DECLARATION

We Brijesh M C, bearing the USN : 4JC11EC068,Deepak N bearing the USN : 4JC11EC070, Nagarjun C S, bearing the USN : 4JC11EC073, Nagendra P, bearing the USN : 4JC11EC074 student of BE in Electronics and Communication Engineering, SJCE, Mysore, do hereby declare that this Thesis entitled **"TEXT BOOK NARRATOR"** has been independently carried out by us under the guidance of Mrs Gayatri H M, Assistant Professor, Department of Electronics and Communication, Sri Jayachamarajendra College Of Engineering, Mysore in partial fullment of the requirements for the award of the degree of Bachelor Of Engineering in Electronics and Communication, Visvesvaraya Technological University, Belgaum.

We also declare that we have not submitted this Thesis to any other University for the award of any degree or diploma.

Place : Mysore
Date :

Brijesh M C
USN : 4JC11EC068

Deepak N
4JC11EC070

Nagarjun C S
USN : 4JC11EC073

Nagendra P
USN : 4JC11EC074

ACKNOWLEDGEMENT

The success of any endeavor depends a lot on the goals set at the onset as well as the constant guidance and motivation received throughout. I take this opportunity to express my deepest gratitude and appreciation towards all those who have helped me directly or indirectly towards the successful completion of this project.

I would like to express my deep sense of respect and gratitude towards my adviser and guide Gayitri H M, Asst. Professor,Department of Electronics and Communication Engineering, who has been the guiding force behind this work. I am greatly indebted to him for his constant encouragement, valuable advice and for propelling me further in every aspect of my academic life. His presence and optimism have provided an invaluable influence on my Carrier and outlook for the future. I consider it my good fortune to have got an opportunity to work with such a wonderful person.

We express our genuine appreciation towards Dr. Shakeeb-Ur-Rahman, Principal SJCE, Mysore. We express our most heartfelt gratitude to Dr.M N Shanmukha swami, Professor and Head, Department of Electronics and Communication, SJCE, Mysore who encouraged us at this venture. He has been very influential and enthusiastic and has helped us a lot in this project with his valuable inputs. We also thank Sri Jayachamarajendra College of Engineering in catering to our academic needs. Our sincere thanks to all Teaching and Non teaching staff of EC dept, SJCE, Mysore and all other diligent members of the college for their support during the course of the project.

ABSTRACT

The growing India needs the development of the dierentially abled people. Our project aims at building a comprehensive system which helps the visually impaired people to acquire the knowledge of the book. The project concept is to use the voice signal processing to detect the input speech using a speech recognition system and accordingly carryout the operations such as NEXT, PREVIOUS or READ. NEXT and PREVIOUS operations are used to turn the page of a book in the forward or backward direction respectively. READ operation makes use of Digital image processing to read out the content of a page.

The software part has been implemented using open source software and a low cost and durable hardware. The whole of the project can be implemented at a very low cost and according to the standards of the current generation. This project can be installed in libraries to help the blind to read the books without the help of others.

LIST OF FIGURES

TABLE OF CONTENTS

1 INTRODUCTION

1.1 Motivation for Project

In an era of rapid evolution of technology, the majority of the focus is on improving the luxury of people who already have enough facilities, while there is a small but considerable amount of population who are differently abled and are in more need of our services.Physically challenged people especially blind people find it very difficult to take up their education along with normal people because of their inability to read due to their lack of vision. Our attempt is towards one such group,the visually impaired or the blind. We are creating a electronic reader for the visually impaired along with automated page turner.

1.2 Aim of the Project

We intend to improve the quality of life of the visually impaired people using image processing and voice processing to read the contents of a book or newspaper. These people will be able to read the contents of the book without others people help by listening to that content. They can turn the pages of the book through voice commands. This will be helpful for the handicapped people as well.

1.3 Project Description

The project creates the hardware which will read the contents of the book for the reader. There is also a feature for turning the pages of the book through voice command. The hardware reads out the contents of a book, blind people can hear to it and gain knowledge of the book.The following voice commands will perform the following functions :

1. Next : The device will turn to next page of the book.
2. Previous : The device will turn to previous page of the book.
3. Read : The device starts reading content of current page.

1.4 Block Diagram

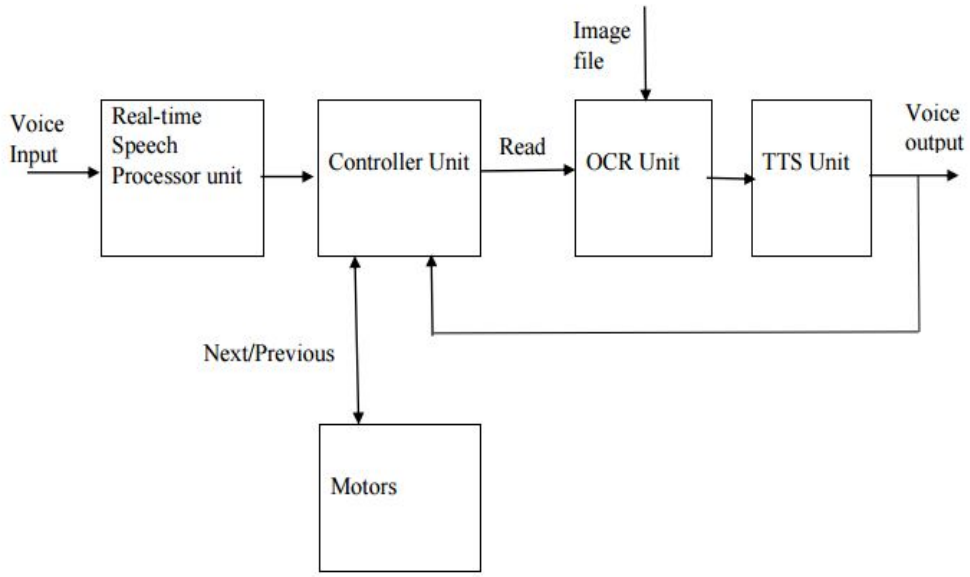

FIGURE 1.1 – Block diagram of textbook narrator

Functional Block Description :

1. Real Time Speech Processing Unit :This block is used for processing the speech signal. It analyses the speech input andproduces control parameters for commands such as next, previous and read correspondingly.
2. Control Unit :The output control parameters of speech processing unit is given as input to this block. According to the control parameters received, it takes decision whether to read the current page or turn the current page to next or previous one.
3. OCR Unit :Control unit transfers the control to the OCR unit. If the Read command is detected, OCR unit takes image file of the page to be read as input from the camera and processes it and converts it into editable document, which is given as input to TTS unit
4. TTS Unit :Editable document is processed and read by this block.

5. Motor :It is used to turn the page of book to next or previous page according to the command detected.

1.5 Organisation of the report

— Chapter 1 Gives a brief introduction of the project work. It also discusses the project description, block diagram, motivation and aim of the project.

— Chapter 2 describes about the literature survey taken from published in IEEE translations and reputed websites related to visually impaired people, the prototype devices concerning the visually impaired.

— Chapter 3 explores the hardware and software requirements of the project.

— Chapter 4 explains the approach and analysis of this project.

— Chapter 5 explains design and implementation for the project.

— Chapter 6 gives the result and summary of this project.

— Chapter 7 gives the conclusion, achievements and future work.

2 LITERATURE SURVEY

2.1 Audio-Supported Reading

The opportunity to access and interact with text, both printed and electronic, continues to be fundamental to education in the information age. It is often stated that students first learn to read so that they can later read to learn, but students who are blind or have low vision are particularly disadvantaged when it comes to working with printed text. While they may readily learn to read, using either braille or magnified print, they may struggle to reach reading rate levels adequate for achieving their true potential as they move through the educational system and on into employment. While some readers of braille obtain rates comparable to those of average print readers, most typically read at rates of one third to one half those of their sighted peers . Readers of enlarged or magnified print fare only slightly better. Remarkably, these comparative averages have remained stable for decades, suggesting that they are unrelated to shifts in teaching practices or opportunity to learn (OtL). Rather, these low reading rates are more likely the direct result of sensory limitations occasioned by blindness or visual impairment. Consequently, there is widespread agreement among educators that, while exceptions exist, students with vision impairments take appreciably longer to complete tasks requiring reading than do their sighted counterparts.

