PALESTINE POLYTECHNIC UNIVERSITY
College of Information Technology and Computer Engineering
Computer Engineering Department

Intelligent Personal Assistant

Team Members

Bahaa Abu Sneineh
181061

Jibreel Natsheh
181039

Supervisor:

Eng. Wael Takrouri

2022 - 2023
Hebron - Palestine

Certification and Anti-Plagiarism Declaration

This is to declare that the graduation project produced under the supervision of Eng. Wael Takrouri having the title "Intelligent Personal Assistant" was prepared by students Jibreel Natsheh and Bahaa Abu Sneineh below in partial fulfillment of the requirements for the degree of Bachelor in Engineering of Computer Systems Engineering and no part hereof has been reproduced illegally (in particular: cut and paste) which can be considered as Plagiarism.
All referenced parts have been used to support and argue the idea and have been cited properly. We certify that we will not commit any plagiarism, cheating, or any other academic integrity violation. We will be responsible and liable for any consequence if violation of this declaration is proven.

Date: 22/12/2022
Graduation project group's student(s):

Name: Jibreel Natsheh
Signature:

Name: Bahaa Abu Sneineh
Signature:

Abstract

This project aims to create intelligent personal assistance using Raspberry Pi, speech recognition, and face recognition technologies. The system consists of two parts, an interactive part, which is the microphone, speaker, robot's head, camera, and displays. These parts provide the robot with the ability to hear, speak, see, move and express his feelings in a similar way to the human's interactions, and that's why it's intelligent.

The second part is the processing unit, which collects the data from the input devices (microphone and camera), processes the data, and sends the output results and actions to the output parts (speaker, displays, and servo motors).

In this project, we designed and implemented both the interactive and the intelligent parts of the system. The system is able to get speech commands from the user. The system also analyzes the speech and performs any predefined command, and is able to respond to the user with speeches that explain the status of each speech command.

Keywords: Intelligent Personal Assistant, Interactive Robot, Speech Recognition, Face Identification.

الملخص

يهدف هذا المشروع إلى إنشاء مساعدة شخصية ذكية باستخدام Raspberry Pi وتقنيات التعرف على الكلام والتعرف على الوجوه. يتكون النظام من جزأين ، جزء تفاعلي ، وهو الميكروفون ومكبر الصوت ورأس الروبوت والكاميرا وشاشات العرض. حيث توفر هذه الأجزاء للروبوت القدرة على الاستماع والتحدث والتحرك والتعبير عن مشاعره بطريقة مماثلة لتفاعلات الإنسان ، وهذا هو السبب في أنها ذكية.

الجزء الثاني هو وحدة المعالجة، التي تجمع البيانات من أجهزة الإدخال (الميكروفون والكاميرا) ، وتعالج البيانات ، وترسل نتائج الإخراج والإجراءات إلى أجزاء الإخراج (مكبر الصوت ، وشاشات العرض ، والمحركات المؤازرة).

في هذا المشروع ، قمنا بتصميم وتنفيذ الأجزاء التفاعلية والذكية للنظام. النظام قادر على الحصول على أوامر الكلام من المستخدم. يقوم النظام أيضًا بتحليل الكلام وتنفيذ أي أمر محدد مسبقًا ، وهو قادر على الرد على المستخدم بخطابات تشرح حالة كل أمر كلام.

الكلمات الرئيسية: مساعد شخصي ذكي ، روبوت تفاعلي ، التعرف على الكلام ، التعرف على الوجه.

Acknowledgment

First and foremost, we would like to express our warmest and deepest gratitude to our supervisor, Eng. Wael Takrouri. This project would not have been possible without his guidance. We are truly indebted to their expertise and countless hours of help.

We would also like to extend our sincere gratitude to everyone who contributed to the success of this project including our families, friends, and the exceptional team of the College of Information Technology and Computer Engineering for their endless support and kind spirits.

Table of Contents

List of Figures

List of Tables

List of Acronyms

AI	Artificial Intelligence
ANN	Artificial Neural Network
API	Application Programming Interface
CSI	Camera Serial Interface
DC	Direct Current
DP	Display Port
DSI	Display Serial Interface
eDP	embedded Display Port
EDR	Enhanced Data Rate
GPIO	General Purpose Input/Output
GPU	Graphics Processing Unit
GUI	Graphics User Interface
HDMI	High-Definition Multimedia Interface
I2C	Inter-Integrated Circuit
I2S	Inter-IC Sound
IoT	Internet of Things
LCD	Liquid-Crystal Display
LPDDR	Low-Power Double Data Rate
MIPI	Mobile Industry Processor Interface
ML	Machine Learning
OLED	Organic Light-Emitting Diodes
PC	Personal Computer
PWM	Pulse Width Modulation
RAM	Random Access Memory
SBC	Single Board Computer
SCL	Serial Clock
SDA	Serial Data
SPI	Serial Peripheral Interface
STT	Speech To Text
SoC	System on Chip
UART	Universal Asynchronous Receiver/Transmitter
UHD	Ultra High Definition
USB	Universal Serial Bus

Chapter 1

Introduction

1.1 Overview

In today's modern world, the working environment in offices has become increasingly important as more people are working remotely or from home. Major tech companies are always introducing new products and services to improve the office experience for employees. It is important to enhance the working environment and experience to increase efficiency and productivity.

However, the needs of disabled people in the office must also be considered. Many people with disabilities face barriers in the workplace, such as a lack of accessibility or accommodations for their needs. By considering the needs of disabled individuals, companies can create a more inclusive and welcoming work environment for all employees.

Motivated by the need to improve the office experience for both able-bodied and disabled individuals, we have decided to use our knowledge and experience gained during the past five years of college to design and implement an Intelligent Personal Assistant that can handle multiple tasks. Our goal is to create a tool that is not only helpful for all employees, but also accessible and accommodating for those with disabilities.

In this chapter, we will discuss the idea behind the project and the motivations behind it, as well as its importance and objectives. We hope that our Intelligent Personal Assistant will not only improve the work experience for all employees, but also help break down barriers and create a more inclusive and welcoming work environment for everyone.

1.2 Motivation and Importance

The increasing reliance on technology has made it difficult for disabled, blind, and elderly individuals to keep up with the rapid pace of innovation. These groups of people often struggle to use all technologies, which can lead to feelings of frustration, isolation, and even despair.

As engineers, it is our responsibility to offer solutions that can help increase the efficiency of people working in the office and improve the level of luxury that they have. By developing an intelligent personal assistant that helps all individuals and especially people with special needs, we can provide a much-needed tool that can help these groups of people live more independent and fulfilling lives.

The importance of this project comes as it has the potential to significantly improve the quality of life for special needs people, while also helping to bridge the technological divide that currently exists.

1.3 Problem Statement

The problem is that special needs people face challenges in using modern technologies, which can limit their ability to perform certain tasks and negatively impact their quality of life and work experience.

