TRIBHUVAN UNIVERSITY
INSTITUTE OF ENGINEERING
PASHCHIMANCHAL CAMPUS
LAMACHAUR, POKHARA

A Project Report
On
"**HOME AUTOMATION USING SPEECH RECOGNITION**"

Submitted by:
Binayak Shrestha [PAS077BEI012]
Prakanda Bhandari [PAS077BEI030]
Pratham Adhikari [PAS077BEI032]
Sandesh Bashyal [PAS077BEI036]

Submitted to:
Department of Electronics and Computer Engineering
Pashchimanchal Campus, Lamachaur-16, Pokhara

ABSTRACT

Home automation using speech recognition is a project carried out with intention to explore the concept of home automation through integration of technology. It is a hardware and software-based project which focuses on economic implementation of home automation. The goal of this project is to explore various technological aspects involved in achieving automation of tasks to improve the overall convenience. To achieve this goal, real-time audio data from the microphone is processed using ADC in Arduino Nano and transmitted using the bluetooth module which is received by the built-in bluetooth module in raspberry pi. The data is then processed by the speech recognition model in raspberry pi which generates text as an output. This text obtained is compared with predefined keywords and incase of a match, the text is converted to signals which is then transferred to appliance(s) through the relay module. The status of appliances are then updated in real-time and shown through a web application.

Keywords : arduino nano, raspberry pi, torchaudio, speech-to-text, wake-word detection algorithm

TABLE OF CONTENTS

ABBREVIATION

ML	*Machine Learning*
AI	*Artificial Intelligence*
Wi-Fi	*Wireless Fidelity*
RNN	*Recurrent Neural Network*
LSTM	*Long-Short Term Memory*
ASR	*Automatic Speech Recognition*
IDE	*Integrated Development Environment*
I2C	*Inter-Integrated Circuit*
SPI	*Serial Peripheral Interface*
ADC	*Analog to Digital Converter*
UART	*Universal Asynchronous Receiver/ Transmitter*
MEAN	*MongoDB, Express, Angular, and Node*
MERN	*MongoDB, Express, React, and Node*
MEVN	*MongoDB, Express.js, Vue.js, and Node.js*
REST	*Representational State Transfer*
API	*Application Programming Interface*
GHz	*GigaHertz*
USB	*Universal Serial Bus*
HDMI	*High-Definition Multimedia Interface*
GPIO	*General Purpose Input/Output*
I/O	*Input/Output*
2D	*2 Dimension*
CPU	*Central Processing Unit*
CUDA	*Compute Unified Device Architecture*
NC	*Normally Closed terminal*
NO	*Normally Open terminal*
COM	*Common terminal*
DC	*Direct Current*

LIST OF FIGURES AND TABLE

CHAPTER 1: INTRODUCTION

1.1 Background

In the contemporary era of smart living, home automation systems have become increasingly popular due to their ability to enhance convenience, efficiency and also catering to individuals with mobility challenges. Integration of speech recognition technology into home automation systems adds an additional layer of user-friendliness, accessibility, leveraging its power to seamlessly interact with various devices within a home environment.

Traditional method with physical switching of appliances, checking if they are on or not brings a level of inconvenience to disabled as well as abled people. This system can accurately interpret the user instructions and generates control signals for appliances. It does so by capturing audio input through a connected microphone, processed, digitized and transmitted via a Bluetooth module linked to the Arduino Nano, establishing a wireless communication link with the Bluetooth module integrated into the Raspberry pi. Upon receiving the audio signals, Raspberry pi utilizes a speech recognition model within it which converts the received voice data into text commands enabling the system to comprehend user instructions accurately. Once the command is interpreted, It then generates necessary control signals and sends them to a 4-channel relay module, allowing a seamless control of different household appliances connected to the system. The status of the appliances are updated in real-time which can be viewed on a dashboard of a web application. This system not only operates the appliances, but also views their status which makes it easier to observe them when we are not physically present near them.

1.2 PROBLEM STATEMENT

The absence of automated controls requires manual operation of various appliances and thus results in a time consuming routine. Managing multiple devices manually can be troublesome especially for elderly and the disabled. The commercial home automation systems are based mainly on Zigbee, Z-Wave, or Wi-Fi protocols. Such automated systems are either ineffective in terms of cost or power consumption.

Additionally, the complexity of processing speech in real-time poses a significant challenge. Existing systems often struggle with accurately interpreting the intent of the speaker. Thus, the technologies that are being applied are prone to various kinds of failures, for instance an individual's speech can be misinterpreted,in case of multiple commands by more than one person the commands might be neglected.

1.3 OBJECTIVES

- To design and develop a speech recognizing home automation system.
- To develop a system that can comprehend user instructions accurately and operates accordingly.