2.2 Human Machine Interface A Smart OCR for the Visually Challenged

Inclusion of the specially enabled in the IT revolution is both a social obligation as well as a computational challenge in the rapidly advancing digital world today. This paper addresses the integration of a complete Malayalam Text Read-out system designed for the visually challenged. The system accepts a page of printed Malayalam text with English numerals scans it into a digital document which is then subjected to skew correction, segmentation, before feature extraction to perform classification. Once classified, the text in Malayalam is

read- out by a text to speech conversion unit. Alternately the user has an option to get Braille prints of selected portions. The system finds interesting applications in libraries, offices where instructions and notices are to be read and also in assisted filling of application forms.

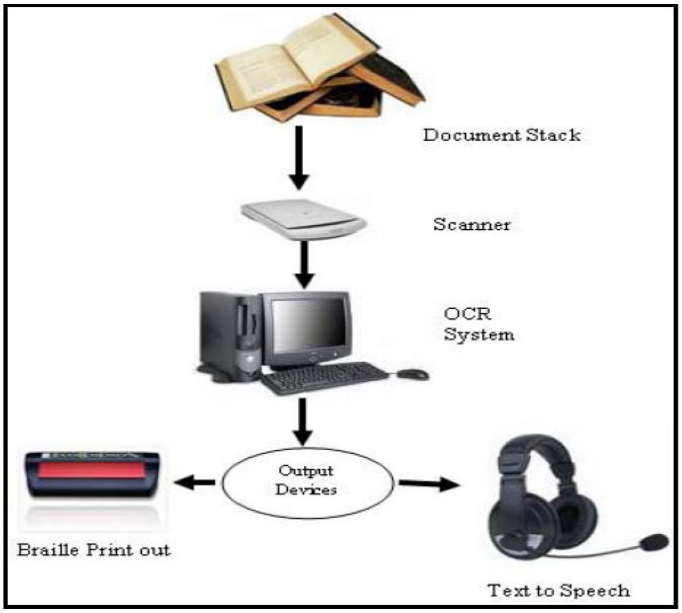

FIGURE 2.1 – Completely Integrated Malayalam OCR System

2.3 Optical character recognition for visually impaired persons

Enclosure of the specially enabled in the IT revolution is both a social responsibility as well as a computational dispute in the rapidly advancing digital world today. The availability of text books for the visually challenged is limited due to the process of conversion. Generally the book in the English will be typed into Braille sheet with the Braille format using the Braille printer. The effort taken for this is very high because for each and every character their corresponding Braille character should be printed. This system enables the automated conversion of scanned images of book directly to the Braille format. The optical character recognition is primarily used for this purpose. Optical character recognition (OCR) method has been used in converting printed text into editable text. OCR is very popular and useful

method in numerous applications. One of the important applications of this tool is its use in Braille Translation. Braille is used as the primary writing and reading system used by the visually challenged since the 19th century. Textto-speech (TTS) is a type of speech synthesis application that is used to create a spoken sound version of the text in a computer document.

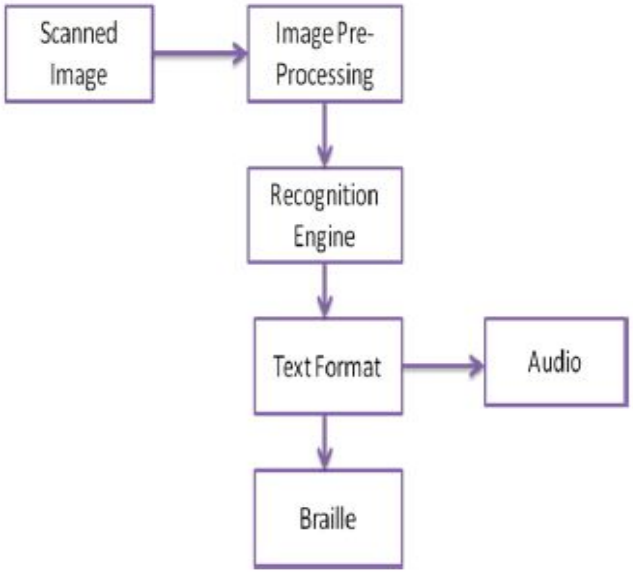

FIGURE 2.2 – Conversion of image to barilie or audio

2.4 Camera-based analysis of text and documents

The increasing availability of high- performance, low-priced, portable digital imaging devices has created a tremendous opportunity for supplementing traditional scanning for document image acquisition. Digital cameras attached to cellular phones, PDAs, or wearable computers, and standalone image or video devices are highly mobile and easy to use ; they can capture images of thick books, historical manuscripts too fragile to touch, and text in scenes, making them much more versatile than desktop scanners. Should robust solutions to the analysis of documents captured with such devices become available, there will clearly be a demand in many domains. Traditional scanner-based document analysis techniques

provide us with a good reference and starting point, but they cannot be used directly on camera-captured images. Camera-captured images can suffer from low resolution, blur, and perspective distortion, as well as complex layout and interaction of the content and background. In this paper we present a survey of application domains, technical challenges, and solutions for the analysis of documents captured by digital cameras. We begin by describing typical imaging devices and the imaging process. We discuss document analysis from a single camera-captured image as well as multiple frames and highlight some sample applications under development and feasible ideas for future development.

3 HARDWARE AND SOFTWARE REQUIREMENTS

Hardware and Software Components are the basic necessity to carry out a project. This chapter gives detailed description about the Hardware and Software components used.

3.1 Hardware Requirements

3.1.1 Atmega16

ATmega16 is an 8 bit high performance microcontroller with low power consumption. Atmega16 is based on enhanced RISC (Reduced Instruction Set Computing) architecture with 131 powerful instructions. Most of the instructions execute in one machine cycle. Atmega16 can work on a maximum frequency of 16MHz

ATmega16 has 16 KB programmable flash memory, static RAM of 1 KB and EEPROM of 512 Bytes. The endurance cycle of flash memory and EEPROM is 10,000 and 100,000, respectively. ATmega16 is a 40 pin microcontroller. There are 32 I/O (input/output) lines which are divided into four 8-bit ports designated as PORTA, PORTB, PORTC and PORTD.

Some of the features of Atmega16 are as follows :

1. High-performance, Low-power AVR 8-bit Microcontroller
2. Advanced RISC Architecture
   - 131 Powerful Instructions Most Single-clock Cycle Execution
   - 32 x 8 General Purpose Working Registers
   - Fully Static Operation
   - LCD ModuleUp to 16 MIPS Throughput at 16 MHz
   - On-chip 2-cycle Multiplier
3. High Endurance Non-volatile Memory segments
   - 16K Bytes of In-System Self-programmable Flash program memory

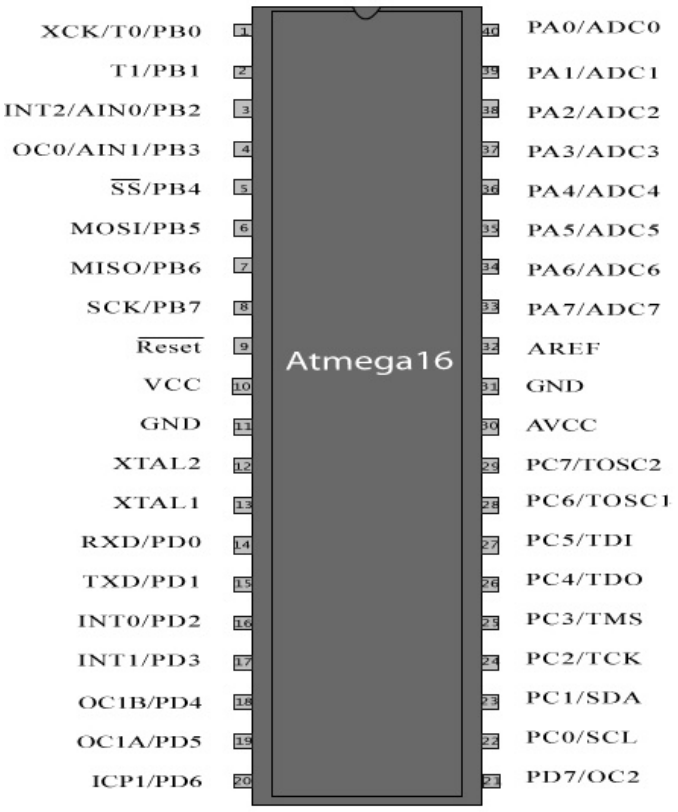

FIGURE 3.1 – ATMega16 pin diagram

— 512 Bytes EEPROM

— 1K Byte Internal SRAM

— Write/Erase Cycles : 10,000 Flash/100,000 EEPROM

— Optional Boot Code Section with Independent Lock Bits

3.1.2 Servomotor

A servomotor is a rotary actuator that allows for precise control of angular position, velocity and acceleration. It consists of a suitable motor coupled to a sensor for position feedback. It also requires a relatively sophisticated controller, often a dedicated module designed specifically for use with servomotors.