As engineers, we have an opportunity to develop solutions that address this problem and help increase the efficiency of people working in offices, as well as improve their level of luxury.

This project aims to create a system that addresses these issues by making it easier for people with special needs to use and interact with different technologies through the use of speech commands and automation scripts.

1.4 Project Description

The project consists of two main parts: an interactive part and an execution part. The interactive part includes a microphone and speaker, webcam, motion servomotors, and an output display. The user interacts with the robot using voice commands, which the robot analyzes and sends to the execution part for processing.

The execution part is a pre-defined and pre-trained machine learning and artificial intelligence system that can recognize speech, recognize faces and gestures, and execute orders.

The robot uses speech recognition, face recognition, and web robots to carry out the commands.

1.5 Project Objectives

- Train the robot on the 'wake up' word.
- Implement the robot's interactive motion using motors.
- The robot uses the output display to express its current mood.
- Train the robot to analyze the speech and recognize specific speech commands and turn it into a command.
- Train the robot to be able to recognize face(s).
- Implement the functions of each command:
  - Play music and do google search.
  - Face Detection and Log-In to the students portal the university.
  - API (Application Programming Interface) calls and returning the values in a human understandable speech.

1.6 Project Requirements

- Functional Requirements
  - The user interacts with the personal assistant using voice commands.
  - The robot detects a wake-up word to start working.
  - It responds to the user speech commands with voice replies and actions (if the commands are defined).
  - It uses the motors to move its arms.
  - It expresses its feeling on an output display.

- Non-Functional Requirements
  - **Scalability**: The system is scalable that is new features and commands can be added.
  - **Responsiveness**: It responds to user commands fast.
  - **Portability**: It is portable and can be placed from one place to another as it's a stand-alone robot and only needs power and internet to do all the tasks.

1.7 Expected Results

After designing and implementing the system it will be able to:

- Recognize the user's voice commands and execute it accordingly.
- Interact with the user using facial expressions, motion, voice, and executions.
- It will be able to recognize the face of the user and use the login credentials that are mapped to the user.
- It will be able to move its arms.
- It will be able to express its feeling on the monitor using emotional emojis.

1.8 Overview of the rest of the report

This report is organized as follows: Chapter 2 provides a literature review and overview of the theoretical background of the project, including the hardware and software components. Chapter 3 discusses the conceptual design of the system, including block diagrams, pseudo-code, detailed hardware connections, and other relevant design information. Chapter 4 explains the implementation of the hardware and software modules of the system and their integration. Chapter 5 presents the test and validation results for the system modules. Finally, Chapter 6 provides a conclusion for the project and discusses potential future work to enhance the system.

Chapter 2

Background

This chapter presents a theoretical background of the system in Section 2.1. the literature review is shown in Section 2.2. In Section 2.3 shows hardware components and Section 2.4 presents a description of the software components and tools used in the project.

2.1 Theoretical Background

This section introduces a list of key technologies which are used in this project such as Face Recognition, Speech Recognition, Gestures Recognition, Text to Speech conversion, and Browser's Robot implementation.

2.1.1 Face Recognition

Face recognition is an important research problem spanning numerous fields and disciplines. This is because face recognition, in addition to having numerous practical applications such as bankcard identification, access control, mugshots searching (search for similar photos/ objects), security monitoring, and surveillance system, is a fundamental human behavior that is essential for effective communications and interactions among people.

An overview of the major human face recognition techniques that apply mostly to frontal faces, and the advantages, and disadvantages of each method are also given. The methods considered are eigenfaces (eigenfeatures), neural networks, dynamic link architecture, hidden Markov model, geometrical feature matching, and template matching. The Approaches are analyzed in terms of the facial representations they used. [9]

2.1.2 Speech Recognition

Speech is the most crucial, widespread, and proficient form of communication method for people to commune with each other. Humans are comfortable with speech hence people would also like to interact with computers via speech, rather than via primitive interfaces such as keyboards and pointing devices. Speech Recognition is the interdisciplinary subfield of computational linguistics that builds up techniques and technologies that facilitate the recognition and translation of spoken words into the textual format by computers. It is also known as "speech to text" (STT). It includes knowledge and research in the linguistics, computer science, and electrical engineering fields. The objective of speech recognition is for a computer to be capable of "perceive speech", "recognizing" and "taking action upon" spoken words.

Speech recognition has many challenges, but most importantly the different types of speech, such as:

- Isolated Words.
- Connected Words.
- Continuous speech.

The process of converting a voice/ speech into a text has many stages, but it can be concluded as illustrated in Figure 2.1 and the following is a description of each stage.

Figure 2.1: Process of converting a voice/ speech into a text

1. Analysis: It is a pre-processing stage where as much as possible of the noise is being removed, where the valuable information is being kept and the unwanted noise is being removed.

2. Feature Extraction: In this stage, the speech is transformed into a stream of vector coefficients that contains the required information to identify the given speech parts.

3. Classification: It is the most important stage, and in this stage, machine learning and artificial intelligence algorithms are being used in order to understand speech and there are 3 approaches in this stage, and it is what we are going to compare in the next sections. [10]

2.2 Literature Review

In this section, we will list some projects similar to the idea of our project.

2.2.1 AI Assistant Robot with Arduino and Python

Authors: Ashraf Minhaj
Release Date: November 15, 2020
It is a Virtual Assistant AI Robot for Computer Automation and Little Chit Chat. This project is a simple example of a personal assistant using Arduino and a PC. where the Arduino is used for the robot/interactive part of the work and the PC is the brain that analyzes and executes the commands and sends the Arduino the required interactions to be executed and he has used the microphone and speakers from the PC is used. [11]

The main difference between our work and this project is that we are using Raspberry Pi instead of Arduino as it is a faster controller than it. Our robot will be a standalone system and doesn't need to be connected to PC. so basically it performs as both, a controller and a brain. Our robot has its own microphone, speakers, and camera.

2.2.2 Speech Recognition and Synthesis with Arduino

Authors: Marcio Tamagushi
Release Date: January 17, 2016
Implements speech recognition and synthesis using an Arduino DUE. This project has done an improvement that it used the Arduino to do more functions, where both the microphone and the speaker were connected with the Arduino, but still it was using the PC to do the voice recognition part of the project. [12]

This project differs from our project in that we are using Raspberry Pi instead of Arduino as we can use built-in services for it, so we don't need to connect it to the PC. also we have other services and technology used that can't be done on regular Arduino like face recognition and so on.

2.2.3 ALLUKA “Smart Virtual Assistant Robot”

Authors: Sawalha, Eman Abu Alrob, Ghadeer
Release Date: 2018
Alluka is an educational robot that helps children and students in the learning process in a very interactive way. The main idea that the robot supports is taking voice commands from the children and producing a movement or speaking out the answer to the question or commands. [13]

The difference is that this project is a more complex assistant. On one side, our robot is not moving (not a placement movement). But on the other hand, our robot performs complex functions such as using face ID to identify the user and log into the website using the required credentials.
Table 2.1 shows a summary for the literature review.

