UTIU
UNIVERSITÀ TELEMATICA
INTERNAZIONALE UNINETTUNO

جامعة حلوان
Helwan University

VISUALLY IMPAIRED ASSISTANT

Presented by

Mennatullah Sabry Saad — Mostafa Hussien Abdelhamed

Ahmed Ehab Ahmed — Mina Jouseph Mitry

Nourane Walid Abdelhakim — Karim Sayed Elkholy

Graduation Project Report in Partial Fulfillment of

Bachelor Degree in communication & IT Engineering at Helwan University, Cairo, Egypt

Bachelor Degree in communication & IT Engineering at Uninettuno University Rome, Italy

Egyptian Supervisor

Dr. Zaki B. Nossair

Italian Supervisor

Dr. Dario Assante

July 2019

AKNOWLEDGMENT

The success and final outcome of this project required a lot of guidance and assistance from many people and we are extremely privileged to have got this all along the completion of our project. All that we have done is only due to such supervision and assistance and I would not forget to thank them.

In a try to express our gratitude to our esteemed doctor Dr. Zaki B. Nossair who directed us in each step in this project, we're thankful for his guidance, patience, enthusiasm, motivation and generosity that kept providing us with the guidance we needed.

Mennatullah, Mostafa, Ahmed,Mina, Nourane, Karim

ABSTRACT

Living normally is the least right for all people. Also, everyone is looking for a private independent life without anyone's help.

Visually Impaired people face a big problem to make everything on their own, so they need an assistant to help them to overcome this problem, and that is why Visually Impaired Assistant has started.

Visually Impaired also need a separate device to perform many functions, like applications to recognize objects and read books. Here comes VIA role to be one device has all the possible ways to help them without using more than one device.

Visually Impaired Assistant is smartglasses which will help visually impaired people to get rid of all their problems and not need any help from any person inside or outside home.

TABLE OF CONTENTS

LIST OF TABLES

CONTACTS

Full name	Phone Contact	Email
Menntullah Sabry Saad	01091349619	menna.ssa95@gmail.com
Mostafa Hussien Abdelhamed	01060869842	mostafa.hussien.abdelhamed@gmail.com
Ahmed Ehab Ahmed	01121191888	ahmed359ehab@gmail.com
Mina Jousef Mitry	01222265278	minamitry218@gmail.com
Nourane Walid Abdelhakim	01023007471	nourane_walid95@yahoo.com
Karim Sayed Elkholy	01000081426	ksayed500@gmail.com

Table 1: contacts

LIST OF FIGURES

TABLE OF ACRONYMS AND DEFINITIONS

VIA	Visually Impaired Assistant
OCR	Optical Character Recognition
NLP	Natural Language Processing
STT	Speech To Text
TTS	Text To Speech
WHO	World Health Organization
ML	Machine Learning
DP	Deep Learning
NN	Neural Network
ANN	Artificial Neural Network
DNN	Deep Neural Network
RNN	Recurrent Neural Network
LSTP	Lebanon Science and Technology Park
GT	Google Translation
GNMT	Google Neural Machine Translation
IDE	Integrated Development Environment
API	Applicant Programming Interface
AI	Artificial Intelligence

KEYWORDS

- Natural Language Processing
- Machine Learning
- Voice Recognition
- Cross Platform
- Speech to Text
- Text to Speech

Chapter 1

Introduction

Chapter 1: Introduction

1.1 Problem Definition

Visually Impaired people face a big problem to live normally as they always need an assistant to do many functions in their lives they cannot do themselves.

First function they cannot do is that dealing with a device with buttons is so confusing as they forget the function of each button.

Second function is that they need to read books; the available books for their impairment are Braille-written books. But the problem is the book which is written in Braille is twice size the normal book and much expensive, also most of books are not found in Braille.

Third function is they cannot recognize objects that they are dealing with daily which makes interacting with area around them difficult.

Fourth function is that they need to take notes for their work, appointment or anything to remember them later which they cannot do.

The most common problem between Visually Impaired that they find a difficulty in using many devices as each device performs a separate function.

1.2 Problem Importance

Visually Impaired people find difficulty in communication and interaction normally, and that's makes their lives more complicated.

Also, as they always need an assistant, this makes their lives so public to the others which is something annoying to any person.

1.3 Current Solutions

- Mobile Applications
    - Color Detection Application
    - Money Detection Application
    - Ask for an assistant Application
- Braille System

1.4 Objectives

1. Construction a simple device that user can easily deal with such as using voice commands instead of buttons.
2. The device should be able to read books and recognizing objects.
3. The device should be able to record notes with voice.
4. Many features should be integrated in one device, so it can be used as a separate device.

1.5 Project Idea

Visually Impaired Assistant (VIA) is smartglasses that will be consisted of many features:

1. *Voice Control:* the device supports some functions with voice:

   - Taking commands from the user to use one feature from VIA's features.

   - Recording notes using voice and ability to play them back using specific command.

2. *Object Recognition:* the device recognizes objects using camera to tell the user its name.

3. *Optical Character Recognition*: the device supports two functions:

   - Take photo with camera and read it

   - Read a PDF file

1.6 Motivation and Justification

1.6.1 Motivation

Our motivation is to help as many as possible persons in their lives especially visually impaired peoples as they face a lot of problems in their daily lives. Also to create a device that contains many features as possible to be more helpful.

Another motivation is to create this device to be environmentally friendly and working with high performance at the same time.

These reasons motivated us to merge between speech recognition, object recognition, and optical character recognition techniques to build up a useful device to be helpful as much as possible.

1.6.2 Educational Value

Working on VIA has added new values not only to our technical overview, but also to our non-technical overviews that has helped us as a team and as individuals.

1.6.2.1 Technical Value

- NLP technology and its contributions.
- STT and TTS technologies and how they work.
- Voice Recognition technology and how to tackle it.

1.6.2.2 Non-technical Value

- Idea generation techniques
- Writing a business plan
- Work division and integration
- Time Management
- Documenting and presenting a large-scale project

1.7 Domain of Application

VIA is an independent device on its own that can be used in some specific fields such as:

- *Entertainment:* users can enjoy reading and recognizing objects
- *Education:* visually impaired people can use the device in their schools or colleges to find, read and learn their academic curriculum.

1.8 Document Overview

In this document we are trying as much as possible to illustrate each and every step we have been through since we started working on **VIA**. Here is the outline of this document:

Chapter 2: A market survey about the idea and the other competitors in the market.

Chapter 3: Explaining any necessary background needed throughout the whole project.

Chapter 4: Stating the system design, features provided and system specification.

Chapter 5: The tools and environment used in our implementation.

Chapter 6: Summarizing our findings and what we will do as future work.

Chapter 2

Market Survery

Chapter 2: Market Survey

2.1 Overview

This chapter to declare how we were able to find the best way to help the visually impaired we had to focus on some points to put our hands in the core of the problem and help them well.

2.2 History & Statistics of the Problem

The World Health Organization (WHO) estimates the number of people with visual impairments globally, based on recent studies, that some 285 million people with visual impairment (blindness or visual impairment) worldwide, including 246 million people with visual impairment and 39 million people are infected Blindness.

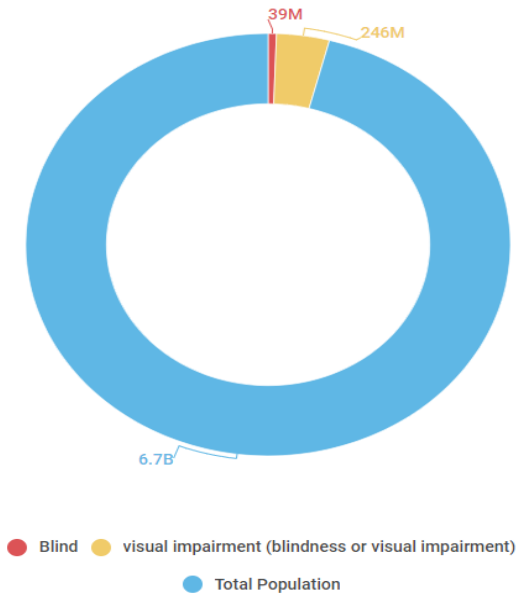


Figure 2-1: Visually impaired over the world

NEW YORK (Reuters Health) - Blind people will triple from about 39 million to 117 million in 2050 as the population expands and increases, according to a recent study conducted by US researchers.