Servomotors are not a specific class of motor although the term servomotor is often used to refer to a motor suitable for use in a closed-loop control system. Servomotors are used in applications such as robotics, CNC machinery or automated manufacturing.

As the name suggests, a servomotor is a servomechanism. More specifically, it is a closed-loop servomechanism that uses position feedback to control its motion and final position. The input to its control is some signal, either analogue or digital, representing the position commanded for the output shaft.

The motor is paired with some type of encoder to provide position and speed feedback. In the simplest case, only the position is measured. The measured position of the output is compared to the command position, the external input to the controller. If the output position differs from that required, an error signal is generated which then causes the motor to rotate in either direction, as needed to bring the output shaft to the appropriate position. As the positions approach, the error signal reduces to zero and the motor stops.

The very simplest servomotors use position-only sensing via a potentiometer and bang-bang control of their motor the motor always rotates at full speed (or is stopped). This type of servomotor is not widely used in industrial motion control, but it forms the basis of the simple and cheap servos used for radio-controlled models.

3.1.3 MAX232 IC

The MAX232 IC is used to convert the TTL/CMOS logic levels to RS232 logic levels during serial communication of microcontrollers with PC. The controller operates at TTL logic level (0-5V) whereas the serial communication in PC works on RS232 standards (-25 V to + 25V). This makes it difficult to establish a direct link between them to communicate with each other.

The intermediate link is provided through MAX232. It is a dual driver/receiver that includes a capacitive voltage generator to supply RS232 voltage levels from a single 5V supply. Each receiver converts RS232 inputs to 5V TTL/CMOS levels. These receivers (R1 and R2)

can accept +30V to -30v inputs. The drivers (T1 and T2), also called transmitters, convert the TTL/CMOS input level into RS232 level.

The transmitters take input from controllers serial transmission pin and send the output to RS232s receiver. The receivers, on the other hand, take input from transmission pin of RS232 serial port and give serial output to microcontrollers receiver pin. MAX232 needs four external capacitors whose value ranges from $1\mu F to 22\mu F$.

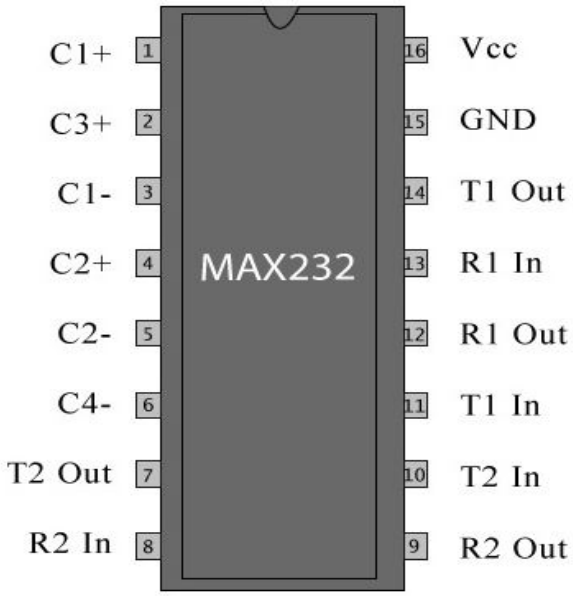

FIGURE 3.2 – MAX232 pin diagram

3.1.4 RS232 Converter and Flat cable

RS232 is a complete standard. This means that the standard sets out to ensure compatibility between the host and the peripheral system by specifying a common voltage and signal levels and common pin wiring configurations and a minimal amount of control information between the host and pepirheral systems and a high level for the driver output is defined as being +5 to +15 volts and a low level for the driver output is defined as being between 5 and 15 volts.

The receiver logic levels were defined to provide a 2V noise margin. The RS232 standard also limits the maximum slew rate at the driver output.

RS232 Flat Cable :

FIGURE 3.3 – RS232 Flat Cable

RS 232 flat cable is the cable that is required in order to access the RS232 header. This cable can be purchased from various sources and is referred to as the ATI/Everex type cable. A widely used rule of thumb indicates that cables more than 50 feet (15 m) long will have too much capacitance, unless special cables are used. By using low-capacitance cables, full speed communication can be maintained over larger distances up to about 1,000 feet (300 m). For longer distances, other signal standards are better suited to maintain high speed. For functional communication through a serial port interface, conventions of bit rate, character framing, communications protocol, character encoding, data compression, and error detection, not defined in RS 232, must be agreed to by both sending and receiving equipment.Since most RS-232 devices do not have auto-matic baud rate detection, users must manually set the baud rate (and all other parameters) at both ends of the RS-232 connection.

3.1.5 Motor Driver IC

L293D is a dual H-bridge motor driver integrated circuit (IC). Motor drivers act as current amplifiers since they take a low-current control signal and provide a higher-current signal. This higher current signal is used to drive the motors. L293D contains two inbuilt

H-bridge driver circuits. In its common mode of operation, two DC motors can be driven simultaneously, both in forward and reverse direction. The motor operations of two motors can be controlled by input logic at pins 2 and 7 and 10 and 15. Input logic 00 or 11 will stop the corresponding motor. Logic 01 and 10 will rotate it in clockwise and anticlockwise directions, respectively. Enable pins 1 and 9 (corresponding to the two motors) must be high for motors to start operating. When an enable input is high, the associated driver gets enabled. As a result, the outputs become active and work in phase with their inputs. Similarly, when the enable input is low, that driver is disabled, and their outputs are off and in the high-impedance state.

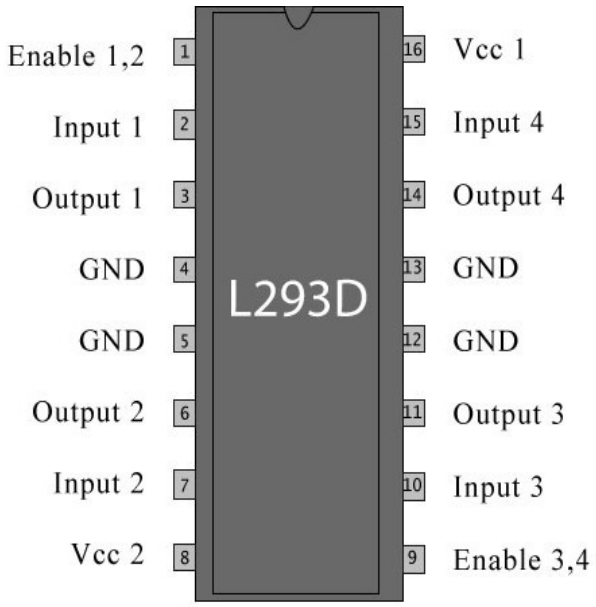

FIGURE 3.4 – L293D pin diagram

3.1.6 Base

The book to be read is placed on this structure. It has been designed to handle any kinds of book independent of size. It has a structure to hold the camera and the motor drivers.This has been designed by us according to the needs of the people.It has tools to keep the book independent of size. The motors are placed on arms which will help to turn the

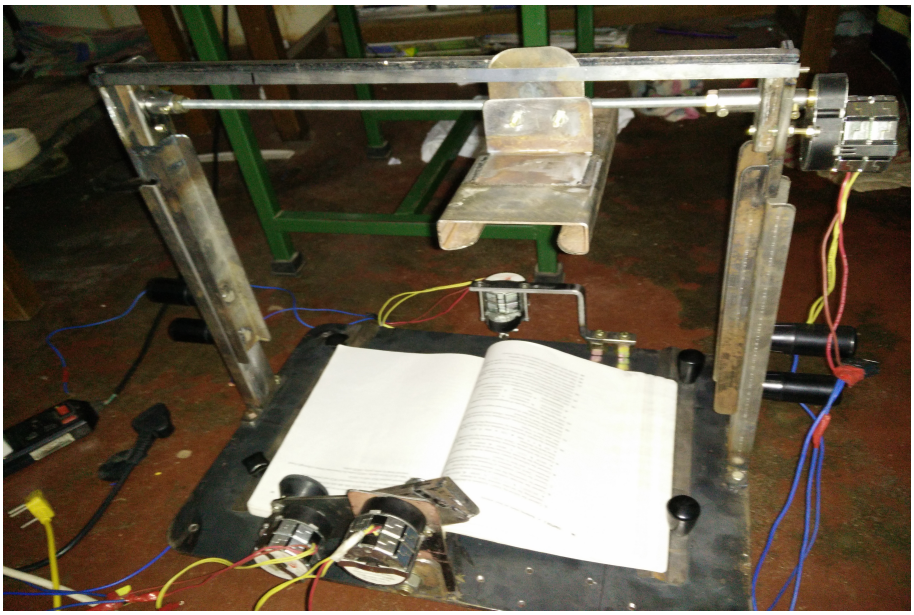

FIGURE 3.5 – Base

pages of the book according to the voice commands given.