Project Name	**Project Description**	**Improvement**
AI Assistant Robot with Arduino and Python	Build an AI robot using an Arduino connected to a PC where it uses the microphone and speaker from the PC	We used Raspberry Pi as one device instead of two devices, and we have added face recognition technology
Speech Recognition and Synthesis with Arduino	Implements speech recognition and synthesis using an Arduino	We have used Raspberry Pi as it’s a faster option than Arduino and can do more functions
ALLUKA “Smart Virtual Assistant Robot”	Robot that supports taking voice commands and producing a movement or speaking out the answer for the question or commands.	We have added more complex and different technologies and commands, like face recognition techniques and web robots.

Table 2.1: Literature Review Summary

2.3 Hardware Components

This part describes the components and devices used in the system.

2.3.1 Single Board Computer (SBC)

We need a small processor or minicomputer that fits in the robot’s body to plug other components into it and to operate the image processing part in our project,

we need this component to process and control the system.
The are many options of this type of processing units and Table 2.2 compares three types of SBC.

Specification	ASUS Tinker Board	LattePanda Alpha	Raspberry Pi 3 model B+ 4GB
Processor	Rockchip rk3288 1.8 GHz Quad-core ARM cortex-a17-32bit	Intel Core m3-7Y30 Processor 1.0 GHz-2.6 GHz	Broadcom bcm2711 Quad-core Cortex A72 (ARM V8) 64 bit SoC @ 1.5 GHz
RAM	2 GB LPDDR3	8 GB Dual Channel	4 GB
GPU	Mali T764@600 MHz	Intel HD Graphics 615 300 MHz-900 MHz	Broadcom Videocore Vi @ 500 MHz
Display	HDMI with H.264 and H2.265 4K decode capability	HDMI/DP/eDP	2x micro -HDMI ports supporting 4K @60Hz displays via HDMI 2.0 MIPI DSI display port
Storage	MicroSD	32 GB	MicroSD + USB boot mode
Ports	Drie4 x USB 2.0 40 x GPIO 1 x HDMI 1 x Camera Serial Interface (CSI) 1 x Display Serial Interface (DSI) 3.5mm audio jack	USB Type-C DC Power Input Battery Port Audio Port USB 3.0x3 MicroSD I2C-I2S-RS232-USB2.0 B2.0	2 x USB 3.0 40 x GPIO 1 x HDMI 2-lane MIPI Camera Serial Interface (CSI) 2-lane MIPI Display Serial Interface (DSI) 3.5mm audio-video jack

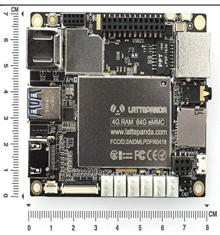
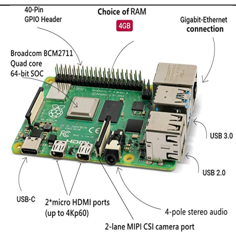
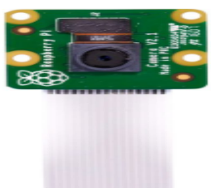
Lan Port	10/100 Gigabit	Gigabit	Gigabit
Wi-Fi and Bluetooth	Wi-Fi 802.11 b/g/n with swappable antenna Bluetooth 4.0 with EDR	802.11ac 1x1 2.4GHz, 5GHz Dual Band Bluetooth 4.2	Wi-Fi 802.11 b/g/n 2.4/5GHz
Cost	$ 150 (19-7-2022) [Amazon]	$ 170 (19-7-2022) [Amazon]	$ 265 (19-7-2022) [Amazon]
Image			

Table 2.2: Differences between ASUS Tinker Board, Raspberry Pi and LattePanda Alpha [14] [15] [16]

We choose Raspberry Pi as it is more available in the market, suitable for a small robot controller that plugs into a monitor, and uses a standard keyboard and mouse, it's capable of doing everything you would expect a PC to do, from browsing the internet and playing videos, to making spreadsheets, word processing, and playing games. Multiple (general-purpose input/output) GPIO pins on the board allow you to connect the board with other peripherals and are capable of generating PWM, I2C, SPI, and UART serial communications.

2.3.2 Camera

In our project, we need a camera to take pictures of the person that uses the robot and send them to the trained machine learning module to detect the face.
There are different options for camera and we came to two options, and Table 2.3 shows the comparison between Raspberry Pi camera and USB web camera.

Specification	Raspberry PI camera	Web Camera
Image		
Input Port	CSI port	USB port

Resolution	supports 1080p @ 30fps (Frame per Second), 720p @ 60fps and 640x480p 60/90	1280 x 720, 1080p which measures 1920 x 1080 pixels, and Ultra-High Definition (UHD, 4K) which has a resolution of 3,840 by 2160
Cost	$ (2-25) (19-7-2022) [Aliexpress]	$ (8-80)(19-7-2022) [Aliexpress]

Table 2.3: Raspberry Pi camera vs Web Camera [17] [18]

We choose the Web Camera, as it's more reliable. After we testing the Raspberry Pi camera, we found that it was very sensitive to light and voltage changes in the power source, and it was prone to damage. Therefore, we determined that the Web Camera was a better option.

2.3.3 Speaker

The speaker is used for the robot to communicate and respond to user commands and Table 2.4 illustrates the differences between the Micro Speaker and the JBL m3 mini speaker.

Specification	Microspeaker	JBL M3 mini speaker
Image		
Cost	$ 2 (19-7-2022) [Aliexpress]	$ 12 (19-7-2022) [Aliexpress]
Size	2*5*3 mm	10*5*10 mm

Table 2.4: Micro Speaker vs JBL M3 mini speaker [19] [20]

We have selected the microspeaker, as it's cheaper in cost, and also it performs well with what we want, so it meets all our requirements.

2.3.4 Microphone

Microphone will be used to input voice commands from the user to the robot and so to analyze the speech and do the required task. There are two options for the microphones and Table 2.5 shows the difference between the two types.

Specification	Wired Microphone	Bluetooth Microphone
Image		
Cost	$ 13 (8-1-2022) According to (Amazon)	$ 6-14 (8-1-2022) According to (AliExpress)
Sensitivity	-28 dB	85dB
Power Consumption	5mW	4mW
Response Frequency	12-60 kHz	20-150 kHz
Connector Type	USB	Bluetooth

Table 2.5: Wired Microphone vs Bluetooth Microphone [1] [2]

We will use wired microphone, as it's more available in the market and it is faster and more accurate than the Bluetooth microphone, so it meets all our requirements.

2.3.5 Servo Motor

Servo motor will be used for the robot to have 1 degree of freedom in order to move the head in a circular motion, in order to be more interactive and Table 2.6 shows the comparison between MG90 and SG90 servo motors.