According to the US medical site "MedicalXpress", the number of people with moderate to severe visual impairment and those who wore glasses or contact lenses or performed a process would triple, from about 246 million to 688 million during the same period.

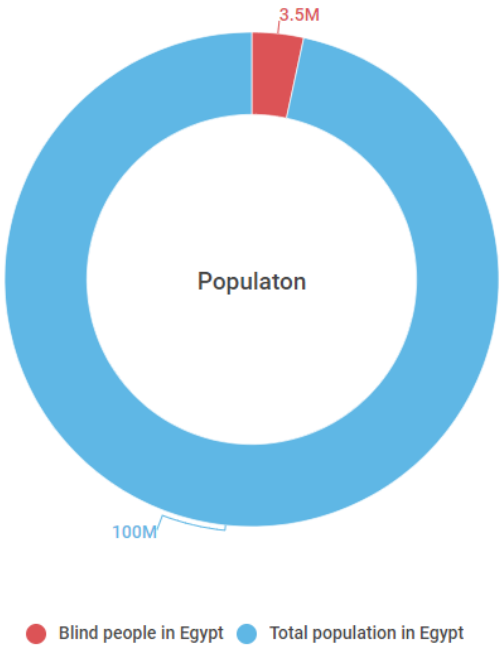


Figure 2-2: Visually impaired in Egypt

When we had a look to the statics of published books, we found that the number of books published annually in the Arab world is estimated at 5000 books, of

which only about 3% are available to the blind in an unorganized manner, compared to 10% in other countries of the world, according to UNESCO. Locally, the number of blind people in Egypt is estimated at 3.5 million

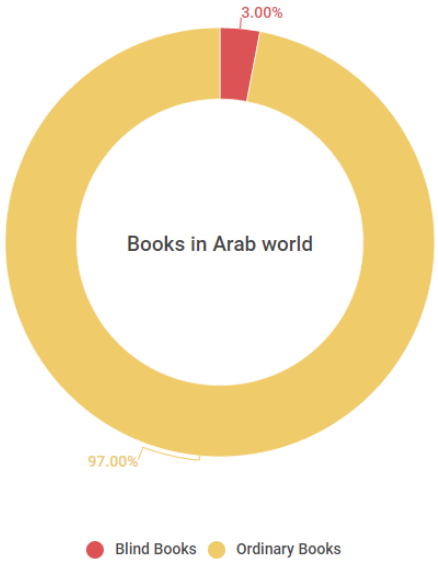


Figure 2- 3: Visually impaired books in arab world

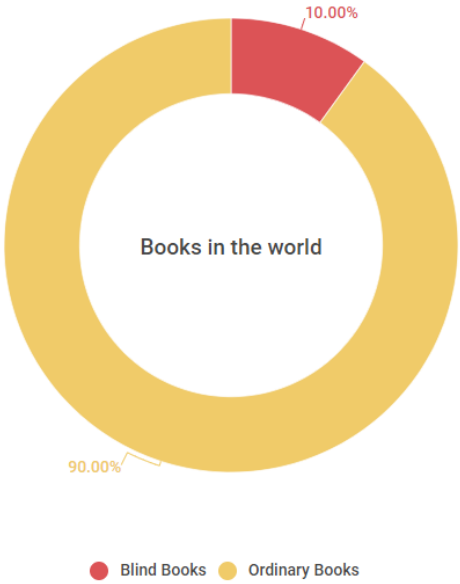


Figure 2-4: Visually impaired books over the world

Also, on survey on blind people asking them what is the best way to deal with ATMs or any other electrical device the result was according to the following graph:

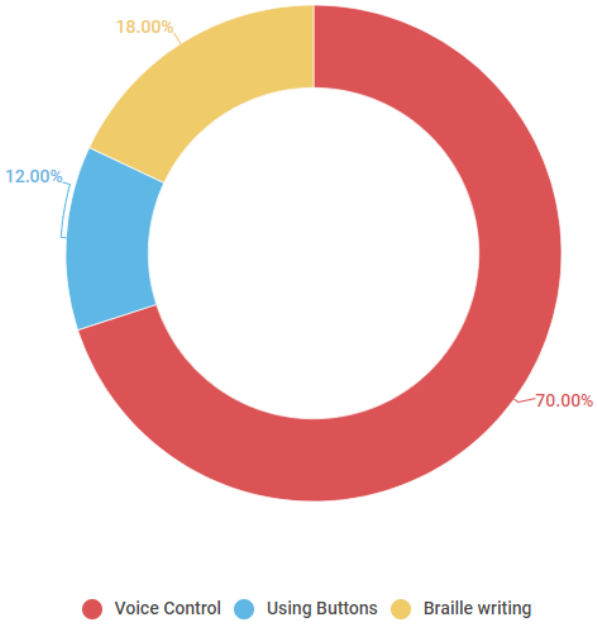


Figure 2-5: Voice Control, Using buttons and Braille writing

This previous result makes I controlled by voice as much as possible, this will be discussed later.

2.3 INTENDED CUSTOMERS

Our project mainly targets two types of customers:

- *Visually Impaired People:* helping them find, read and learn books easily without needing to know any braille language or needing someone to read for them, help them to discover the environment around them and

detect objects. **VIA** will simply make it easy for them to do these features on their own.

- *Ordinary People:* who mostly enjoy reading but easily get bored of holding books or reading softcopies either on their laptops or their mobile devices or those who enjoy listening to stories rather than reading them or for those who couldn't read. **VIA** will make reading much faster, easier and enjoyable for them.

2.4 COMPETITOR ANALYSIS

2.4.1 Be My Eyes

Be My Eyes is a free application that connects completely and partially visually impaired people with sighted volunteers and company representatives for visual assistance through a live video call.

Figure 2-6: Be My Eyes

2.4.2 Aira

Aira supports transformative remotely assistance technology that connects the visually impaired with a network of certified workers via wearable smart glasses and an enhanced reality dashboard that allows workers to see what the visually impaired person sees in real time.

Figure 2-7: Aira

Workers, serving as visual interpreters for the visually impaired, help users accomplish a wide range of daily tasks and activities - from navigating busy streets to recognizing faces and literally traveling the world.

2.5 WHAT IS UNIQUE ABOUT VIA?

Of course there were many previous projects serving visually impaired people, and there are many competitive edges between our project and the competitors as they have some services in common but our project is unique in some point:

- VIA is a separable individual device that has many features collected in one device so the user will not need any devices.

- VIA supports more features, which will talk about later, like object recognition, reader, recording notes and voice control over all these features which solves a lot of visually impaired people's problem.

- One of the most important unique points about VIA that the user will keep his life private as he will not need anyone's help.


Chapter 3
Necessary Background

Chapter 3: Necessary Background

3.1 Overview

In this chapter, we are going to discuss the scientific background of the software and the methods used in understanding the question to get most accurate results.

This chapter discusses the algorithms and techniques we used to build our application. We will illustrate the different backgrounds we used in building our project as: Image Processing, Deep Learning, and Machine Learning.

3.2 Digital Image Processing

In computer science, digital image processing is using image processing on digital images. Digital image processing has a lot of advantages on analog image processing. It allows many algorithms to be applied to the input data and can avoid problems such as the build-up of noise and signal distortion during processing. Since images are outlined over two dimensions (perhaps more) digital image processing may be modeled in the compose of multidimensional systems.

