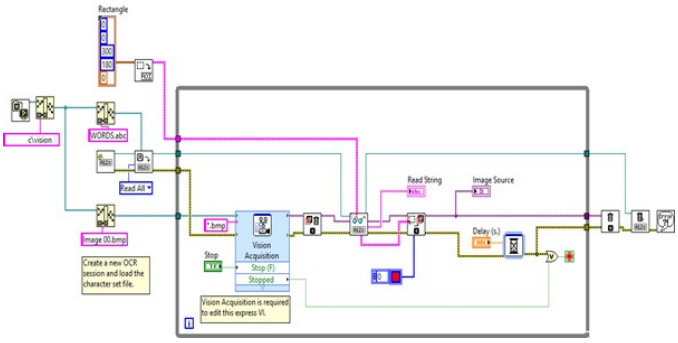


Figure4.2:labVIEW block diagram

A.Image acquisition

A camera is a device that optically scans images, printed text, handwriting, or an object, and converts it to a digital image. In this paper mobile camera is been used as a web camera using ip camera application in which the image are been sent to labview through wifi and saved as BMP images

B. Binarization (pre-processing)

With the advancement of technology and widespread use of colour and gray scale scanners, most images scanned now are grayscale. The reasons for not using colour images are the non-colour nature of some texts such as books, the long time needed for scanning, and the large volume needed for storing color images and lack of appropriate methods for segmentation of colour images. On the contrary, because of the complexity of the OCR operation, the input of the character recognition phase in most methods is binary images. Therefore, in the preprocessing phase, grayscale images are to be converted to binary images. The most common method is using a threshold. In this method, the pixels lighter than the threshold are turned to white and the remainder to black pixels. An important point to notice in here is to determine the threshold. In some methods in which the used pictures are very similar to each other, a fixed threshold is used. So binarization is the process of converting a grayscale image (0 to 255 pixel values) into binary image (0 to1 pixel values) by thresholding. The binary document image allows the use of fast binary arithmetic during processing, and also requires less space to store.

D. Segmentation Process

Segmentation of text is a process by which the text is partitioned into its coherent parts. The text image contains a

ISSN (ONLINE): 2454-9762
ISSN (PRINT): 2454-9762
Available online at www.ijarmate.com

International Journal of Advanced Research in Management, Architecture, Technology and Engineering (IJARMATE) Vol. 3, Special Issue 6, March 2017

number of text lines. Each line again contains a number of words. Each word may contain a number of characters. The following segmentation scheme is proposed where lines are segmented then words and finally characters. These are then put together to the effect of recognition of individual characters. The individual characters in a word are isolated. Spacing between the characters can be used for segmentation.
Line segmentation is the process of identifying lines in a given image. Steps for the line Segmentation is as follows
1. Scan the BMP image horizontally to find first ON pixel and remember that y coordinate as y1.
2. Continue scanning the BMP image then we would find lots of ON pixel since the characters would have started.
3. Finally we get the first OFF pixel and remember that y coordinate as y2.
4. y1 to y2 is the line.
5. Repeat the above steps till the end of the image.
As it's known that there is a distance between one word to another word. This concept will be use here for word segmentation. After the line segmentation scan the image vertically for word segmentation. Steps for the word Segmentation is as follows
1. Scan the BMP image vertically for the recognized line segment, to find first
ON pixel and remember that x coordinate as x1. Treat this as starting coordinate for the word.
2. Continue scanning the BMP image then we would find lots of ON pixel since the word would have started.
3. Finally we get the successive five (this is assumed word distance) OFF pixel column and remember that x coordinate as x2.
4. x1 to x2 is the word.
5. Repeat the above steps till the end of the line segment.

E. Template-Matching and Correlation Techniques.

These techniques are different from the others in that no features are actually extracted. Instead the matrix containing the image of the input character is directly matched with a set of prototype characters representing each possible class. The distance between the pattern and each prototype is computed, and the class of the prototype giving the best match is assigned to the pattern. The technique is simple and easy to implement in hardware and has been used in many commercial OCR machines. However, this technique is sensitive to noise and style variations and has no way of handling rotated characters..