Specification	MG90	SG90
Dimension	40.7 x 19.7 x 42.9 mm approx.	23 x 11.5 x 24 mm approx.
Stall torque	9.4 kgf · cm (4.8 V), 11 kgf · cm (6 V)	4.8 v : 1.2kg-cm
Operating speed	0.17 s/60º (4.8 V), 0.14 s/60º (6 V)	0.10sec/60°
Operating voltage	4.8 V - 7.2 V	3.0 -7.2 v
Running Current	500 mA (900 mA (6V))	4.8 v : 1.2kg-cm
Temperature range	0 ºC -55 ºC	-30 ºC - 60 ºC
Cost	$ 20 (19-7-2022) [Amazon]	$ 20 (19-7-2022) [Amazon]
Image		

Table 2.6: MG90 vs SG90 [3] [4]

We will use SG90 servo, as it's more available in the market and it performs well with what we want, so it meets all our requirements.

2.3.6 Small Display Screen for Facial Expressions

Display screen unit will be used to express the robot's current mood using facial emotion displayed on the display. We searched for options for the display screen and Table 2.7 lists and compares the options.

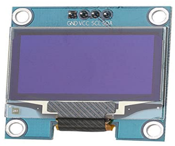
Specification	OLED Display	LCD 2X16
Image		
Cost	$ 9 (19-7-2022) [Amazon]	$ 12 (19-7-2022) [Amazon]
Working Voltage	3.3V	5V
Dimension	3.82 x 2.28 x 0.75 inches	4.9 x 1.7 x 0.4 inches

Table 2.7: OLED Display vs LCD 2x16 [5] [6]

We have selected the OLED display, as it's more available in the market and it meets all our requirements.

2.3.7 Display Screen

Display screen will be used to show the browser when we open the student's portal and Table 2.8 shows the comparison between the LED display with the TFT display.

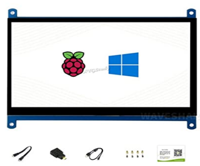
Specification	7inch LED Display	TFT Display
Image		
Cost	$ 100 (23-11-2022) [Amazon]	$ 30 (23-11-2022) [Amazon]
Connection Port	HDMI Port	GPIO Ports
size	7 inches	3.5 inches

Table 2.8: LED display with TFT display [7] [8]

We have selected the LED display, as it uses HDMI port while the TFT display's connection doesn't comply with the other components where it uses all the pins so we will not be able to connect the servo motors and the OLED display.

2.4 Softwares Components

This section illustrates the software and the data used in the project.

2.4.1 Python

Python is a very popular high-level interpreter-based programming language. It was first released in 1991 but continued to evolve to this day. Python programs are very simple and easy to read. Few lines in python often correspond to tens of lines in other programming languages like C. This means that writing programs in Python is time-saving practice. In Addition, programs written in Python are shorter in length and more readable. Furthermore, Python scripts can run on many platforms including Windows, Linux, and Unix. Due to its popularity, Python has a very large community and a lot of resources. Python users can easily add code packages to their programs simply by importing them at the top of the script. Programmers can also download non-standard packages with PIP, which is a package manager for Python.

2.4.2 Face Recognition Algorithm

Following are the types of algorithms that are used in Face Recognition and Table 2.8 compare these algorithms:

Technique	Description	Advantages	Disadvantages
Eigenface	Any face images could be approximately reconstructed by a small collection of weights for each face and a standard face picture (eigenpicture). The weights describing each face are obtained by projecting the face image onto the eigenpicture.	• Fast. • Simple. • Practical method.	It does not provide invariance over changes in scale and lighting conditions.
Neural Networks	One of the first artificial neural networks (ANN) techniques used for face recognition is a single-layer adaptive network called WISARD which contains a separate network for each stored individual.	The feature extraction step may be more efficient than the linear way in Eigenface.	(1) Slow at training.(2) Slow at recognizing when we have too many individuals.(3) Not suitable for single model image.
Graph Matching	Presented a dynamic link structure for distortion invariant object recognition which employed elastic graph matching to find the closest stored graph.	Computation- ally expensive.	Superior to other face recognition techniques in terms of facial expression, luminati change and rotations.

Table 2.9: Comparison between different Face Recognition algorithms [9]

After comparing the above algorithms we have chosen Neural Networks, as it's more accurate, despite that it's slow, it's still efficient for us as we have a small number of users. But after testing the algorithm on the Raspberry Pi, we found that Neural Networks takes longer time to process, so we have switched to Eigenface algorithm as it's much faster, and to solve the scaling issues in Eignface approach, we have preceded some image processing on the pictures before using it in the algorithm, where was detecting the face and crops the picture on the face only.

2.4.3 Speech Recognition

Following are the types of algorithms that are used in Voice Recognition:

Approach	Description
Acoustic Phonetic Approach	In the Acoustic Phonetic approach speech recognition was based on discovering speech sounds and endowing them with appropriate labels. This is the foundation of the acoustic-phonetic approach which hypothesizes that there exist fixed, distinguishing phonetic units called phonemes and these units are largely regarded as a set of acoustics properties presenting a speech. The primary stage in the acoustic-phonetic approach is a spectral analysis of the speech combined with a feature detection that alters the spectral measurements to a set of features that portray the broad acoustic properties of the different phonetic units. The subsequent step is a segmentation and labeling phase in which the speech signal is segmented into stable acoustic regions, followed by attaching one or more phonetic labels to each segmented region, resulting in a phoneme lattice characterization of the speech. The last step in this approach attempts to determine a valid word (or string of words) from the phonetic label sequences produced by the segmentation to labeling.
Pattern Recognition Approach	Pattern classification mainly embodies the training or development of a system (given a feature vector) that will divide a large number of individual examples into groups called classes. As the source of the speech is often due to a large number of causes, the available speech signal results from the combination of the audio channel, noise, additive noise, etc. Pattern Classification (or recognition)is the process of comparing the unknown test pattern with each sound class reference pattern and computing a measure of similarity between them. After completing the training of the system at the time of testing patterns are classified to recognize the speech.
Artificial Intelligence Approach	The approach of artificial intelligence is the most enlarged and effective technique, which supports error-free and accurate speech recognition and is used for decoding. It is because; artificial intelligence includes certain algorithmic techniques, which foster logical conversion and transformation of speech into understandable patterns, and vice versa. The artificial intelligence approach is a mixture of the pattern recognition approach and acoustic-phonetic approach so it is called a hybrid approach of pattern recognition and acoustic-phonetic approach. It is due to the fact that it incorporates the concepts and ideas of pattern recognition methods and acoustic-phonetic approach.

	It has been established that artificial intelligence is also mentioned as a knowledge based approach and it uses the information, which is related to spectrogram, phonetic, and linguistic. Artificial intelligence approaches play an essential role indifferent activities of speech recognition, including designing of recognition algorithms, demonstration of speech units, and representation of proper and appropriate inputs. It is considerable to bring into the notice that, some of all methods of speech recognition, artificial intelligence is the most reliable and proficient methods.