3.2.1 Tasks of Digital Image Processing

Digital image processing deal with:

- Classification
- Feature Extraction
- Multi-scale signal analysis

- Pattern recognition
- Projection

Techniques are used in image processing:

- Image restoration
- Independent component analysis
- Linear filtering
- Neural networks
- Partial differential equations
- Pixelation
- Point feature matching
- Principal components analysis
- Self-organizing maps

3.3 Deep Learning

Deep learning (also detected as deep hierarchical learning or structured learning) is a section of a fringe family of machine learning methods based on ANNs. Learning can be unsupervised, supervised or semi-supervised
Deep learning architectures such as deep belief networks, deep neural networks, convolutional neural networks and recurrent neural networks have been applied to fields including speech recognition, computer vision , audio recognition ,natural language processing, bioinformatics ,social network filtering, drug design, machine translation, material inspection, board game programs and medical image analysis where they have generated results comparable to and in some cases superior to individual scientist.
Artificial Neural Network (ANN) was motivated by information processing and distributed communication nodes in biological systems. ANN have many differences from biological brains. Specifically, neural networks tend to be symbolic and static, while the biological brain of most living organisms is analog and dynamic.

3.2.1 Application of deep learning

3.3.1.1 Automatic Speech Recognition

The automatic speech recognition is the first and most convincing successful case of deep learning. LSTM Recurrent Neural Networks can learn "Very Deep Learning" tasks that involve many-second intervals including speech events separated by many discrete time steps, that one time step corresponds to about 10 ms. LSTM with forget gates is competitive with traditional speech recognizers on many tasks.
The original success in speech recognition is based on small-scale recognition tasks based on TIMIT. The data set includes 630 speakers from eight major dialects of American English, where every speaker reads ten sentences .Its small size helps many configurations be tried. Most importantly, the TIMIT task contains phone-sequence recognition, which, is not the same as word-sequence recognition, allows weak phone bigram language models. This helps that the strength of the acoustic modeling aspects of speech recognition be analyzed in an easy way.

3.3.1.2 Image Recognition

A mutual evaluation set for image classification is the MNIST database data set. MNIST is composed of handwritten digits and includes sixteen thousands training examples and 10,000 test examples. As with TIMIT, its small size lets actors test many configurations. A comprehensive list of results on this set is available.
Deep learning-based image recognition has become "superhuman"; manufacture more valid results than human contestants. Deep learning-trained vehicles now interpret 360° camera views.

3.3.1.3 Virtual Art Processing

Closely concerning to the progression that has been made in image recognition is the increasing application of deep learning techniques to different visual art tasks. Deep Neural Networks have proven themselves capable
a) Identifying the form period of a given painting
b) Neural Style Transfer - capturing the form of a given artwork and applying it in a visually pleasing method to an arbitrary photograph or video

c) Producing striking imagery based on random visual input fields

3.3.1.4 Natural Language Processing

Neural networks have been used for performing language models since the early 2000s. LSTM helped to ameliorate language modeling and machine translation Other key techniques in this domain are negative sampling and word embedding. Word embedding, such as word2vec, can be thought of as a representational layer in a deep learning architecture that converts an single word into a positional representation of the word related to other words in the dataset; the position is acted as a point in a vector space. Using word embedding as an Recurrent Neural Network input layer allows the network to parse sentences and phrases using an effective compositional vector grammar. A compositional vector grammar can be thought of as probabilistic context free grammar (PCFG) implemented by a Recurrent Neural Network . Recursive auto-encoders construct atop word embedding can assess sentence similarity and detect paraphrasing. Deep neural architectures provide the best results for information retrieval, spoken language understanding, writing style recognition, machine translation, constituency parsing, sentiment analysis, contextual entity linking, and Text classification.
Recent developments popularize word embedding to sentence embedding. Google Translate (GT) uses a big end-to-end long short-term memory network. Google Neural Machine Translation (GNMT) uses a translation method of an example-based machine in which the system "learns from millions of examples. It translates the entire sentence rather than parts of the sentence. Google Translate provide over 100 languages.

3.3.1.5 Drug Discovery and Toxicology

A big percentage of candidate drugs fail to win regulatory approval. Insufficient efficacy (on-target effect), unanticipated toxic effects or undesired interactions (off-target effects) are causes of these failures .Research has investigated use of deep learning to predict off-targets, toxic effects and the biomolecular targets of environmental chemicals in nutrients, household products and drugs.

3.3.1.6 Customer Relationship Management

Deep reinforcement learning is used to approximate the value of possible direct marketing actions, defined in terms of RFM variables. The estimated value function was presented to have a natural interpretation as customer lifetime value.

3.3.1.7 Recommendation System

A recommendation system is used deep learning to extract meaningful features for a latent factor model for content-based music recommendations. Many views deep learning has been applied for learning user preferences from multiple fields. The model uses a mixture collaborative and content-based approach and improves recommendation in many tasks.

3.3.1.8 Bioinformatics

An auto encoder Artificial Neural Network was used in bioinformatics, to guess gene ontology annotations and gene-function relationships .In medical informatics, deep learning was used to guess low quality based on data from wearables and guesses of health complications from electronic health record data. Deep learning has also displayed efficacy in healthcare.

3.3.1.9 Medical Image Analysis

Deep learning has been displayed to generate additional results in medical application such as lesion detection, organ segmentation , image enhancement and cancer cell classification.

3.3.1.10 Mobile Advertising

Finding the appropriate mobile audience for mobile advertising is always challenging, since a lot of data points must be considered and assimilated before a target segment can be made and used in ad serving by any ad server. Deep learning has been used to interpret large, multiple dimensioned advertising

datasets. A lot of data points are collected during the request/serve/click internet advertising cycle. This information can make the basis of machine learning to enhance ad selection.

3.3.1.11 Image Restoration

Deep learning has been successfully applied to inverse problems such as super-resolution imprinting, film colorization and denoising. These applications include learning methods such as Deep Image Prior which trains on the image that needs restoration and Shrinkage Fields for Effective Image Restoration which trains on an image dataset.

3.3 Neural Network

3.4.1 Artificial Neural Network

Artificial neural network (connectionist systems) is computing systems motivated by the biological neural network that constitute animal brains. Such systems learn (progressively improve their ability) to make tasks taking into consideration examples, generally without task-specific programming.
For example, in image recognition, they might learn to detect images that have cats by analyzing example images that have been manually categorized as no cat or cat and using the analytic results to detect cats in other images.
They have found most use in applications difficult to express with a classic computer algorithm using rule-based programming.
An ANN is based on a combination of connected units called artificial neurons. Each connection (synapse) between neurons can transfer a signal to another neuron. The receiving (postsynaptic) neuron can process the signal and then signal downstream neurons connected to it. Neurons may have state, generally represented by real numbers, typically between zero and one. Neurons and synapses may have a weight also that differ as learning proceeds, that can decrease or increase the strength of the signal that it sends downstream.

Typically, neurons are well-planned in layers. Various layers may perform different kinds of transformations on their inputs. Signals transmit from the first (input), to the last (output) layer, possibly after traversing the layers multiple times.
The original objective of the neural network approach was to solve problems in the same way that a human brain solve it. Over time, attention focused on matching specific mental abilities, leading to deviations from biology such as passing information in the reverse direction or back-propagation and adjusting the network to reflect that information.
As of 2017, neural networks typically have a range from thousand to million units and millions of connections. Regardless of this number being several order of magnitude less than the number of neurons on a human brain, these networks can perform a lot of tasks at a level beyond that of humans

3.4.2 Deep Neural Network

A deep neural network is an artificial neural network with many layers between the first (input) and last (output) layers. The DNN finds the right mathematical handling to turn the input into the output, so it can be a non-linear relationship or a linear relationship. The network moves between the layers calculating the probability of every output. For example, a deep neural network that is drilled to recognize dog breeds will go over the given image and calculate the probability that the dog in the image is a certain breed. The actor can check the results and select which probabilities that the network should display (above a certain threshold, etc.) and return the chosen label. Each mathematical manipulation as such is considered a layer, and complex DNN have multiple layers.
DNN can pattern complex non-linear relationships. DNN architectures produce compositional models where the object is expressed as layered composition of primitives. The extra layers enable composition of features from the further downs layers, potentially modeling complex data with fewer units than a similarly performing shallow network.
Deep architectures contain many different of a little basic approach.
Every architecture has found success in specific fields. It is not always possible to compare the performance of much architecture, except if they have been evaluated on the same data sets.
DNNs.0 are typically feedforward networks in which data flows from the first (input) layer to the last (output) layer without returning back. At first, the DNN creates a map of virtual neurons and assigns random weights, or numerical values, to connections between them. The weights and inputs are multiplied and return an output between zero and one. If the network didn’t carefully recognize a

particular pattern, an algorithm would correct the weights. That way the algorithm can make many parameters more influential, until it determines the right mathematical manipulation to fully process the data.