F. IMAQ

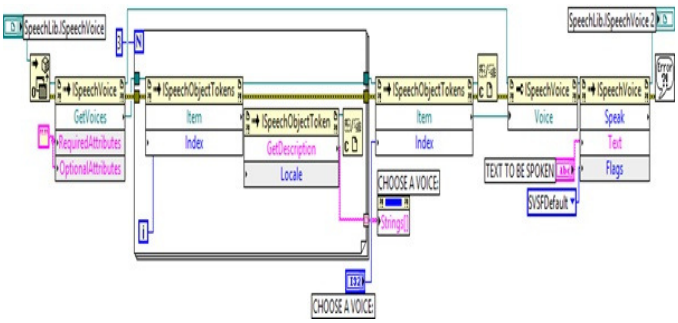
IMAQ Vision for LabVIEW—a part of the Vision Development Module—is a library of LabVIEW VIs that you can use to develop machine vision and scientific imaging applications. The Vision Development Module also includes the same imaging functions for LabWindows™/CVI™ and other C development environments, as well as ActiveX controls for Visual Basic. Vision Assistant, another Vision Development Module software product, enables you to prototype your application strategy quickly without having to do any programming. Additionally, NI offers Vision Builder AI: configurable machine vision software that you can use to prototype, benchmark, and deploy applications.

V. TTS (TEXT TO SPEECH)

A text to speech (TTS) synthesizer is a system that can read text aloud automatically, which is extracted from Optical Character Recognition (OCR). A speech synthesizer can be

Figure5.1: TTS block Diagram

implemented by both hardware and software. Speech synthesis is the artificial production of human speech. A computer system used for this purpose is called a speech synthesizer. A text-to-speech (TTS) system converts normal language text into speech. A synthesizer can incorporate a model of the vocal tract and other human voice characteristics to create a completely "synthetic" voice output. A text to speech processor consists of two parts. The front part performs two major tasks. First, it converts the text containing the words and numbers into the equivalent of written out words. This is called as text normalization or tokenization. The front part assigns phonetics transcriptions (it is the visual representation of the speed sounds) to each word. The process of assigning phonetic transcriptions to words is called text-to-phoneme conversion.



ISSN (ONLINE): 2454-9762
ISSN (PRINT): 2454-9762
Available online at www.ijarmate.com

International Journal of Advanced Research in Management, Architecture, Technology and Engineering (IJARMATE) Vol. 3, Special Issue 6, March 2017

Figure 5.2: Labview TTS Block diagram

Text analysis is used to analysis the text in order to compare the text with the Microsoft ActiveX SDK application. In order to use the application the application is invoked first then the text is been compared with the voicelib file in order to get the exact voice for the text then the char voice is been recognized and play back is been done using the laptop speaker. The figure given below shows the block diagram of the text to speech synthesizer.

A. Speech Synthesis

Speech synthesis is the artificial production of human speech. Synthesizing is the very effective process of generating speech waveforms using machines based on the phonetical transcription of the message. Recent progress in speech synthesis has produced synthesizers with very high intelligibility but the sound quality and naturalness still remains a major problem.

B. Speech production

Continuous speech is a set of complicated audio signals which makes producing them artificially difficult. Speech signals are usually considered as voiced or unvoiced, but in some cases they are something between these two. Voiced sounds consist of fundamental frequency (F0) and its harmonic components produced by vocal cords (vocal folds). The vocal tract modifies this excitation signal causing formant (pole) and sometimes anti-formant (zero) frequencies. Each formant frequency has also amplitude and bandwidth and it may be sometimes difficult to define some of these parameters correctly. The fundamental frequency and formant frequencies are probably the most important concepts in speech synthesis and also in speech processing in general. With purely unvoiced sounds, there is no fundamental frequency in excitation signal and therefore no harmonic structure either and the excitation can be considered as white noise. The airflow is forced through a vocal tract constriction which can occur in several places between glottis and mouth. Some sounds are produced with complete stoppage of airflow followed by a sudden release, producing an impulsive turbulent excitation often followed by a more protracted turbulent excitation. Unvoiced sounds are also usually more silent and less steady than voiced ones. Speech signals of the three vowels (/a/ /i/ /u/) are presented in time- and frequency domain in Figure: 5.2.