Table 2.10: Voice Recognition algorithms

We choose the Pattern Recognition Approach, as it is more accurate and faster than other approaches, and meets our need for a fast and real-time voice recognition module.

Chapter 3

System Design

This chapter gives a description of the system, a detailed design, and important information about the design.

3.1 Detailed Description of The System

Our system will be designed to help the user achieve tasks using vocal commands. The functions that the robot will do is play music, login to the University's website using Face ID, answer the 'How is the weather now?' question, and give some interactive response to the user and this will be implemented by 3 subsystems:

1. **Interactive Subsystem** in which the user will directly interact with it and it will contain the following part:
    - **Microphone**: where the robot will use it to get the commands from the user.
    - **Speaker**: which the robot will use to give interactions, warnings, and answers to the user.
    - **Monitor**: which will show the currently open web page/application.
    - **OLED screen**: to show the robot's facial expressions.
    - **Servo Motor Robot Arms**: to do some interactive moves.
    - **Raspberry Pi**: where it connects all systems together.
2. **Speech Recognition Subsystem**: this system is the intermediate module that is responsible for:
    - Convert the spoken speech into text using Python tools and libraries.

- Extract the intended commands from the text words using keywords matching formulas, and machine learning could be used here if the regular formulas were not working properly.
- Trigger the required function to be executed and get the returned output out of this function, and the output will be one of the followings:
  - Text to be converted and sent to the speaker.
  - A facial expression is to be presented on the OLED monitor.
  - A rotation/s to be executed by the servo motor to move the head.

3. **Executing Subsystem** that executes the commands and the following are the commands:

   - Play music on Youtube.
   - Search for a any topic on Google.
   - Login in the student portal website using Face ID using the following techniques:
     - Identify the user using a picture to be taken using the camera and identifying it using a pre-trained face recognition module.
     - Extract the mapped login credentials.
     - Use a 'Python Web Robot' to open the Student's Portal and log in using the credentials provided.

3.2 Block Diagram

The following subsection presents the block diagram of the hardware components and the relationship between them as shown in Figure 3.1.

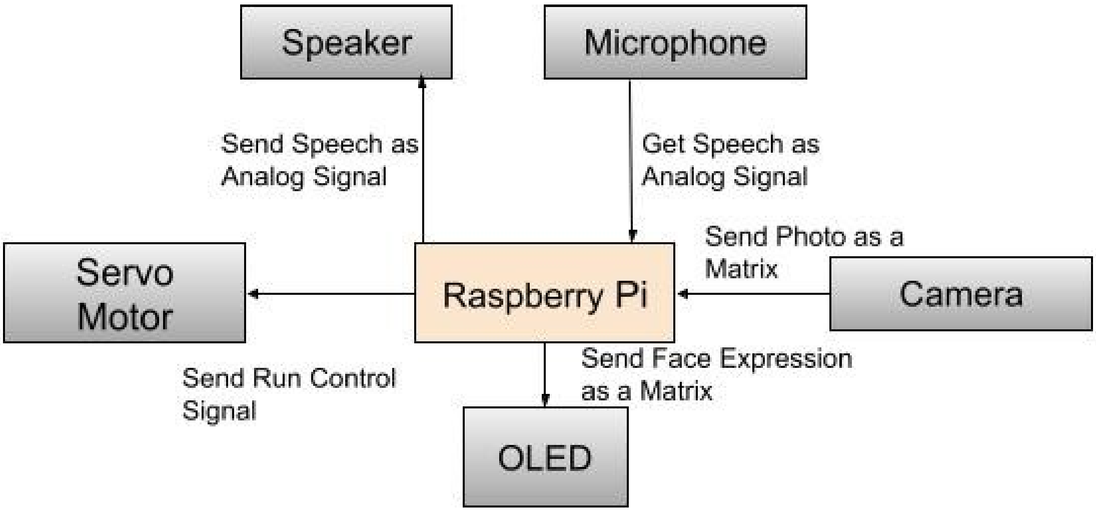


Figure 3.1: System Block Diagram

3.3 System Schematics

The system schematics and the connections of the hardware components is shown in Figure 3.2.

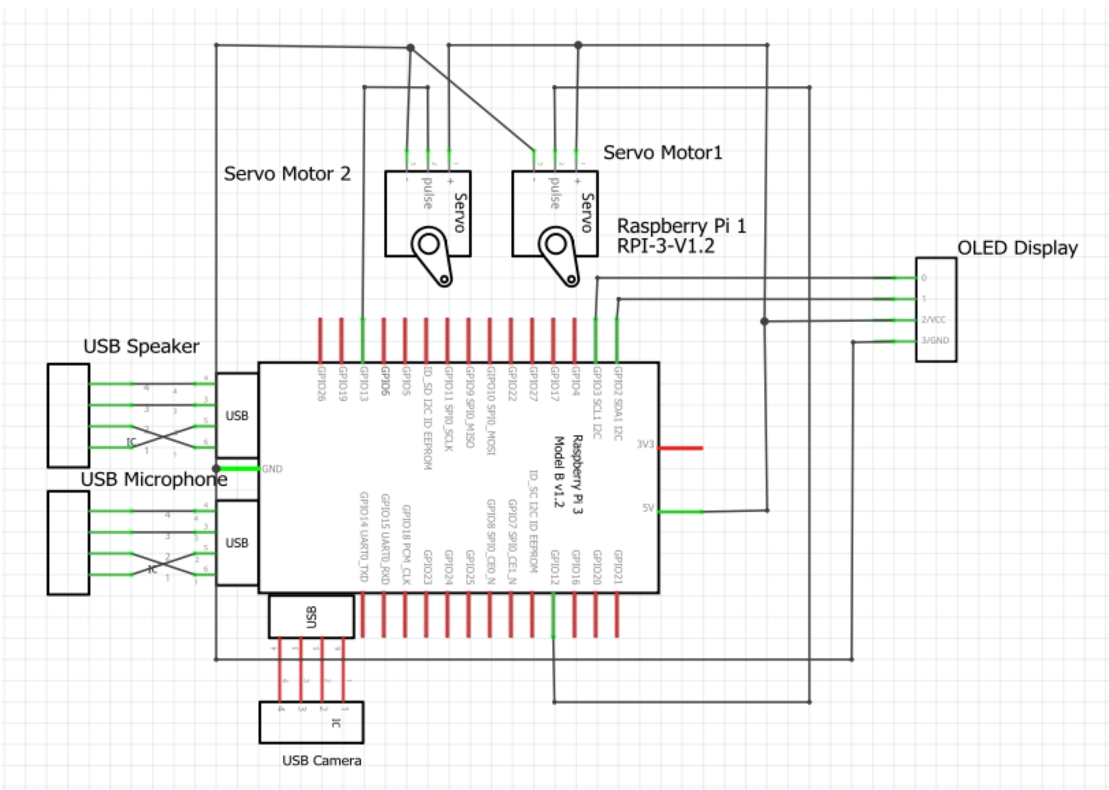


Figure 3.2: System Schematic Diagram

Following is a detailed description for the system schematics and connections:

1. The ground pin and the 5V pin of the microcontroller will be connected to Raspberry Pi.
2. The lines Vcc, GND, and PWM lines are connected to the 5V, GND and GPIO13 of the Raspberry Pi.
3. The lines Vcc, GND, and OLED lines are connected to the 5V, GND, GPIO2, and GPIO3 of the Raspberry Pi.
4. Camera is connected to the camera port in the Raspberry Pi.
5. Speaker and microphone are connected via the Raspberry Pi USB port.