3.3 Machine Learning

It is the study of statistical models and algorithms that computer systems use to implement a special function effectively without using explicit instructions, relying on patterns and inference instead. It is displayed as a subset of AI. Machine learning algorithms construct a mathematical model based on training data, known as "sample data", in order to make guesses or decisions without being explicitly programmed to implement the function. Machine learning algorithms are used in a large variety of applications, such as computer vision and email filtering where it is infeasible to develop an algorithm of specific instructions for implementing the task. Machine learning is relative to computational statistics, which focuses on applying predictions using computers. The study of mathematical optimization delivers theory, application domains and methods to the domain of machine learning. Data mining is a domain of study within machine learning, and focuses on exploratory data analysis through unsupervised learning. Machine learning is also referred to as predictive analytics in its application across business problems.

3.5.1 Types of Learning Algorithms

3.5.1.1 Supervised Learning

Supervised learning algorithms construct a mathematical model of a set of data that include both the inputs and the desired outputs the data is known as sampled data, and consists of a set of training examples. Each training example has one or many inputs and a desired output, also known as a supervisory signal. In the mathematical model, each training example is displayed by a vector or array, at sometimes it is called a feature vector, and the training data is displayed by a matrix. Through reduplicate optimization of an objective function, supervised learning algorithms learn a function that can be used to guess the output associated with new inputs. An optimal task will allow the algorithm to

correctly determine the output for inputs that were not a piece of the training data. An algorithm that enhances the accuracy of its outputs or predictions over time is said to have learned to implement that task.
Similarity learning is a wide range of supervised machine learning relative to regression and classification, but the objective is to learn from examples using a common function that measures how similar or related two objects are. It has applications in recommendation systems, visual identity tracking, and face verification and ranking.

3.5.1.2 Reinforcement learning

Reinforcement learning is a wide range of machine learning included the way that the software agents have to take actions in an environment so as to increase some notion of additional reward. The domain is studied in many other disciplines, such as control theory, game theory , information theory, operations research, multi-agent systems, simulation-based optimization, statistics , genetic algorithms and swarm intelligence due to its generality.
In machine learning, the environment is typically displayed as a Markov Decision Process Multiple reinforcement learning algorithms use dynamic programming techniques. Reinforcement learning algorithms do not assume knowledge of an typical mathematical model of the Markov Decision Process , and are used when typical models are infeasible. Reinforcement learning algorithms are used in autonomous vehicles or in learning to play a game versus a human opponent.

3.5.1.3 Feature learning

Many learning algorithms have the goal of discovering better representations of the inputs provided during training. Classic examples contain principal components analysis and cluster analysis. Feature learning algorithms, also known as representation learning algorithms, always fails to preserve the information in their input but also convert it in a way that makes it useful always as a pre-processing step before implementing classification or predictions. This technique help in reconstruction of the inputs coming from the unknown data-generating distribution, while not being necessarily loyal to configurations that are improbable under that distribution.

3.5.1.4 Decision trees

Decision tree learning uses a decision tree as a predictive model to go to conclusions about the target value of the item (represented in the leaves) from observations about an item (represented in the branches). It is one of the predictive modeling approaches used in data mining, machine learning and statistics. Tree models where the target variable can take a discrete set of values are called classification trees; in these tree structures, leaves represent class labels and branches represent conjunctions of features that direct to those class labels. Decision trees where the target variable can take continuous values (typically real numbers) are called regression trees. In decision analysis, a decision tree can be used to show and explicitly display decisions and decision making.

3.5.1.6 Support vector machines

Support vector machines (SVM), also defined as support vector network, are a set of combined supervised learning methods used for classification and regression. Given a set of training examples, every marked as belonging to one of two groups , an Support Vector Machine training algorithm constructs a model that guess if a new example falls into one group or the other. An Support vector machines training algorithm is a non-probabilistic, binary, linear classifier, although methods such as Platt scaling exist to use Support vector machines in a probabilistic classification setting. Support vector machines can efficiently perform a non-linear classification using what is called the kernel trick, implicitly mapping their inputs into high-dimensional feature spaces to performing linear classification.

3.5.1.7 Bayesian networks

Bayesian network, belief network or directed acyclic graphical model is a probabilistic graphical model that displays a group of random variables and their conditional independence with a directed acyclic graph (DAG). For example, a Bayesian network could display the probabilistic relationships between symptoms

and sickness. Given symptoms, the network can be used to compute the probabilities of the presence of different sickness. Efficient algorithms have an existence that implements inference and learning. Bayesian networks that model sequences of variables, like speech signals or protein sequences are known as dynamic Bayesian networks. Generalizations of Bayesian networks that can display and solve decision problems under uncertainty are known as influence diagrams.

3.6 Classification

Classification techniques in data mining have the ability of processing a large amount of data. It can be used to predict a set of class labels and classifies data based on training set and class labels and it can be used for classifying a new available data. This definition could cover any context in which some decision or forecast is made on the basis of existence available information. Classification procedure is a detected method for repeatedly making such decisions in new situations. Here if we suppose that problem is a concern with the construction of a procedure that will be applied to a continual sequence of cases in which each new case must be assigned to one of a set of predefined classes on the basis of observed features of data. Generation of a classification procedure from a set of data for which the exact classes are defined in advance is termed as pattern recognition or supervised learning. Contexts in which a classification function is essential include, for example, assigning individuals to credit status on the basis of financial and other personal information, and the initial diagnosis of a patient's disease in order to select immediate treatment while awaiting right test results. Many of the most serious problems arising in industry, commerce and science can be named as classification or decision problems. Three main historical strands of research can be identified as statistical, machine learning and neural network. All groups have some objectives in common. They have all attempted to create procedures that would be able to handle a lot of problems and to be extremely general used in practical settings with proven success.

3.6.1 Algorithms of classification technique

3.6.1.1 K-nearest Neighbors Algorithm

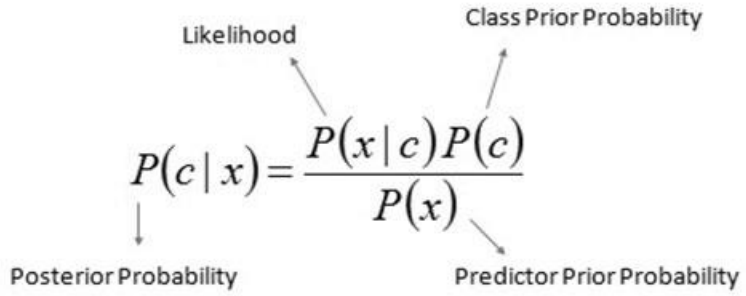
The closest neighbor (Nearest Neighbors) rule distinguishes the classification of unknown data point on the basis of its closest neighbor whose class is already known. M. Cover and P. E. Hart purpose k nearest neighbor (K Nearest Neighbors) in which nearest neighbor is computed on the basis of estimation of k that indicates how many nearest neighbors are to be considered to characterize class of a sample data point. It makes employment of the more than one closest neighbor to determine the class in which the given data point belongs to and and consequently it is called as K Nearest Neighbors. These data samples are needed to be in the memory at the run time and hence they are referred to as memory-based technique. T. Bailey and A. K. Jain enhance K Nearest Neighbors which is concentrated on weights. The training points are allocated weights depended on their distances from sample data point. But at the same time the computational complexity and memory requirements vestige the primary concern dependably. To overcome memory limitation size of data set is decreased. For this the repeated patterns which don't include extra data are also discarded from training data set. To further improve the information focuses which don't influence the result are additionally deleted from training data set. The nearest neighbor training data set can be organized utilizing various systems to improve over memory limit of k nearest neighbor . The k nearest neighbor implementation can be done using ball tree, k-d tree, nearest feature line (Nearest Feature Line), principal axis search tree and orthogonal search tree. The structure training data of the tree is further divided into nodes and techniques like Nearest Feature Line and tunable metric divide the training data set depending on planes.