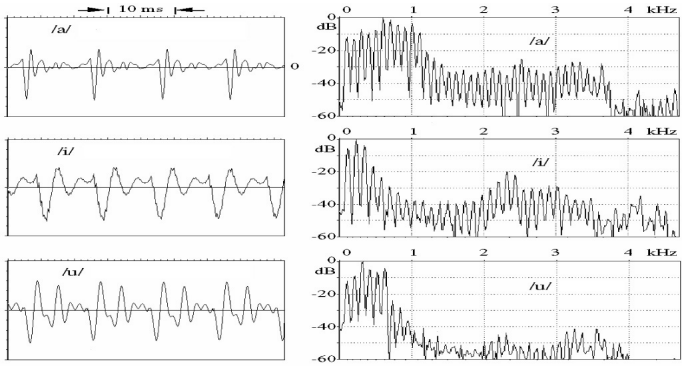


Figure5.2: The time- and frequency-domain presentation of vowels /a/, /i/, and /u/.

The fundamental frequency is about 100 Hz in all cases and the formant frequencies F1, F2, and F3 with vowel /a/ are approximately 600 Hz, 1000 Hz, and 2500 Hz respectively. With vowel /i/ the first three formants are 200 Hz, 2300 Hz, and 3000 Hz, and with /u/ 300 Hz, 600 Hz, and 2300 Hz. The harmonic structure of the excitation is also easy to perceive from frequency domain presentation. For determining the fundamental frequency or pitch of speech, for example a method called cepstral analysis may be used [9]. Cepstrum is obtained by first windowing and making Discrete Fourier Transform (DFT) for the signal and then logaritmizing power spectrum and finally transforming it back to the time-domain by Inverse Discrete Fourier Transform (IDFT). The procedure is shown in Figure 1.2.

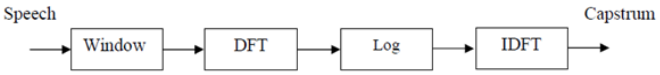


Figure 5.3 Cepstral analyses.

Fundamental frequency or intonation contour over the sentence is important for correct prosody and natural sounding speech. The different contours are usually analysed Window DFT Log IDFT Speech Cap strum from natural speech in specific situations and with specific speaker characteristics and then applied to rules to generate the synthetic speech.

C. ActiveX (SAPI)

The Speech Application Programming Interface or SAPI is an API developed by Microsoft to allow the use of speech recognition and speech synthesis The Speech API can be viewed as an interface or piece of middleware which sits between applications and speech engines (recognition and synthesis). In SAPI versions LabVIEW application could directly communicate with engines. The API included an abstract interface definition which applications and engines conformed to. Applications could also use simplified higher-level objects rather than directly call methods on the engines.

ISSN (ONLINE): 2454-9762
ISSN (PRINT): 2454-9762
Available online at www.ijarmate.com

International Journal of Advanced Research in Management, Architecture, Technology and Engineering (IJARMATE) Vol. 3, Special Issue 6, March 2017