3.4 Body Design

The robot's body was designed to contain the Raspberry Pi and the other hardware components and connections as shown in figure 3.3.

Figure 3.3: Robot's Body Design

We used SolidWorks 2020 Software to design the body to be printed using Anet A8 3D printer.

We designed the body to have the shape of a robot, with arms and a head that can rotate on their axes. The arms have a place to connect to the servo motor, and the body has a place for the touch screen display. The face has a place for the camera and OLED display.

3.5 Pseudo code of the system

The pseudocode of the system is illustrated in Algorithm 1, when the system is powered up:

Algorithm 1

```
procedure Begin
    DisplayOLED with SmilingFace
    RobotHeadRotate
    while True do
        if There is an input command from user then
            ANALYZE VoiceCommand to Get Command
            if Command is Wake-up word then
                RobotTalk with "Hey, How can I Help you?"
                Set status to active
            else if Command is PlayMusic and status is active then
                PlayMusicClip with MusicClipLink
            else if Command is LoginPortal and status is active then
                CaptureUser with Cam
                IdentifyUser with UserCapture to get User
                if User not exists then
                    DisplayOLED with SadFace
                    RobotTalk with "sorry, you don't have account"
                else
                    CALL GetUserCredentials with User
                    OpenStudentsPortal
                    LoginRobot with Creds
                    RobotTalk with "Welcome <user-name>"
            else if Command is WeatherQuestion and status is active then
                CALL WeatherAPI to get degree
                if degree not between 15,25 then
                    DisplayOLED with SadFace
                RobotTalk with "it is x degrees today"
            else if status is active then
                Robot Says:"Sorry, I don't know what do you mean, please try again
            DisplayOLED with SmilingFace
```

Chapter 4

Implementation

This chapter provides a detailed description of the implementation of our project, including the various hardware components and software parts of the system.

4.1 Hardware Implementation

In this section, we describe the process of implementing and assembling all of the hardware components of our system. Our system consists of a Raspberry Pi 3 Model B+ to control all of the components and operations.
The process of implementing the project began by connecting all of the components to the Raspberry Pi and writing code to control each component. The next step was to assemble the components into the shape of a robot.
We designed the robot's body using Solid Works 2022 and printed it using a 3D printer, as the design shows in Chapter 3 in Figure 3.3 and Figure 4.1 illustrates the printed parts of the robot body.

Figure 4.1: The printed robot's body parts

The microphone, speaker, camera, and display screen are connected to the Raspberry Pi via USB ports, with the display screen also connected to the HDMI port. The OLED display is connected to the Raspberry Pi using the I2C pins SDA, SCL, and power, and the servo motor is connected to the power and GPIO 12, as shown in Chapter 3 in Figure 3.2 and Figure 4.2 illustrates the robot's body after assembling.

Figure 4.2: The robot after assembling the parts

4.2 Software Implementation

This section provides an overview of the implementation details of the various software components of the system. It explains the choices made for different parameters, features, and functions of each software component in the system.

4.2.1 Speech Recognition Module

The speech recognition module consists of the microphone as an input device, the Raspberry Pi as the digitizing device and controller, and the speech recognition software. The **speech_recognition** Python library was used to capture the input from the microphone, utilizing the PyAudio driver, and the Google API, a predefined voice recognition engine based on the Pattern Recognition Approach described in sections 2.1.2 and 2.4.3.

The script below captures speech from the microphone and converts it into text:

```python
# run the app
while True:
    try:
        with sr.Microphone() as source:  # get input from mic
            print("Talk >>")
            listener.energy_threshold = 100
            listener.adjust_for_ambient_noise(source, 2)
            voice = listener.listen(source)
            command = listener.recognize_google(voice).lower()  # use google API
            # all words lowercase- so that we can process easily
            print(command)
            # look for wake up word in the beginning
            if robot_name in command or isAwake == True:
                if isAwake is False:
                    print("[wake-up word found]")
                    isAwake = True
                isAwake = process(command)
    except Exception as e:
        print("Exception: " + str(e))
```

We set the *energy_threshold to 100* and used the 'adjust_for_ambient_noise' function to reduce noise and control the response time. The source and 2 parameters are used in this function to specify the audio source and duration, respectively.

4.2.2 Face Recognition Module

The face recognition module consists of the camera as an input device, the Raspberry Pi as the digitizing device and controller, and the face recognition software. The Eigenface algorithm was used, as described in section 2.4.2. To improve the accuracy of the module, we applied image processing to the photos before using them in the Eigenface module. We used the **cv2** and **NumPy** libraries to detect and crop the face, and the **face_recognition** library, which is based on the Eigenface algorithm, to extract and encode the facial features, mapping them to the user name, student number, and portal password.

The script below shows how the image is processed and matched to the correct user:

```
# Grab a single frame of video
frame = video_capture.read()
# Resize frame of video to 1/4 size for faster face recognition processing
small_frame = cv2.resize(frame, (0, 0), fx=0.25, fy=0.25)
# Convert the image from BGR color (which OpenCV uses) to RGB color (which face_recognition uses)
rgb_small_frame = small_frame[:, :, ::-1]
# Find all the faces and face encodings in the current frame of video
face_locations = face_recognition.face_locations(rgb_small_frame)
face_encodings = face_recognition.face_encodings(rgb_small_frame, face_locations)
face_names = []
for face_encoding in face_encodings:
       # See if the face is a match for the known face(s)
       matches = face_recognition.compare_faces(known_face_encodings, face_encoding)
       name = "Unknown"
       face_distances = face_recognition.face_distance(known_face_encodings, face_encoding)
       best_match_index = np.argmin(face_distances)
       if matches[best_match_index]:
            match = known_face_names[best_match_index]
            return match
        return "Unkown"
```

4.2.3 Google and Youtube Search Module

In this module, we used the **pywhatkit** Python library, which is an easy-to-use library that offers various helpful methods such as **playonyt**, which takes a text as a parameter and searches for it on YouTube, playing the first search result, and

search, which takes a text as a parameter and searches for it on Google using the default browser to display the results.
The script below shows how these library functions are used to process a search term:

```
pywhatkit.playonyt(extension)
pywhatkit.search(extension)
```