3.6.1.2 Naïve Bayes Algorithm

The Naive Bayes Classifier technique is based on Bayesian theorem and is particularly used when the dimensions of the inputs is high. The Bayesian Classifier is able to calculate the most possible output based on the input. It is also possible to add new raw data at runtime and have a improved probabilistic classifier. A naive Bayes classifier considers that the existence (or absence) of a particular attribute of a class is not related to the existence (or absence) of any other feature when the class variable is given. As an example, a fruit may be assigned to be an apple if it is red, round. Although if these features depend on

each other or upon the presence of other features of a class, a naive Bayes classifier considers all of these properties to not dependent contribute to the probability that this fruit is an apple.

Calculating the posterior Probability as follows:



$$P(c \mid \mathrm{X}) = P(x_1 \mid c) \times P(x_2 \mid c) \times \cdots \times P(x_n \mid c) \times P(c)$$

3.6.1.3 Adaptive decision boundary

A decision boundary or decision surface is a hyper surface that split the underlying vector space into two groups, each class has one group In a statistical-classification problem with two classes. The classifier will classify all the points at one side of the decision boundary as belonging to one class and all those data on the other side as belonging to the other class.

3.7 Regression

Regression analysis is a set of statistical processes for estimating the relationships between variables in statistical modeling, . It contains a lot of techniques for modeling and analyzing many variables, when the constraint is on the relationship between a dependent variable and one or more independent variables (or 'predictor'). More particularly, regression analysis aids one understand how the typical value of the dependent variable (or 'criterion variable') changes when any one of the independent variables is varied, while the other independent variables are held fixed.
Most mutually, regression analysis estimates the subjunctive expectations of the
non -independent variable given the independent variables – that is, the average value of the non -independent variable when the independent variables are fixed. Less mutually, the constrain is on a quantile, or another location parameter of the conditional allocation of the dependent variable given the independent variables. A function of the independent variables called the regression function is to be estimated in all cases. It is also of interest to characterize the variation of the dependent variable around the prediction of the regression function using a probability distribution in regression analysis, . A relative but different approach is Necessary Condition Analysis (Necessary Condition Analysis), which estimates the maximum (better than average) value of the dependent variable for a given value of the independent variable (ceiling line rather than central line) in order to identify what value of the independent variable is necessary but not sufficient for a given value of the dependent variable.
Regression analysis has a use for forecasting and prediction where its use has fundamental overlap with the domain of machine learning. Regression analysis is also used to understand which between the independent variables are related to the dependent variable, and to develop the forms of these relationships. Regression analysis can be used to infer causal relationships between the independent and dependent variables in restricted circumstances . So this can lead to illusions or artificial relationships, so caution is advisable.
A lot of techniques for carrying out regression analysis have been created. Mutual methods such as linear regression and ordinary least squares regression are parametric, in that the regression function is defined in terms of a finite number of unknown parameters that are predicted from the data. Nonparametric regression indicates to techniques that help the regression

function to lie in a specified group of functions, which may be infinite-dimensional.
The performance of regression analysis methods in practice depends on the style of the data producing process, and the way that it relates to the regression approach being used. Since the true form of the data-generating process is generally not known, regression analysis often depends to some extent on making assumptions about this process. These assumptions are always testable if a better quantity of data is available.

Chapter 4

Design Specifications

Chapter 4: Design Specifications

4.1 Feature Provided

4.1.1 Voice Interaction and Understanding

The system shall recognize user's speech and understand speech intent.

4.1.2 Voice Control

The system shall work according to the user's voice commands.

4.1.3 Recording Voice Note

The system shall record user's speech and save it.

4.1.4 Object Recognition

The system shall detect and recognize the objects in an image.

4.1.5 Optical Character Recognition

The system shall convert an image to text using two methods:

4.1.5.1 Using Camera

The system shall open the camera and capture an image to read

4.1.5.2 Searching in database

The system shall search in database by book's name and read it to user

4.2 Context Diagram

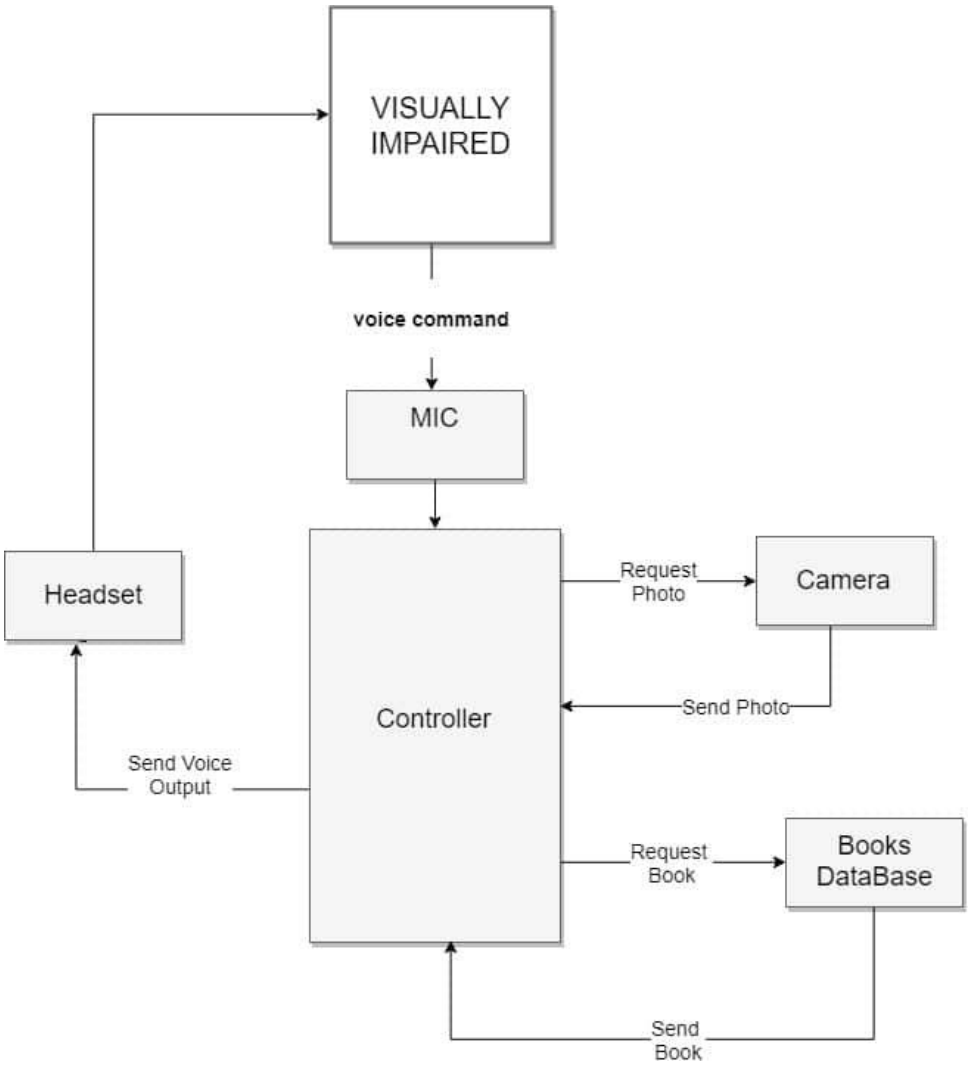


Figure 4-1: Context Diagram

4.2.1 Overview

The diagram shows how the user will manage the operation by inserting command using voice, to perform one of the three functions whether to recognize objects, read a book, or record a voice note.