VI. Automatic Page turner

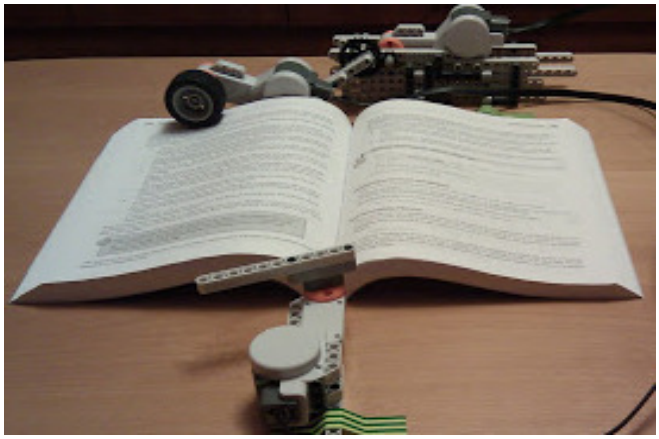


Figure:5.1 Automatic page turner

Automatic page turner is used for flipping and turning the book pages automatically from one page to another to capture the content of text in the next page this is been don using two geared DC motor , one is from edge flipping and another is for turning the page. The motor are been controlled using PWM which is been generated using myRIO .

VI RESULTS

This paper work describes OCR based Speech Synthesis System to produces a wave file output that can be used as a good mode of communication between people. The system is implemented on LabVIEW 14 platform. The above figure shows the result of the application. Figure 7.1 shows the recognization of the character from a book and it is been heard as David voice which is a default voice used in Microsoft applications. Similarly passport, handwritten text can also be recognized using this application which are been shown in figure 7.2 and 7.3 respectively.

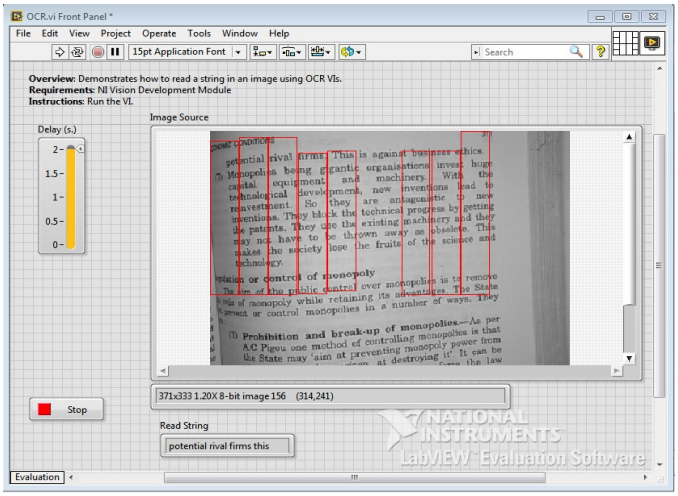


Figure 7.1: Text recognized from a book using camera

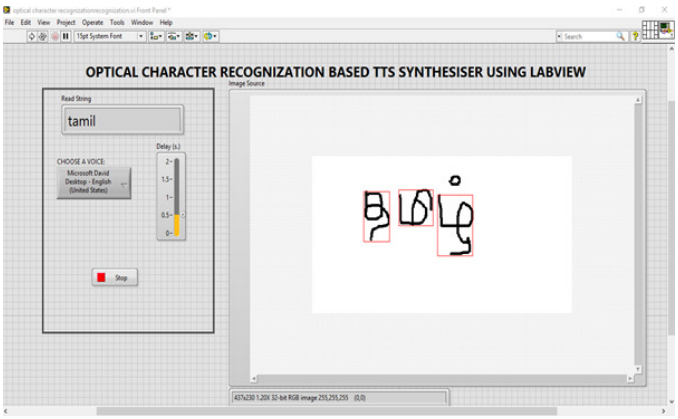


Figure: 7.2: OCR from image drawn in MSpaint

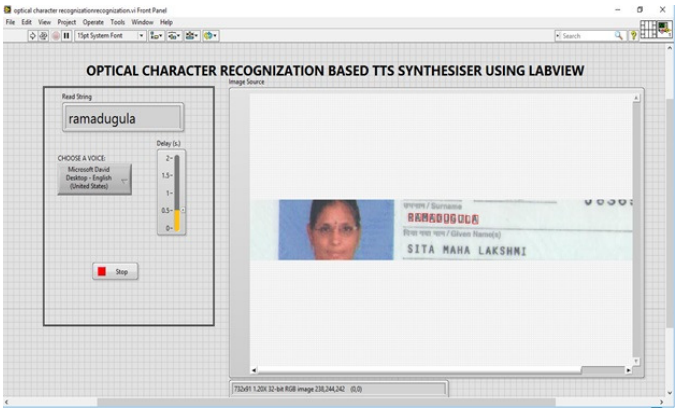


Figure 7.3: OCR from passport image

There is two session of system first is OCR and second is Speech Synthesis. In OCR printed or written character documents are scanned and image is acquired by using IMAQ