4.2.4 Web Robot Module

We used the **selenium** library to control a web browser, specifically the Chromium browser, to open the student portal and log in using the credentials obtained from the face recognition module. The methods used to control the screen are JavaScript-based commands, and the elements are identified using XPATH selectors.
The script below shows how **selenium** is used to control the web page session:

```
browser = webdriver.Chrome(executable_path='./chromium-browser')

browser.get('https://www.ppu.edu/p/ar/portals')
browser.find_element_by_xpath("//a[@href='http://reg.ppu.edu']").click()
browser.find_element_by_id('txtUserName').send_keys(studentNumber)
browser.find_element_by_id('txtPassword').send_keys(password)
browser.find_element_by_id('txtPassword').send_keys(Keys.RETURN)
```

4.2.5 Robot Speech Module

We used the **pyttsx3** Python library, which is an online text-to-speech library, and the **pyaudio** library to connect to the speaker and play the text as speech.
The script below shows the function where we pass the text as a parameter and it is sent to the speaker module:

```
def talk(sentence):
    """ talk / respond to the user """
    engine.say(sentence)
    engine.runAndWait()
```

4.2.6 Robot Facial Expressions Module

We used the **Adafruit_GPIO.SPI**, **Adafruit_SSD1306**, and **PIL** Python libraries to control the OLED display, and the **c2v** library to convert the selected image for the current expression into a 2-D matrix to be displayed on the OLED. The script below demonstrates how the **displayEmotion** function takes the emotion path as a parameter and displays it on the OLED:

```python
def displayEmotion(emotion):
    file_path = oled_emotion_paths.get(emotion)
    image = Image.open(file_path).resize((disp.width, disp.height),
    ↪ Image.ANTIALIAS).convert('1')
    disp.image(image)
    disp.display()
```

4.2.7 Robot Arms Rotation Module

We used the **RPi.GPIO** Python library to control the rotations of the servo motors and move the robot arms by 90 degrees.
The script below demonstrates how we rotate the servos to achieve this movement:

```python
def runServo():
    servo.start(0)
    print ("Rotating at intervals of 12 degrees")
    duty = 2
    while duty <= 17:
        servo.ChangeDutyCycle(duty)
        time.sleep(0.08)
        duty = duty + 1
    print ("Turning back to 0 degrees")
    servo.ChangeDutyCycle(2)
    time.sleep(1)
    servo.ChangeDutyCycle(0)
    servo.stop()
    GPIO.cleanup()
    print ("Everything's cleaned up")
```

By combining all of the modules, we will have a robot that can understand and execute speech commands.

Chapter 5

Testing and Validation

This chapter will cover the testing and validation process for each module individually and for the full system as an integrated system.

5.1 Face Recognition Module

We tested the face recognition module on various devices and cameras and found it to be fast and simple with no issues. However, the camera on the Raspberry Pi did not have good quality when the system was running, leading to decreased accuracy in the recognition module with an error rate of up to 20%. Despite this, the module is still valid and meets our requirements and Table 5.1 displays the results after testing the face recognition module.

Situation	Camera Quality	Number of Trials	Success Rate
Single Module	High quality	20	100%
Full System	Bad quality	30	80%

Table 5.1: Test results for the face recognition module

5.2 Speech Module

The speech module has been verified and tested with a variety of statements, numbers, and conversations of different lengths, and the response time was not noticeable.

We also tested the speaker and found that the sound quality was sufficient for our needs in a room or office, but may not be sufficient for larger areas. However, as it meets our requirements, we will consider it a valid component.

5.3 Arms Motion Module

We tested the servo motors to work in parallel and we can control the rotation angel by setting the PWM value, and after testing we found that they functioned correctly, although there is mechanical blockages that prevents the arms from rotating correctly due to it's weight, but overall so we can consider it a valid component.

5.4 Speech Recognition Module

The speech recognition module was tested in real-time using 30 sentences of varying lengths in both low and high noise environments.

- In a low noise room, the system correctly recognized 26/30 sentences without mistakes and 4/30 sentences had 1-2 incorrect words.
- In a high noise room, the system correctly recognized 18/30 sentences without mistakes, 6/30 sentences had 1-2 incorrect words, and 6/30 sentences had more than 2 incorrect words, leading to misunderstandings.

Table 5.2 displays the test results for the speech recognition module.

Situation	Number of Trials	Success Rate
Low noise room	30	86%
High noise room	30	60%

Table 5.2: Test results for the speech recognition module

Chapter 6

Conclusion and Future Work

This chapter summarizes the results of the project and discusses potential future work.

6.1 Conclusion

This project aimed to improve the ability of officers and disabled individuals to use and interact with various technologies by enabling them to complete tasks using speech commands and creating automation scripts for daily tasks that had faster performance compared to traditional methods.

We successfully built an intelligent personal assistant that can understand speech commands using minimal hardware and software configuration. This was achieved with the use of a single microcomputer and a simple collection of low-cost hardware and software components.

During this project, we learned that it can be challenging to create a standalone system using basic and incompatible hardware, as this results in reduced quality for each component when running them together. For more compatible systems, it is more efficient to use a laptop or PC.

To minimize costs, we used simple hardware such as a Raspberry Pi 3 Model B+ and a microphone, speaker, camera, and display in this project. However, we had to replace several of these devices multiple times to achieve a more compatible system that met our project requirements. Additionally, we were restricted by the specific microcomputer we initially used, which severely limited the progress of the project. To improve accuracy and reduce error rates, we could upgrade to a higher microcomputer, power source, and camera to improve the robot's face recognition capabilities and overall system efficiency.

Overall, most of the components in the system were very accurate and performed their duties well. However, the face recognition module had an error rate due to poor camera performance, and the speech recognition module sometimes struggled in noisy environments with the microphone we used.

6.2 Future Work

There are several features that can be added to this project to improve and enhance the system. The following are some examples:

- Adding additional mobility to the robot, such as moving its head.
- Creating a GUI to allow users to define new commands and register additional users without modifying the source code.
- Implementing gesture recognition to control the volume.
- Adding an Arabic language version to the system.
- Adding a feature that allows the robot to read screen content loudly to assist blind individuals
- Adding feature of controlling house appliance like lights and AC (Air Condition).