4.3 Data Flow Diagram

4.3.1 Speech Recognition

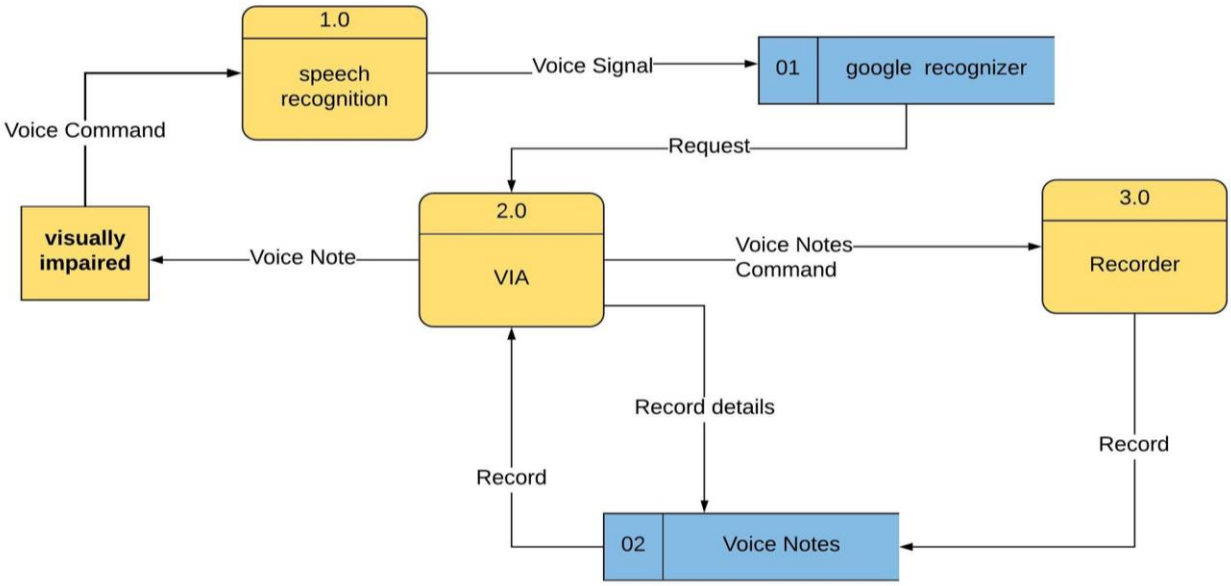


Figure 4-2: Data Flow-1

4.3.2 Optical Character Recognition

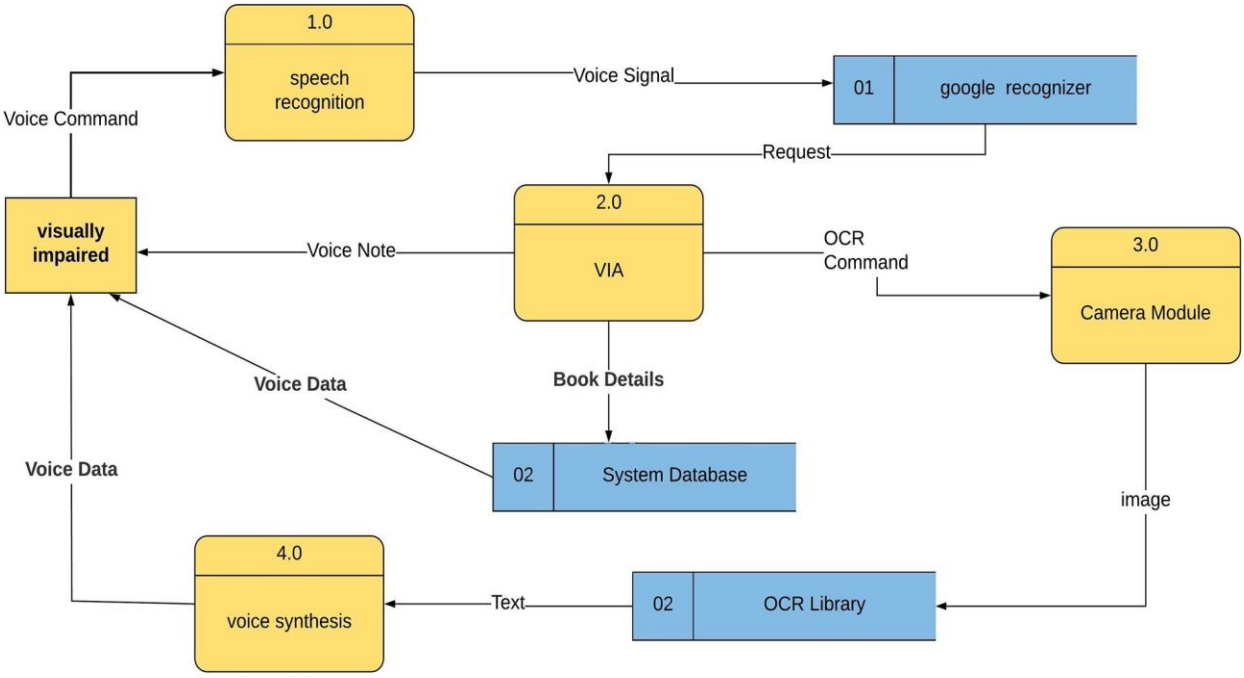


Figure 4-3: Data Flow-2

4.3.3 Object Recognition

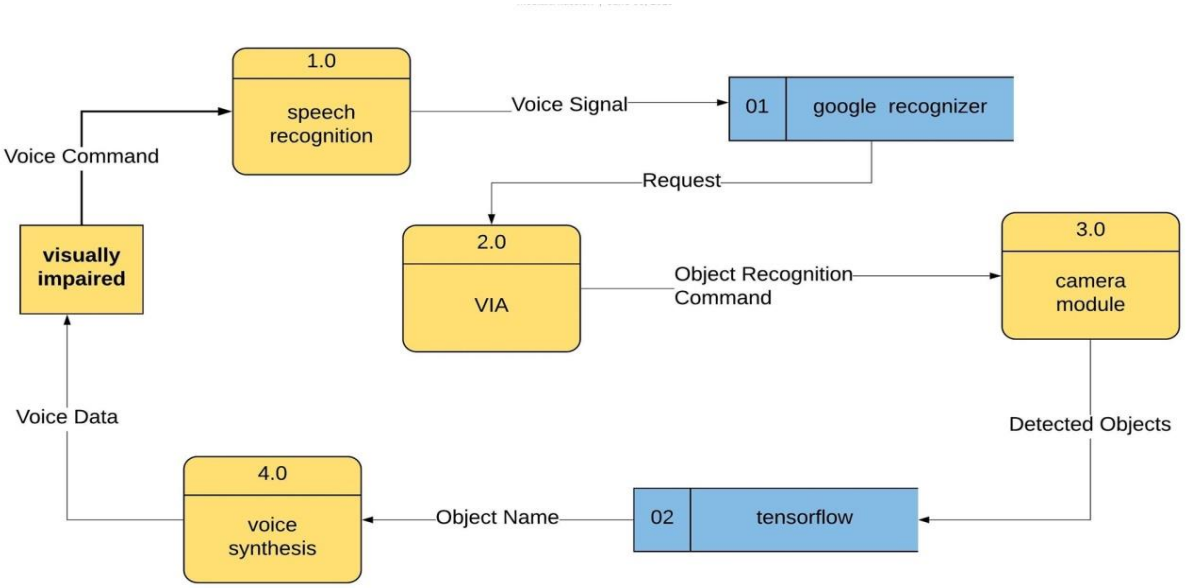


Figure 4-4: Data Flow-3

4.4 Sequence Diagram

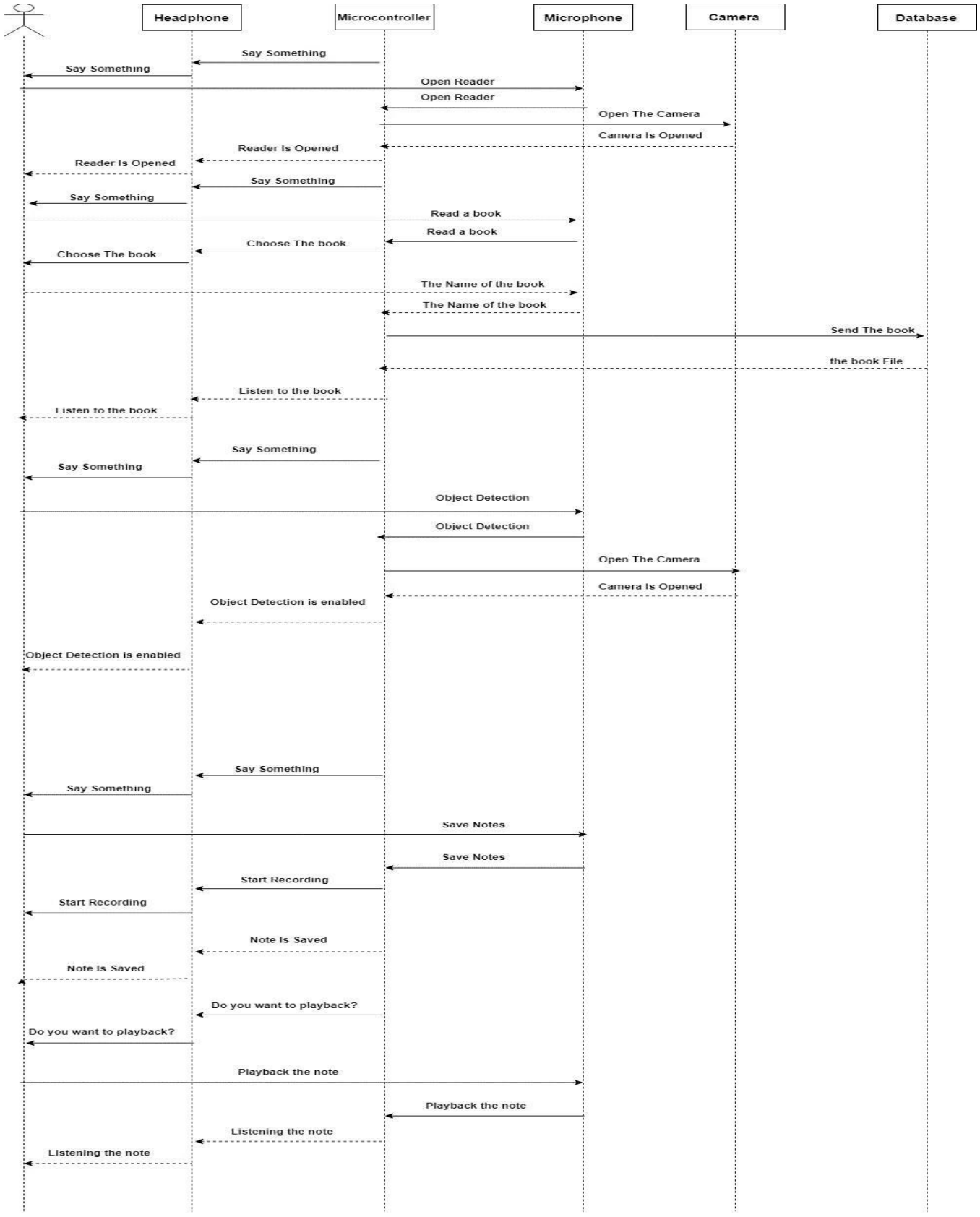


Figure 4-5: Sequence Diagram

4.5 Use Case Diagram

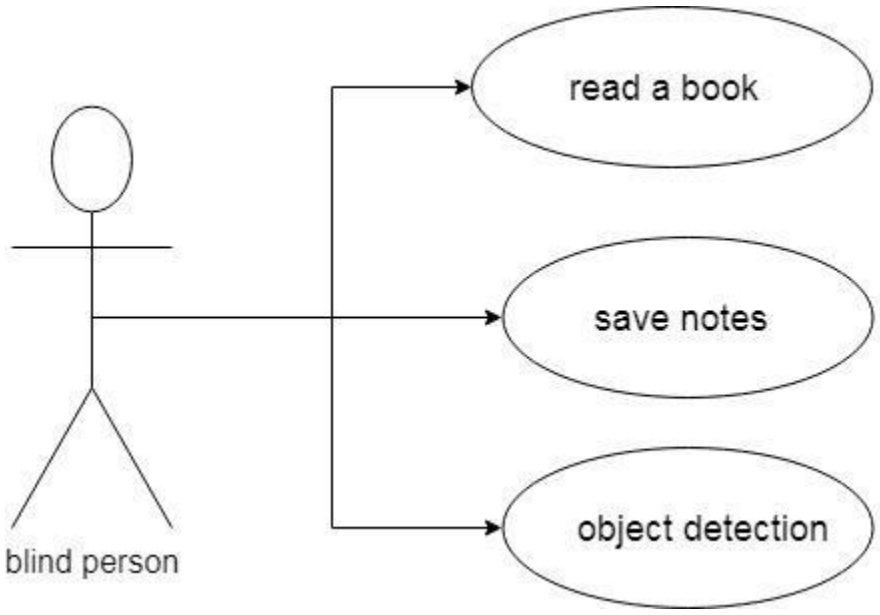


Figure 4-6: Use Case Diagram

Use Case ID	UC-1
Use Case Name	Save notes
Actor	Visually Impaired
Preconditions	Say save note
Post-conditions	Listen to the saved note
Normal flows	1. The visually impaired choose save note command 2. The device start saving the voice note 3. The visually impaired can listen to the saved notes

Table 3: Use Case 1

Use Case ID	UC-2
Use Case Name	Reader-1
Actor	Visually Impaired
Preconditions	Say open reader
Post-conditions	Listen to the text exist on image
Normal flows	1. The visually impaired choose open reader command 2. The device take an image 3. Convert photo to text 4. Convert text to voice 5. The visually impaired will listen to the text on the taken image

Table 4: Use Case 2

Use Case ID	UC-3
Use Case Name	Reader-2
Actor	Visually Impaired
Preconditions	Say search for the book
Post-conditions	Listen to the book
Normal flows	1. The visually impaired choose search for book command 2. The device searches in database for book 3. The visually impaired listen to the book that searched for

Table 4: Use Case 3

Use Case ID	UC-4
Use Case Name	Object Recognition
Actor	Visually Impaired
Preconditions	Say recognize object
Post-conditions	Listen to the recognized objects
Normal flows	1. The visually impaired choose recognize object command 2. The device turn on camera 3. Start recognition objects 4. The visually impaired listen to the name of object in front of him

Table 5: Use Case 4

4.6 Functional Description

4.6.1 Speech Recognition

The system shall work according to the user's voice commands and save notes if it is commanded.

4.6.1.1 Classify command