ISSN (ONLINE): 2454-9762
ISSN (PRINT): 2454-9762
Available online at www.ijarmate.com

International Journal of Advanced Research in Management, Architecture, Technology and Engineering (IJARMATE) Vol. 3, Special Issue 6, March 2017

Vision for LabVIEW and then characters are recognized using segmentation and template matching methods developed in LabVIEW. In second section recognized text is converted into speech. The ACTIVE X sub pallet in Communication pallet is used to exchange data between applications. ActiveX technology provides a standard model for interapplication communication that different programming languages can implement on different platforms. Microsoft Speech Object 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, OCR recognized text to be read, and adjust the rate and volume of the selected voice. The application developed is user friendly, cost effective and gives the result in the real time. Moreover, the program has the required flexibility to be modified easily if the need arises.

VII APPLICATION

- Digitizing library books and old valuable books
- Data entry for business documents, e.g. check, passport, invoice, bank statement and receipt
- Automatic insurance documents key information extraction
- Extracting business card information into a contact list
- More quickly make textual versions of printed documents, e.g. book scanning for Project Gutenberg
- Make electronic images of printed documents searchable, e.g. Google Books
- Converting handwriting in real time to control a computer pen computing
- Defeating CAPTCHA anti-bor systems, though these are specifically designed to prevent OCR
- Assistive technology for blind and visually impaired users

VII CONCISION

Although lot of work has been done in the field of OCR and Speech Synthesis individually, but it is first OCR based Speech Synthesis System using FPGA implementation in myrio which provides more reliability and accuracy in recognization At the highest level, FPGAs are reprogrammable silicon chips. Using prebuilt logic blocks and programmable routing resources, you can configure these chips to implement custom hardware functionality without ever having to pick up a breadboard or soldering iron. And under labview platform its more easier to configure the FPGA implementation using graphical programming language.

REFERENCES

[1] Paramver singh, Rashpinder, Gurjinder singh "LabVIEW based real time alphanumeric character recognization" - IEEE,2015
[2] N.Bhatia, "Optical character recognition technique:a review,"International Journal of Advanced Research in Computer Science and Software engineering, vol. 4, pp. 1219-1223, 2014.
[3] A.Chopra, A.A.Ghadge, O.A.Padwal, K.S.Punjabi, G.S.Gurjar, "Optical charcter recognition," International journal of advanced research in Computer and Commuination Engineering, vol. 3, pp. 4956-4958, 2014.
[4] Deepa V.Jose, Alfateh Mustafa, Sharan R "A Novel Model for Speech to Text Conversion," published in International Refereed Journal of Engineering and Science (IRJES) ISSN (Online) 2319-183X, (Print) 2319- 1821Volume 3, Issue 1 (January 2014), pp.39-41
[5] B.M. Sagar, Shobha G, R. P. Kumar, "OCR for printed Kannada text to machine editable format using database approach" WSEAS Transactions on Computers Volume 7, Pages 766-769, 2008.
[6] 6. Smith R., "An Overview of the Tesseract OCR Engine," Document Analysis and Recognition, 2007. ICDAR 2007. Ninth International Conference on, vol.2, no., pp.629-633, 2007
[7] Y.M. Alginahi., "Thesholding and Character Recognition in Security Documents with Watermarked Background," Computing: Techniques and Applications", 2008. DICTA 08.Digital Image, vol., no., pp.220-225, 2008.