3.2 Software Requirements

3.2.1 AVR Studio

AVR Studio was created by Atmel in order to help developers to create applications for AVR micro controllers using C/C++ programming languages. This piece of software comes with a large number of tutorials, which allow the users to get familiar with the application. The program stands as a complete pack for programmers that use C++ and other programming languages.

The menu of this application is easy-to-use and offers access to powerful tools for both beginners and experienced developers, making it easy for the users to find their way through C/C++ programming. Some of the key features are : cycle correct simulator with advanced debugging functionality, rich SDK that enables tight integration of customer plug-

ins and compatibility with many Microsoft Visual Studio plug-ins. Also the tool provides a "split window" button that allows the users to work on more than one project at a time.

All in all AVR Studio is a complete tool for programmers which develop, test and debug C/C++ applications; you should give this tool a try as it comes in handy for programming AVR microcontrollers.

Here AVR studio is used to create programs for the microcontroller in order to control the arms according to the instructions from the speech recognition programme.

3.2.2 Java and Sphinx

Java is a general-purpose computer programming language that is concurrent, class-based, object-oriented, and specifically designed to have as few implementation dependencies as possible. It is intended to let application developers "write once, run anywhere" (WORA), meaning that compiled Java code can run on all platforms that support Java without the need for recompilation. Java applications are typically compiled to bytecode that can run on any Java virtual machine (JVM) regardless of computer architecture. As of 2015, Java is one of the most popular programming languages in use, particularly for client-server web applications, with a reported 9 million developers.[citation needed] Java was originally developed by James Gosling at Sun Microsystems (which has since been acquired by Oracle Corporation) and released in 1995 as a core component of Sun Microsystems' Java platform. The language derives much of its syntax from C and C++, but it has fewer low-level facilities than either of them.

Sphinx is a continuous-speech, speaker-independent recognition system making use of hidden Markov acoustic models (HMMs) and an n-gram statistical language model. It was developed by Kai-Fu Lee. Sphinx featured feasibility of continuous-speech, speaker-independent large-vocabulary recognition, the possibility of which was in dispute at the time (1986). Sphinx is of historical interest only; it has been superseded in performance by subsequent versions.

The original and reference implementation Java compilers, virtual machines, and class libraries were originally released by Sun under proprietary licences. As of May 2007, in compliance with the specifications of the Java Community Process, Sun relicensed most of its Java technologies under the GNU General Public License. Others have also developed alternative implementations of these Sun technologies, such as the GNU Compiler for Java (bytecode compiler), GNU Classpath (standard libraries), and IcedTea-Web (browser plugin for applets).

The Java Speech API (JSAPI) is an application programming interface for cross-platform support of command and control recognizers, dictation systems, and speech synthesizers. Although JSAPI defines an interface only there are several implementations created by third parties.

Sphinx is a continuous-speech, speaker-independent recognition system making use of hidden Markov acoustic models (HMMs) and an n-gram statistical language model. It was developed by Kai-Fu Lee. Sphinx featured feasibility of continuous-speech, speaker-independent large-vocabulary recognition, the possibility of which was in dispute at the time (1986). Sphinx is of historical interest only ; it has been superseded in performance by subsequent versions.Sphinx is a pure Java speech recognition library. It provides a quick and easy API to convert the speech recordings into text with the help CMUSphinx acoustic models. It can be used on servers and in desktop applications. Beside speech recognition Sphinx helps to identify speakers, adapt models, align existing transcription to audio for timestamping and more.

3.2.3 Abbyy Fine Reader

ABBYY Fine Reader OCR software helps individuals turn scans of paper documents, PDF files, and digital photographs into searchable and editable formats. Unmatched text recognition accuracy and conversion capabilities virtually eliminate retyping and reformatting. Intuitive use and one-click automated tasks let you do more with this OCR software in fewer

steps. Up to 190 languages supported for text recognition and conversion - absolute record on OCR/PDF market. Version 12 supports OCR for 190 languages and has a built-in spell checker for 48 of them.

The program allows you to extract text data from digital images (photos, scan results, PDF-files). The resulting recognition can be saved in different file formats : Microsoft Word, Microsoft Excel, Microsoft Powerpoint, Rich Text Format, HTML, PDF, CSV and text (plain text) files. Starting with version 11 files can be saved in a format djvu.

Of the features can be noted

1)Made of high-level procedure supplements the unknown word in the dictionary of the program : offers several opportunities early forms of expression, and on the basis of this system is its construction possible declination (you can adjust the wrong shape). Thus, immediately added to a single word, and all its forms.

2)Training program for human handwriting recognition or non-standard fonts, through the mechanism of standards.The first version of FineReader was released in 1993 .

3.2.4 Ivona Text To Speech

IVONA Reader is a personal easy-to-use text reader. Converts any written text from your PC into spoken words. This text-to-speech reader allows computer to read any text aloud. It can read documents, news, RSS, books, web pages and emails.

It is composed of two parts : a front-end and a back-end. The front-end has two major tasks. First, it converts raw text containing symbols like numbers and abbreviations into the equivalent of written-out words. This process is often called text normalization, pre-processing, or tokenization. The front-end then assigns phonetic transcriptions to each word, and divides and marks the text into prosodic units, like phrases, clauses, and sentences. Phonetic transcriptions and prosody information together make up the symbolic linguistic representation that is output by the front-end. The back-endoften referred to as the synthesizerthen converts the symbolic linguistic representation into sound.

IVONA uses Unit Selection with Limited Time-scale Modification (USLTM). Unit selection synthesis uses large databases of recorded speech. During database creation, each recorded utterance is segmented into some or all of the following : individual phones, syllables, morphemes, words, phrases, and sentences. The division into segments is done using a specially modified speech recognizer. An index of the units in the speech database is then created based on the segmentation and acoustic parameters like the fundamental frequency (pitch), duration, position in the syllable, and neighboring phones. At runtime, the desired target utterance is created by determining the best chain of candidate units from the database (unit selection).

3.2.5 Serial Communication USART

Communication between two entities is important for the information flow to take place. In general the information transport system can be parallel in which the complete byte of data is sent at a time, with each bit having a separate dedicated line or it can be serial where only one communication line is available which is shared by all the bits sequentially. The pros and cons of these two systems are equivalent and selection between the two depends on the application. Data can be exchanged using parallel or serial techniques. Setup for parallel data transfer is not cost effective but is a very fast method of communication. Serial communication is cost effective because it requires only a single line of connection but on the other hand is a slow process in comparison to parallel communication.

There are two methods for serial data communication :

1. Synchronous Communication
2. Asynchronous Communication

In Synchronous communication method complete block (characters) is sent at a time. It does'nt require any additional bits (start, stop or parity) to be added for the synchronization of frame. The devices are synchronized by clock. And in asynchronous communication data transmission is done byte by byte i.e., one byte at a time. The additional bits are added

to complete a frame. In synchronous communication the frame consists of data bits while in asynchronous communication the total number of bits in a frame may be more than the data bits.