The system shall receive the user command and classify it to take an action.

4.6.1.2 Convert Speech to Text

The system shall convert a speech to text.

4.6.1.3 Record Voice

The system shall record a voice note using microphone and save it in database.

4.6.1.4 Send voice

The system shall send an output voice to be heard by the user.

4.6.2 Object Recognition

The system shall detect and recognize the objects in an image.

4.6.2.1 Capture an image

The system shall open the camera and capture an image.

4.6.2.2 Detect Objects

The system shall apply the image to image processing algorithms to detect objects in the image.

4.6.2.3 Recognize Objects

The system shall apply the objects to neural network classifier with many layers to recognize the objects in the image and return the recognized object as a text.

4.6.2.4 Send Voice Output

The system shall take the object name and convert it into speech and then send it as an output.

4.6.3 Optical Character Recognition

The system shall read books for user by using optical character recognition.

4.6.3.1 Detect Characters

The system shall apply the image to image processing algorithms to detect characters in the image.

4.6.3.2 Convert image to text

The system shall apply neural network classifier to convert the text in the image to text can be read by the device.

4.6.3.3 Search for books

The system shall search for the needed book in the database containing lists of books.

4.6.3.4 Convert PDF to text

The system shall convert PDF file into text

4.6.3.5 Read the text

The system will convert text to speech and send it as an output.

4.4 Non-functional Description

4.4.1 Easiness

The device is user friendly and easy to use; it will provide the user the steps in order to make the best use of the device. In addition, any feature in the device will not exceed two commands from the user to perform.

4.4.2 Response Time

The device should respond as fast as possible with delay a very few seconds after the user performs one of the system actions.

4.4.3 Cost

The device price will be affordable to most of categories as much as possible.