References

[1] "Mini usb microphone for laptop and desktop computer, with gooseneck universal usb sound card, compatible with pc and mac, plug play, ideal condenser mic for remote work, online class, cgs-m1," Jul 2022. [Online]. Available: https://www.amazon.com/Microphone-Gooseneck-Universal-Compatible-CGS-M1/dp/B08M37224H/ref=sxin_15_pa_sp_search_thematic_sspa?content-id=amzn1.sym.9e69e792-0ff0-4e1a-b10c-a41b9d9b3ffc%3Aamzn1.sym.9e69e792-0ff0-4e1a-b10c-a41b9d9b3ffc&cv_ct_cx=raspberry+pi+microphone&keywords=raspberry+pi+microphone&pd_rd_i=B08M37224H&pd_rd_r=74841803-4818-41ff-88c2-ad499543600d&pd_rd_w=ubxIs&pd_rd_wg=nGrYQ&pf_rd_p=9e69e792-0ff0-4e1a-b10c-a41b9d9b3ffc&pf_rd_r=9GKH52YQFG0BXZNBB1KZ&qid=1659357201&sr=1-4-a73d1c8c-2fd2-4f19-aa41-2df022bcb241-spons&psc=1&spLa=ZW5jcnlwdGVkUXVhbGlmaWVyPUEzRDdKWkJUN1UyVUpLJmVuY3J5cHRlZElkPUEwN

[2] "Wireless bluetooth microphone handheld karaoke mic audio for children musical stage toy music singing speaker kids gift," Jul 2022. [Online]. Available: https://ar.aliexpress.com/item/1005002247781548.html?gatewayAdapt=glo2ara

[3] "10pcs mg90s metal gear digital 9g servo for rc helicopter plane boat car mg90 90s 9g," Jul 2022. [Online]. Available: https://www.amazon.com/10pcs-MG90S-Metal-Digital-Helicopter/dp/B07TT4W659

[4] "Universal 0.12 seconds / 60 degrees sg90 servo, motor, for planes 90°180° airplane accessories mini robotssg90," Jul 2022. [Online]. Available: https://www.amazon.com/Universal-Seconds-90%C2%B0-180%C2%B0-Airplane-Accessories/dp/B08NJGVV3H

[5] "Hiletgo 1.3" iic i2c serial 128x64 ssh1106 ssd1306 oled lcd display lcd module for arduino avr pic stm32," Jul 2022. [Online]. Available: https://www.amazon.com/

HiLetgo-Serial-SSH1106-Display-Arduino/dp/B01MRR4LVE/ref=sr_1_2_sspa?keywords=oled%2Bdisplay&qid=1659637159&sr=8-2-spons&spLa=ZW5jcnlwdGVkUXVhbGlmaWVyPUExVElHS0xYOTFLRVNPJmVuY3J5cHRlZElkPUEwO th=1

[6] "Sunfounder iic i2c twi 1602 serial lcd module display compatible with arduino r3 mega 2560 16x2," Jul 2022. [Online]. Available: https://www.amazon.com/SunFounder-Serial-Module-Display-Arduino/dp/B019K5X53O/ref=sr_1_1_sspa?keywords=16x2%2Blcd%2Bdisplay&qid=1659637049&sr=8-1-spons&spLa=ZW5jcnlwdGVkUXVhbGlmaWVyPUExUEZLS01ZOEhRMVBEJmVuY3J5cHRlZElkPUEwN th=1

[7] "Elecrow raspberry pi screen touchscreen 7 inch mini hdmi monitor lcd screen 1024x600 compatible with raspberry pi 400 4 3b+ 3b windows pc," Nov 2022. [Online]. Available: https://www.amazon.com/ELECROW-Display-1024X600-Function-Raspberry/dp/B01GDMDFZA

[8] "or raspberry pi 4 touch screen with case, 3.5 inch touchscreen with fan, 320x480 monitor tft lcd game display," Nov 2022. [Online]. Available: https://www.amazon.com/Raspberry-Touchscreen-320x480-Monitor-Display/dp/B07WQW6H9S

[9] "Face recognition: A literature review," Jul 2022. [Online]. Available: https://www.researchgate.net/publication/233864740_Face_Recognition_A_Literature_Review

[10] "Survey paper on different speech recognition algorithm: Challenges and techniques," Jul 2022. [Online]. Available: https://www.researchgate.net/publication/320547133_Survey_paper_on_Different_Speech_Recognition_Algorithm_Challenges_and_Techniques

[11] "Ai assistant robot with arduino and python," Jul 2022. [Online]. Available: https://create.arduino.cc/projecthub/ashraf_minhaj/ai-assistant-robot-with-arduino-and-python-ac4c52

[12] "Speech recognition and synthesis with arduino," Jul 2022. [Online]. Available: https://create.arduino.cc/projecthub/msb4180/speech-recognition-and-synthesis-with-arduino-2f0363

[13] "Alluka "smart virtual assistant robot"," Jul 2022. [Online]. Available: https://repository.najah.edu/handle/20.500.11888/14113

[14] "Tinker board." Jul 2022. [Online]. Available: https://www.amazon.com/Rockchip-Computer-Bluetooth-Connectivity-Dual-Display/dp/B08V4B9TG9/ref=sr_1_2?keywords=Tinker+Board&qid=1659355578&sr=8-2

[15] "Lattepanda," Jul 2022. [Online]. Available: https://www.amazon.com/LattePanda-4G-64GB-Development-without/dp/B01ID4HYE4/ref=sr_1_1_sspa?keywords=lattepanda+alpha&qid=1659355765&sr=8-1-spons&psc=1&spLa=ZW5jcnlwdGVkUXVhbGlmaWVyPUEyU0lLSFpNRVhSWEtPJmVuY3J5cHRlZElkPUEwOT
=

[16] "Raspberry," Jul 2022. [Online]. Available: https://www.amazon.com/Vemico-Raspberry-Starter-Heatsinks-Screwdriver/dp/B09QGZ94M8/ref=sr_1_1_sspa?crid=ZVXGOO3SEVCL&keywords=Raspberry+Pi+4+model+B+4GB&qid=1659355895&sprefix=raspberry+pi+4+model+b+4gb%2Caps%2C240&sr=8-1-spons&psc=1&spLa=ZW5jcnlwdGVkUXVhbGlmaWVyPUEyV0lJRlVTNzZBMlVKJmVuY3J5cHRlZElkPUEwMT
=

[17] "Pi camera module v2.1 - raspberry pi camera v2.1," Jul 2022. [Online]. Available: https://www.distrelec.de/en/raspberry-pi-camera-v2-raspberry-pi-pi-camera-module-v2/p/30134462

[18] "Webcam 1080p full hd web camera for laptop pc computer video meeting usb webcam built-in microphone," Jul 2022. [Online]. Available: https://www.alibaba.com/product-detail/Webcam-1080p-Full-HD-Web-Camera_1600172261047.html

[19] "2pcs/lot lcd monitor/tv speaker horn 3w 4r 4028 2840 loud speaker 4 ohms 3 watt 4r 3w 40*28mm," Jul 2022. [Online]. Available: https://ar.aliexpress.com/item/1005003316110659.html?gatewayAdapt=glo2ara

[20] "Jbl m3 mini portable bluetooth speaker," Jul 2022. [Online]. Available: https://www.shoponly.pk/shop/accessories-and-gadgets/wireless-speaker/jbl-m3-mini-portable-bluetooth-speaker/?v=d71bdd22c8bb