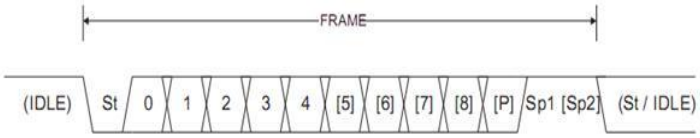

FIGURE 3.6 – Data Bit Frame

Atmega16 is equipped with three different kinds of serial communication peripheral systems :

1. Serial USART
2. SPI (Serial Peripheral Interface)
3. TWI (Two wire Interface)

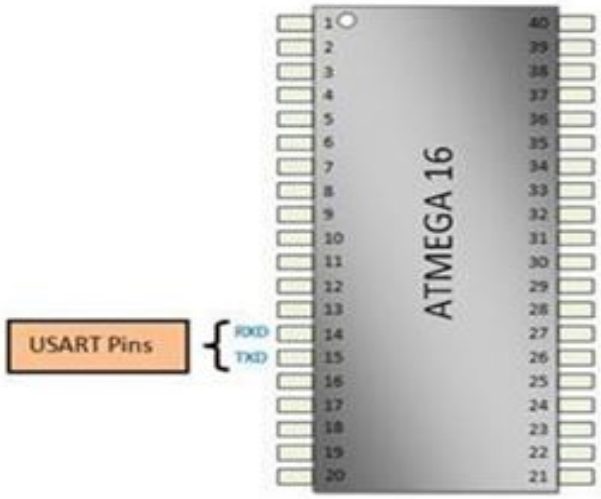

FIGURE 3.7 – USART Pins of Controller

SERIAL USART (universal synchronous asynchronous receiver and transmission/ trans-mitter) : Serial USART provides full-duplex communication between the transmitter

and receiver. Atmega16 is equipped with independent hardware for serial USART communication. Pin-14 (RXD) and Pin-15 (TXD) provide receive and transmit interface to the microcontroller.

Atmega16 USART provides asynchronous mode of communication and do not have a dedicated clock line between the transmitting and receiving end. The synchronization is achieved by properly setting the baud rate, start and stop bits in a transmission sequence.

Start bit and stop bit : These bits are use to synchronize the data frame. Start bit is one single low bit and is always given at the starting of the frame, indicating the next bits are data bits. Stop bit can be one or two high bits at the end of frame, indicating the completion of frame.

4 APPROACH AND ANALYSIS

4.1 Speech Recognition

In computer science and electrical engineering, speech recognition (SR) is the translation of spoken words into text. It is also known as "automatic speech recognition" (ASR), "computer speech recognition", or just "speech to text" (STT).

Some SR systems use "speaker-independent speech recognition"while others use "training" where an individual speaker reads sections of text into the SR system. These systems analyze the person's specific voice and use it to fine-tune the recognition of that person's speech, resulting in more accurate transcription. Systems that do not use training are called "speaker-independent" systems. Systems that use training are called "speaker-dependent" systems.

Speech recognition applications include voice user interfaces such as voice dialling (e.g. "Call home"), call routing (e.g. "I would like to make a collect call"), domotic appliance control, search (e.g. find a podcast where particular words were spoken), simple data entry (e.g., entering a credit card number), preparation of structured documents (e.g. a radiology report), speech-to-text processing (e.g., word processors or emails), and aircraft (usually termed Direct Voice Input).

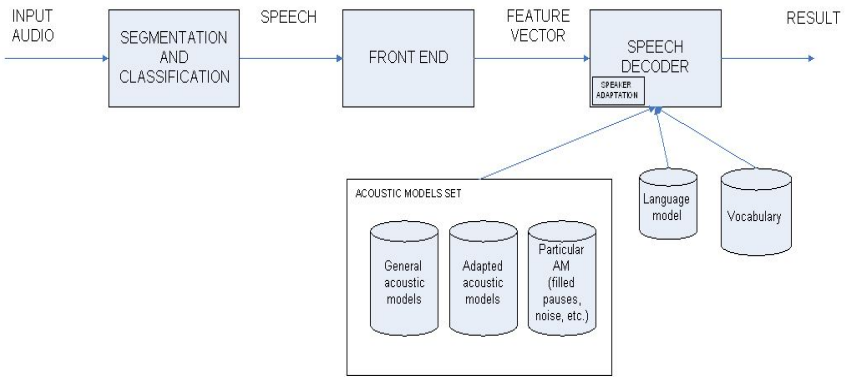

FIGURE 4.1 – Speech Recognition System

The term voice recognition or speaker identification refers to identifying the speaker, rather than what they are saying. Recognizing the speaker can simplify the task of translating speech in systems that have been trained on a specific person's voice or it can be used to authenticate or verify the identity of a speaker as part of a security process.

From the technology perspective, speech recognition has a long history with several waves of major innovations. Most recently, the field has benefited from advances in deep learning and big data.

Despite its limitations, present speech recognition technology can be a very useful tool for a variety of applications, as long as designers and users fully understand the boundaries and weaknesses of such systems. It is regrettable that the desire to hype up a new product or generation of speech recognition engine sometimes leads to blatantly misleading statements or misrepresentation of the realities of speech recognition and its role in real-world delivery.

Speech recognition is already used for live subtitling on television, as dictation tools in the medical and legal profession, and for off-line speech-to-text conversion or notetaking systems. For all these applications, human editing of the output is needed to achieve really good levels of accuracy. In addition, and as already mentioned, there are an increasing number of small vocabulary or specialised command and control applications, from sat-nav systems and voice command in smartphones, to home automation.

In this project the speech recognition is implemented using sphinx which is a open source software.

4.2 OCR

Optical character recognition (OCR) is the process of converting scanned images of machine printed orhandwritten text (numerals, letters, and symbols), into machine readable character streams, plain (e.g. textles) or formatted (e.g. HTML les). As shown in Figure below, the data path in a typical OCR system consistsof three major stages :

1. document digitization
2. character/word recognition
3. output distributionIn

the rst stage, the scanner optically captures text in documents and produces document images.Recent advances in scanner technology have made high resolution document scanning widely available.Unlike early black-and-white template matching methods, modern feature based optical recognition methodsrequire image spatial resolutions of at least 200 dots per inch (dpi) and can benet from gray-scale textimagery when it is available. Lower resolutions and simple bi-tonal thresholding tend to break thin lines orll gaps, thus distorting or invalidating character features needed in the recognition stage.

The second (and the most interesting) stage is responsible for character and/or word recognition indocument images. The process involves four operations :

1. optional image analysis : image quality assessment, text line detection, word and character extraction,etc.
2. optional image enhancement : removing speckle and other image noise, lling holes and breaks, etc.
3. character/word recognition : usually based on their shapes and other features
4. optional contextual processing : to limit the feature search space

In the third (final) stage, the output interface communicates the OCR results to the outside world.For example, many commercial systems allow recognition results to be placed directly into spread sheets,databases, and word processors. Other commercial systems use recognition results directly in further auto-mated processing and when the processing is complete, the recognition results are discarded. In any case,the output interface is vital to the OCR systems because it communicates results to the user and program domains outside of the OCR system.

In this project the optical character recognition is implemented using Ivona text to speech which is a open source software.

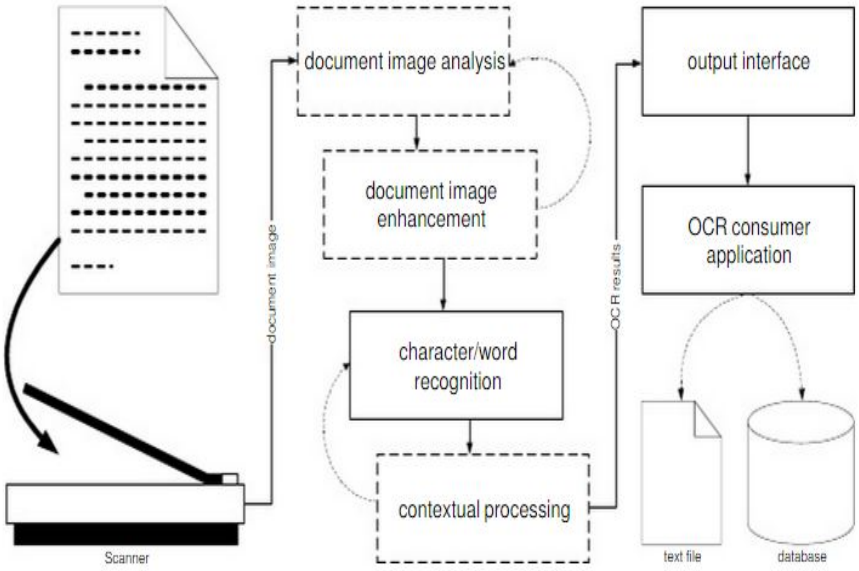

FIGURE 4.2 – A typical OCR system

4.3 TTS

The process of reading, speaking and the issues involved in computers to do this is known as speech synthesis. That is generation of speech from orthographic text is called as Text to speech (TTS) synthesis. TTS system is having an enormous range of applications and advantages. The main use of TTS system is a reading system for visually impaired community, where TTS system would read some text from book and convert it into corresponding speech. The early TTS system is of course sounded very mechanical, but its adoption by visually impaired community was surprising as there is no other option for them or having a real person do the reading were often not possible. In recent a year many sophisticated systems exist that facilitates human computer interaction for visually impaired, in which TTS system is successful in navigating this community towards success. The mainstream adoption of TTS