4.4.4 Size

The size of the device will be small like normal glasses to be suitable for 24/7 usage.

4.4.5 Efficiency

The device efficiency is higher as much as possible to be working accurate with user as it is a daily-use device.

Chapter 5

Development Tools and Environment

Chapter 5: Development Tools and Environment

5.1 Overview

This chapter presents the tools and programs that have been used in building **VIA**. It starts with the IDEs used. Then, the libraries used and their importance in the project.

5.2 IDEs

5.2.1 PyCharm

Figure 5-1: PyCharm

PyCharm is an IDE (Integrated Development Environment) which is used in computer programming, especially for python language. It wass developed by Czech company JetBrains.
It is a cross-platform IDE which works on Windows, MacOS and many Linux versions. It provides a graphical debugger, code analysis, An integrated unit tester; integration with version controls systems and supports web development through Django library.

5.3 LIBRARIES AND FRAMEWORKS

5.3.1 TensorFlow

TensorFlow is an open-source free software library for dataflow, and differentiable programming across a range of tasks.

Figure 5-2: TensorFlow

It is a symbolic math library. Also is used for machine learning applications such as: neural networks.

It is used for research also for production at Google.

TensorFlow was developed for Google internal use by their brain team.

Role in the project: used to recognize objects

5.3.2 OpenCv

OpenCV (Open Source Computer Vision Library) is an open-source software library in machine learning and computer vision. OpenCV was built up to support a common infrastructure for computer vision applications, and to accelerate the use of machine perception in the commercial products.

Figure 5-3:OpenCV

The library has a lot of algorithms reaching to more than 2500 optimized algorithms, which includes: a comprehensive set of both classic and state-of-the-art machine learningand computer vision algorithms.

These algorithms can be used to perform many functions like to detect and recognize faces, classify human actions in videos, track camera movements, identify objects, track moving objects, extract 3D models of objects, produce 3D point clouds from stereo cameras, find similar images from an image database,

stitch images together to produce a high resolution image of an entire scene.

Role in the project: apply image processing techniques on the captured images.

5.3.3 PyTesseract

Python-tesseract is an optical character recognition (OCR) appliance for python. That is, it will recognize and “read” the text typed in images. Pytesseract is a wrapper for Google’s Tesseract-OCR Engine. It is also useful as a stand-alone invocation script to tesseract, as it can read all image types provided by the Python Imaging Library, including jpeg, png, gif, bmp, tiff, and others, whereas tesseract-ocr by default supports only tiff and bmp.

Figure 5-4: PyTesseract

In addition if used as a script, Python-tesseract will print the recognized text instead of writing it to a file.

Role in the project: extract text from an image.

5.3.4 Tesseract

Tesseract is an engine of optical character recognition for various operating systems. It is available for Windows, Linux and Mac OS X.

It is free charge software that released under the Apache License, Version 2.0 and development has been sponsored by Google since 2006. In 2006, Tesseract was considered one of the most accurate open source OCR engines available then.

Figure 5-5: Tesseract

Tesseract supports the ability to recognize more than 100 languages out of the box.

Role in the project: engine helps working of PyTesseract

5.3.5 PDFMiner

PDFMiner is a tool used to extract information from PDF documents. Unlike other PDF-related tools, it focuses totally on getting and analyzing text data.

Figure 5-6: PDFMiner

PDFMiner allows obtaining the exact location of text in a page, as well as other information such as lines or fonts. It includes a PDF converter that can transform PDF files into many text formats (such as HTML). It has an extensible PDF parser that can be used for other many purposes than text analysis.

Role in the project: it converts PDF files into text.

5.3.6 PyTTSX

PYTTSX (Python Text-To-Speech) is a cross-platform text to speech library which is independent platform that converts text to speech in python. The major advantage of this library for text-to-speech conversion is that it works offline.

Figure 5-7: Python Text to Speech

Role in the project: convert text into speech.

5.3.7 PyAudio

PyAudio supports Python bindings for PortAudio, the cross-platform audio I/O library. With PyAudio, you can easily use Python to record and play audio on a variety of platforms.

Role in the project: recording and playing audio.

Figure 5-8: PyAudio

5.3.8 SpeechRecognition

SpeechRecognition is a library that perform speech recognition role in python. Its support lies in several engines and APIs, online and offline like Google Cloud Speech API, Microsoft Bing Voice Recognition, IBM Speech to Text etc.

Role in the project: converting speech into text.

Figure 5-9: SpeechRecognition

5.4 HARDWARE USED

5.4.1 Raspberry Pi

Raspberry Pi is a small single-board computer.

The Raspberry Pi 3 Model B is the earliest model of the third-generation Raspberry Pi.

Specification

- Quad Core 1.2GHz Broadcom BCM2837 64bit CPU
- 1GB RAM
- BCM43438 wireless LAN and Bluetooth Low Energy (BLE) on board
- 100 Base Ethernet
- 40-pin extended GPIO
- 4 USB 2 ports
- 4 Pole stereo output and composite video port
- Full size HDMI
- CSI camera port for connecting a Raspberry Pi camera
- DSI display port for connecting a Raspberry Pi touchscreen display

Figure 5-10: Raspberry Pi

- Micro SD port for loading your operating system and storing data
- Upgraded switched Micro USB power source up to 2.5A

5.4.2 Pi Camera

The Raspberry Pi camera (PI Camera) module can be used to take high-definition video, as well as photographs.

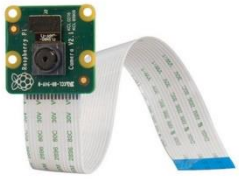


Figure 5-11: Pi Camera

▪ Specification

- 5MP sensor
- Wider image, capable of 2592x1944 stills, 1080p30 video
- 1080p video supported
- CSI
- Size: 25 x 20 x 9 mm

5.4.3 Microphone

Input device used to deliver user's command to the device.

Figure 5-12: Microphone

5.4.4 Power Bank

Power bank role is to be the power supply of the Raspberry Pi.

Figure 5-13: Power Bank

5.4.5 Speaker

Output device used to deliver the device response to the user.

Figure 5-14: Speaker

Chapter 6

Conclusion and Future Work

Chapter 6: Conclusion and Future Work

6.1 Introduction

In this chapter we share and summarize our steps, results, knowledge and experience while developing **VIA**.

6.2 Phases of the Project

Phase 1 – Getting ready for the project by searching for the best idea to help many people as possible.

Phase 2 – After finding the idea, we started searching to study the basic knowledge we need to learn

Phase 3 – After Studying, we started dividing the whole team to 3 sub-teams to work on the 3 main features in parallel

Phase 4- Each team starts to search for the libraries and frameworks that will achieve the target of each feature.

Phase 5 – For long time of working each team build the feature their separately.

Phase 6 – All the features were integrated to be working together with the core feature, the voice commands.

6.2 Conclusion

Visually Impaired Assistant is built up for people who have problems in vision. The system's main purpose is to help visually impaired people to live their live normally without needing help from anybody. Our project main features are to interact with the users through voice helping them to:

1. Recognize the objects around them.
2. Read anything they want to.
3. Choose their favorite books from database and read it for them through voice.
4. Save notes for them, so that they can listen to it again.

In our project we have built and trained our model using different algorithms to reach the min acceptable accuracy for offering reliable user interaction. We have tried logistic regression, SVM and boosted decision trees and finally reached a good enough accuracy for classification of 90% using logistic regression as a classifier.

6.3 Future Work

Our goal is to visually impaired Peron can live without any help from the other people.

So we suggest:

1. Provide Arabic language support for both voice recognition and book reading.

2. Include a book store where users can buy a book if it is not available for free.

3. Improve the performance and reduce response time of the device.

4. Add smart home commands by which user can control lights, air condition and other home appliances; so the home will be more controllable and cozy.

5. Add a GPS feature to the device to guide the user outside home.

6. Provide a warning system to warn the user whether there are stairs in front of him or in case of the user is about to hit anything.

References

https://www.wikipedia.org/

https://github.com/

https://stackoverflow.com/

www.tutorialspoint.com/