system has been severely limited by its quality. In recent years the considerable advances in quality have changed the situation in such a way that TTS system is common and used for many applications. TTS system is used in Telecommunication, Industrial applications, User guide for illiterate masses and wide variety of applications. It is very happy to say that research in speech synthesis has contributed an enormous amount to our understanding of language. TTS systems have made a good testing ground for many models and theories. The automatic conversion of written to spoken language is commonly called Text - to - Speech or simply TTS. The input is text and the output is a speech waveform. A TTS system is almost always divided into two main parts. The rst of these converts text into what we will call a linguistic specication and the second part uses the specication to generate a waveform. This division of the TTS system into these two parts makes a lot of sense both theoretically and for practical implementation. The rst end is typically language specic, whilst the waveform generation component can be largely independent of the language ; The conversion of text into a linguistic specication is generally achieved using a sequence of separate process and variety of internal immediate representations.

A text-to-speech system (or "engine") is composed of two parts : a front-end and a back-end. The front-end has two major tasks. First, it converts raw text containing symbols like numbers and abbreviations into the equivalent of written-out words. This process is often called text normalization, pre-processing, or tokenization. The front-end then assigns phonetic transcriptions to each word, and divides and marks the text into prosodic units, like phrases, clauses, and sentences. The process of assigning phonetic transcriptions to words is called text-to-phoneme or grapheme-to-phoneme conversion. Phonetic transcriptions and prosody information together make up the symbolic linguistic representation that is output by the front-end. The back-endoften referred to as the synthesizerthen converts the symbolic linguistic representation into sound. In certain systems, this part includes the computation of the target prosody (pitch contour, phoneme durations), which is then imposed on the output speech.

In this project the text to speech is implemented using Abbyy fine reader which is a

open source software.

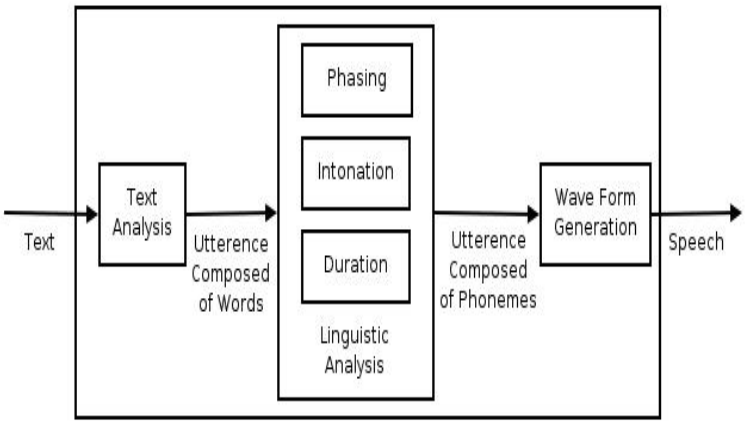

FIGURE 4.3 – A typical TTS system

4.4 Interfacing

The speech recognition system implemented using sphinx which runs in the personal computer continuously in order to recognises voice commands and it produces output according to the command.The database is added to the sphinx software in order to differentiate between the commands by having a speech graphs for each of the word it can recognise the words by different people once it is trained for the word.

The output from the sphinx programme tells the hardware through serial communication to read or pause the page or to turn the page to previous or next according to the specified commands. The hardware is implemented using a particular base and with motors to turn the page.

The OCR implemented using Abbyy fine reader converts the contents of the book which is captured by the camera using ip webcam into a editable document. This document is presented to the TTS system which is implemented using Ivona text to speech. The interfacing of all these is done with the help of autohot key software.

5 DESIGN AND IMPLEMENTATION

5.1 Circuit Diagrams

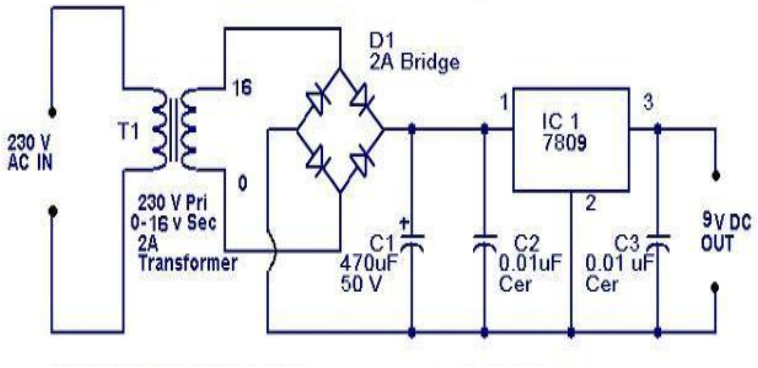

FIGURE 5.1 – Interfacing db-9 connector to ATmega16 via Max 232

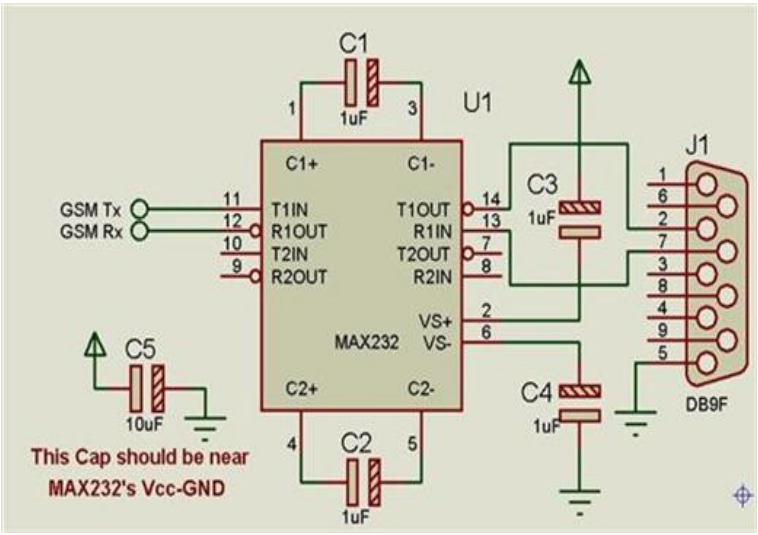

FIGURE 5.2 – Interfacing db-9 connector to ATmega16 via Max 232

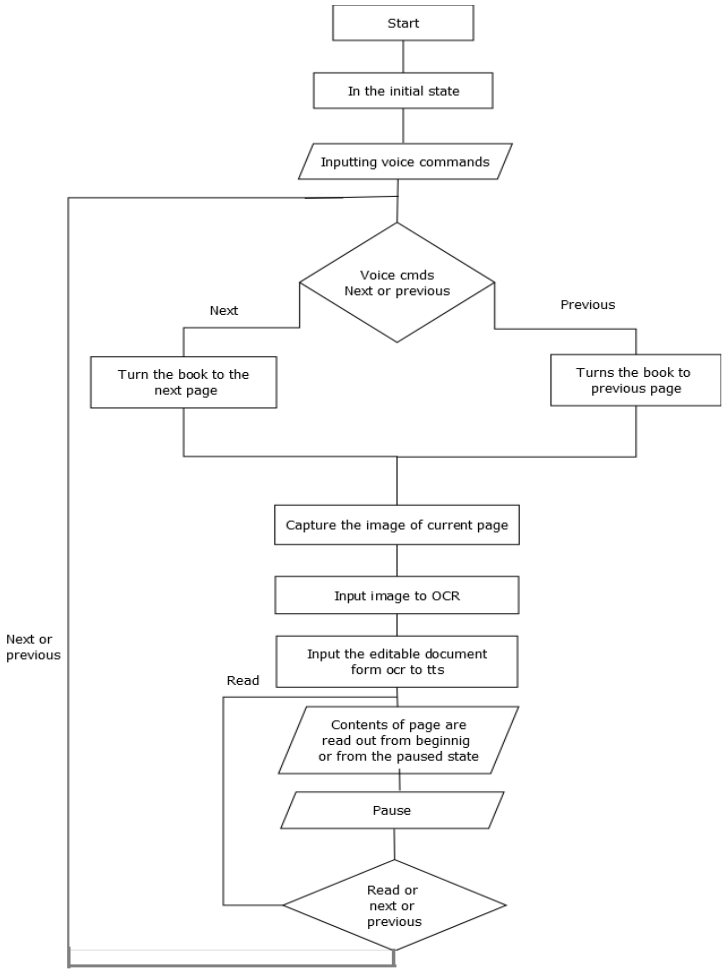

FIGURE 5.3 – Flow chart

5.2 Procedure

Voice samples were collected from different people for the words 'Next','Previous','Play' and 'Pause' . For each person about 10 samples were collected . The samples were recorded at a particular sampling frequency.

The sphinx java code(appendix) has been written in order to differentiate between different words of all the persons . The program creates the database of all the samples and extracts index from each of the and forms a reference vector. The bit to be recognized is inputted in real time mode in the presence of noise. The code recognizes the word and

provides different code for different samples . It produces one code for next and a different code for previous . Those code produced will be sent to the microcontroller via the serial communication . Code for serial communication is in appendix .

The microcontroller receives the code from the matlab via serial communication . It then proceeds depending on the code received . If next is received it will turn to the next page . If previous is received it will turn back to the previous page . The code for this is in appendix . The arms which are controlled by the DC motors which are driven by microcontroller depending on word .

6 RESULTS AND SUMMARY

6.1 Implementation Results

The circuit was switched on. The required word (previous or next or play or stop) is input through voice commands in real time mode. The sphinx java programme differentiates between the words and performs the action of turning or reading the content of the page depending on the word loaded into the program.

6.2 Advantages

Pages of the book can be turned without using hands through voice commands.The contents of the page is read out so that the physically impaired people and also to some extent the people with no functional hands so that can gain the knowledge of the book. This can be help for the blind people.

6.3 Disadvantages

1. Roller material's should be selected such that good friction must exist between roller and pages or else the page may not turn or many pages will be turned at once
2. If pages are too delicate they may be damaged.
3. Voice recognition must be trained to the words to be used.
4. The books of huge size cannot be read used because these will be too thick to turn the pages.

6.4 Cost of the project

Commodity	**Quantity**	**Cost(Rs)**
ATMega16	1	200
Roller and Arms	1	300
Gear Motors	3	300
Serial Communication Unit	1	300
Gear motor driver ICs	4	5000
Power supply unit	1	500
Web cam	1	1000
Others		200

TOTAL=7800

7 CONCLUSION AND FUTURE SCOPE

This chapter gives the detail discussion of the entire project and concludes the nal of the project, future enhancement and applications of TTS and OCR system developed as a part of typing aid for visually impaired.

7.1 Conclusion

Text book narrator system developed considering the physically disabled people successfully covers building the hardware and software. The reading of contents of the book was done eciently and works ne with all the test cases it was subjected. The concatenation based speech recognition system is developed using JAVA sphinx programming, where it uses text processed decimal code as input. This concatenation algorithm is developed and implemented by the databases DB1, DB2 and DB3 and results are simulated. The naturalness of speech output was considered while developing the speech output. The overall performance of the project was as expected and performed well in varied test cases. The project has wide impact on the society as well in terms of its applications and impact.

7.2 Future Extensions

The applications of narrator systems are increasing rapidly in day to day life. In the same way technology is growing faster and faster as researches is going on in improving the quality of the synthesized speech. Hence these are some of the future quality improvement techniques to for developed OCR and TTS system in the project.

1. Synthesized speech quality can he improved much by using standard speech database.
2. This can be implemented for any of the Language with suitable speech database.
3. Text normalization can be done for special symbols and Abbreviations.
4. This can be made to identify the special symbols and some notations.

5. THis can be made to read some mathematical notation and representations.

7.3 Applications

The application eld of TTS and OCR is expanding fast whilst the quality of TTS systems used in typing aid is also increasing steadily. Speech synthesis systems are also becoming more aordable for common customers, which makes these systems more suitable for everyday use.

1. E-readers for the blind This project was built for the main purpose of helping the physically disabled people so that they can gain knowledge of the book without much of a difficulty.They can come to know the contents simply by listening to it.
2. Audio files can be generated for the book using this system. People will be able to use this on the go in any places. This is not only helpful for the blind but also for normal people .
3. Telecommunication and multimedia The newest applications in speech synthesis are in the area of multimedia. Synthesized speech has been used for decades in all kind of telephone enquiry systems, but the quality has been far from good for common customers. Today, the quality has reached the level that normal customers are adopting it for everyday use. Electronic mail has become very usual in last few years. However, it is sometimes impossible to read those E-mail messages when last few 31 for example abroad. With synthetic speech e-mail messages may be listened to via normal telephone line. Synthesized speech may also be used to speak out short text messages (SMS) mobile phones. For totally interactive multimedia applications an automatic speech in recognition system is also needed. The automatic recognition of uent speech is still far away, but the quality of current systems is at least so good that it can be used to give some control commands, such as yes/no, on/o, or ok/cancel.

4. Man-Machine communication This application not only helps visually impaired community this also helps the illiterate masses present in various part of the country. We know that India is having a population of around 110 cores according to statistical survey and number of illiterates present in India is around 25 to 30.

8 REFERENCES

Références

[1] Bindu Philip1 and R. D. Sudhaker Samuel *Human Machine Interface A Smart OCR for the Visually Challenged* International Journal of Recent Trends in Engineering, Vol 2, No. 3, November 2009.

[2] C. Vignesh and MR. M. Senthil Kumaran *Optical character recognition for blind and visually impaired persons* International Journal of Emerging Technology in Computer Science and Electronics (IJETCSE) ISSN : 0976-1353 Volume 14 Issue 1 APRIL 2015.

[3] Jian Liang, David Doermann, Huiping Li *Camera-based analysis of text and documents : a survey* IJDAR (2005) Digital Object Identier (DOI) 10.1007/s10032-004-0138-z.

[4] Arun Kumar C and Shreekanth T *A Comprehensive review on Concatenation Based Text to Speech Synthesis for Indian Language* Int J. ElecElectr.EngTelecoms, Vol. 3, No. 2, April 2014, ISSN 2319 2518.

[5] Ravina Mithe, Supriya Indalkar, Nilam Divekar *Optical Character RecognitionAn Assistive Reading System for Visually Impaired using OCR and TTS* International Journal of Computer Applications (0975 8887) Volume 95 No 2, June 2014.

[6] Akshay Sharma , *Optical Character Recognition* International Journal of Recent Technology and Engineering (IJRTE) ISSN : 2277-3878, Volume-2, Issue-1, March 2013.

[7] Ravina Mithe, Supriya Indalkar, Nilam Divekar *Optical Character Recognition* International Journal of Recent Technology and Engineering (IJRTE) ISSN : 2277-3878, Volume-2, Issue-1, March 2013.

[8] Boersma and Weenink *PRAAT : A Tool for Phonetic Analysis and Sound Manipulations* 1992-2001, www.praat.org.

[9] Paul Taylor *Text-to-Speech Synthesis* Cambridge University Press. X. Huang, A.Acero and H.W. Hon, Spoken Language Processing A Guide to Theory, Algorithm and System Development, New Jersey : Prentice Hall, 2001.

[10] John F. Pitrelli, Raimo Bakis, Ellen M. Eide, Raul Fernandez, Wael Hamza, and Michael A. Picheny *The IBM Expressive Text-to-Speech Synthesis System for American English* IEEE Transactions on Speech and Audio Processing, VOL. 14, NO. 4, JULY 2006.

[11] T. Giannakopoulos *A method for silence removal and segmentation of speech Signals, implemented in Matlab, University of Athens, Greece* Text-to-Speech Synthesis using syllable-like units IIT, Madras.

[12] WWW.Disabled-world.com/assistivedevices/visual/

9 APPENDIX

APPENDIX A

ATmega 16

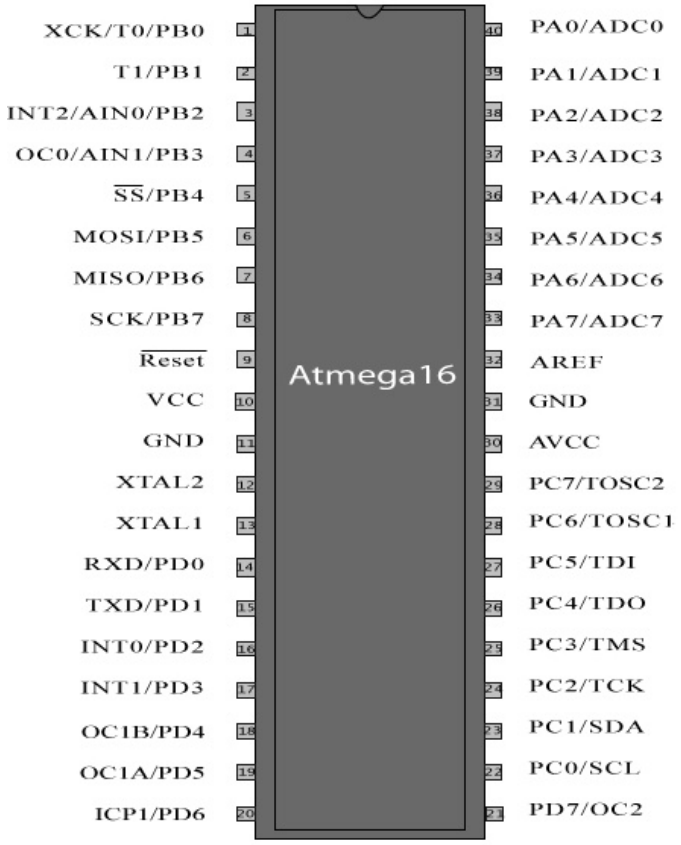

FIGURE 9.1 – ATMega16 pin diagram

Vcc : Digital supply voltage GND : Ground

Port A/B/C :

Port A/B/C is an 8-bit bi-directional I/O port with internal pull-up resistors (selected for each bit). The Port A/B/C output buffers have symmetrical drive characteristics with both high sink and source capability. As inputs, Port A/B/C pins that are externally pulled

low will source current if the pull-up resistors are activated. The Port A/B/C pins are tri-stated when a reset condition becomes active, even if the clock is not running. Fig A.1 shows the pin diagram of Atmega16.

Port D :

Port d is a 3 bit bi-directional I/O port with internal pull-up resistors (selected for each bit). The Port D output buffers have symmetrical drive characteristics with both high sink and source capability. As inputs, Port D pins that are externally pulled low will source current if the pull-up resistors are activated. The Port D pins are tri-stated when a reset condition becomes active, even if the clock is not running. If the RSTDISBL Fuse is programmed, PD0 is used as an I/O pin. Note that the electrical characteristics of PD0 differ from those of other pins of Port D. If the RSTDISBL fuse is unprogrammed, PD0 is used as a reset input. A low level on this pin for longer than the minimum pulse length will generate a reset, even if the clock is not running. Shorter pulses are not guaranteed to generate a Reset. Depending on the clock selection Fuse setting, PD1 can be used as input to the inverting oscillator amplifier and input to the internal clock operating circuit. Depending on the clock selection fuse settings, Pd2 can be used as output from the Inverting Oscillator amplifier.

AVcc :AVcc is the supply voltage pin for the A/D converter, D/A Converter, Current source. It should be externally connected to Vcc, even if the ADC, DAC are not used. If the ADC is used, it should be connected to Vcc through a low-pass filter.

ARef :This is the analog reference pin for the A/D converter.

APPENDIX B

USART Registers

To use the USART of Atmega16, certain registers need to be configured.

UCSR : USART control and status register. It is basically divided into three parts UCSRA, UCSRB and UCSRC. These registers are basically used to configure the USART.

UBRR : USART Baud Rate Registers. Basically use to set the baud rate of USART
UDR : USART data register.

1. UCSRA : (USART Control and Status Register A)

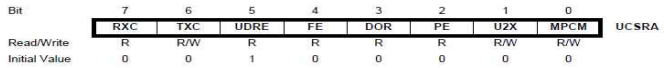

FIGURE 9.2 – UCSRCA

RXC (USART Receive Complete) : RXC flag is set to 1 if unread data exists in receive buffer, and set to 0 if receive buffer is empty.

TXC (USART Transmit complete) : TXC flag is set to 1 when data is completely transmitted to Transmit shift register and no data is present in the buffer register UDR.

UDRE (USART Data Register Empty) : This flag is set to logic 1 when the transmit buffer is empty, indicating it is ready to receive new data. UDRE bit is cleared by writing to the UDR register.

2. UCSRB : (USART Control and Status Register B)

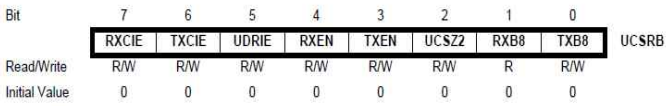

FIGURE 9.3 – UCSRCB

RXCIE : RX Complete Interrupt Enable,

When 1 - RX complete interrupt is enabled.

When 0 - RX complete interrupt is disabled.

TXCIE : TX Complete Interrupt Enable,

When 1 - TX complete interrupt is enabled

When 0- TX complete interrupt is disabled

UDRIE : USART Data Register Empty Interrupt Enable,

When 1 - UDRE flag interrupt is enabled.

When 0 - UDRE flag interrupt is disabled.

3. UCSRC : (USART Control and Status Register C)

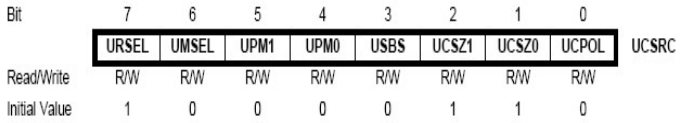

FIGURE 9.4 – UCSRC

The transmitter and receiver are configured with the same data features as configured in this register for proper data transmission.

URSEL : USART Register select. This bit must be set due to sharing of I/O location by UBRRH and UCSRC.

UMSEL : USART Mode Select,

When 1 - Synchronous Operation

When 0 - Asynchronous Operation

UPM[0 :1] : USART Parity Mode, Parity mode selection bits. USBS : USART Stop Select Bit,

When 0- 1 Stop Bit

When 1 - 2 Stop Bits

UCSZ[0 :1] : The UCSZ[1 :0] bits combined with the UCSZ2 bit in UCSRB sets size of data frame i.e., the number of data bits. The table shows the bit combinations with respective character size.

4. UDR : (USART Data Register)

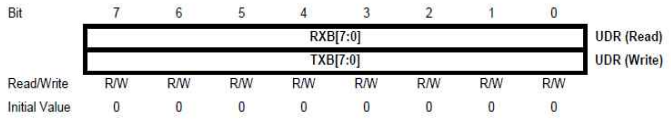

FIGURE 9.5 – UDR

The USART Data receive and data transmit buffer registers share the same address referred asUSART UDR register, when data is written to the register it is written in transmit data buffer register (TXB). Received data is read from the Receive data buffer register (RXB).

5. UBRRH and UBRRL (USART Baud Rate Registers)

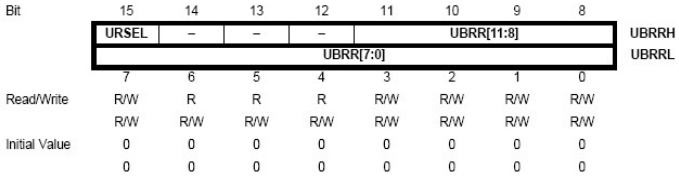

FIGURE 9.6 – UBRRH and UBRRL

The UBRRH register shares the same I/O address with the UCSRC register, The differentiation is done on the basis of value of URSEL bit. When URSEL=0 ; write operation is done on UBRRH register.

When URSEL=1 ; write operation is done on UCSRC register.

The UBRRH and UBRRL register together stores the 12-bit value of baud rate, UBRRH contains the 4 most significant bits and UBRRL contains the other 8 least significant bits. Baud rates of the transmitting and receiving bodies must match for successful communication to take place

APPENDIX C
CODES