CHAPTER 2: LITERATURE REVIEW

Home automation has evolved significantly as it enhances convenience and efficiency in household tasks. There are various technologies to achieve home automation such as sensor based automation, Hub based systems, machine learning, cloud based automation, etc.There have been several research projects intended to automate various tasks. In recent times there has been an increase in the use of artificial intelligent robots, neural networks, Natural language understanding and generation,voice recognition and so on in various fields [1]. Thus, there has been an increase in the use of technology to manage household devices. There have been commercial and research projects on smart homes and speech recognition systems through use of protocols like Zigbee [2], Z-Wave, or Wi-Fi protocols. To build a reliable, compact, fast and low cost home automation system bluetooth communication protocol can be used [3]. So, home automation has been successful through use of speech recognition [1][2] and still evolving till the date [8].

To do any ML task, the first and foremost need is access to the dataset and that's the case for the speech recognition too. Speech recognition has always been a state-of-the-art task in the field of ML and AI. Speech recognition is the recognition of what has been said by the speaker to understand the intent of the speaker [1]. We have datasets and toolkits like Kaldi implemented in PyTorch framework [4], Deep Speech [5], and many more. However, its implementation on the edge devices like Raspberry Pi is challenging but has been made possible as in [6]. Speech recognition is done by wake-word detection algorithm and the speech-to-text model. We have sequence-to-sequence networks like RNN, LSTM, and self-attention based models like transformers to perform ASR [7].

CHAPTER 3: RELATED THEORY

3.1 Microphone

A microphone is a device that translates sound vibrations into electrical signals. It enables users to capture audio and transmit it to various devices such as recording medium or over a loudspeaker. Microphones are used in many applications such as telephones, sound engineering, live and recorded audio engineering, radio and television broadcasting, etc., for communication purposes.

3.2 Arduino Nano

The Arduino Nano is a small, breadboard-friendly microcontroller board based on the ATmega328. It is equipped with 30 male I/O headers which can be programmed using the Arduino Software IDE (Integrated Development Environment). It can communicate with a computer and other microcontrollers.It consists of flash memory of 32KB capacity of which 2 KB is used for Bootloader.It supports Inter-Integrated Circuit (I2C), Serial Peripheral Interface (SPI) and Universal Asynchronous Receiver / Transmitter (UART).It has 10 bit ADC.

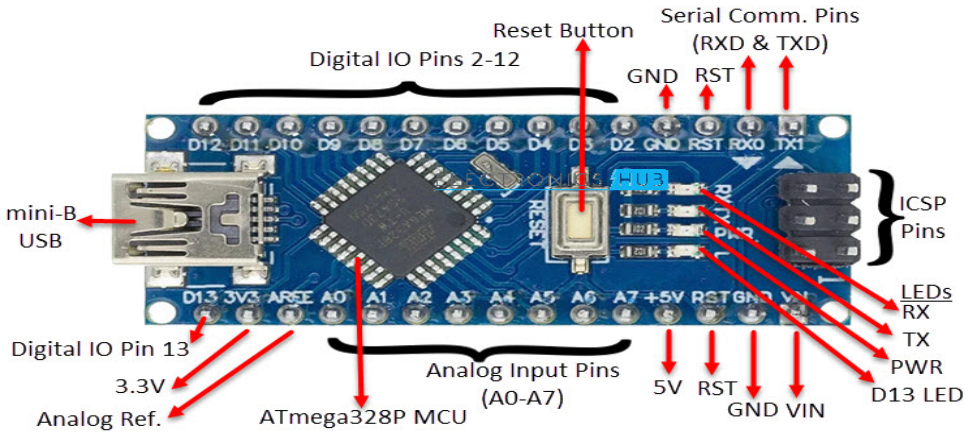

Figure 1: Arduino Nano

3.3 Bluetooth module

The Bluetooth module acts as the interface between devices and the Bluetooth network.It uses short-range radio waves to establish connection between devices. Bluetooth modules operate on frequencies within the 2.4 GHz range.Bluetooth is a wireless communication protocol that allows devices to exchange data over the radio waves.

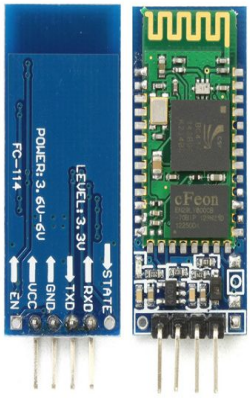

Figure 2: Bluetooth module

3.4 Raspberry Pi 4 Model B

The Raspberry Pi is a credit card-sized single-board computer that provides a powerful processing platform in a compact form factor. It is equipped with a range of input/output interfaces, including USB ports, HDMI, Ethernet, GPIO (General Purpose Input/output), and a camera interface. The processor speed of Raspberry Pi 4 Model B is 1.5 GHz. It comes with onboard wireless networking and bluetooth.

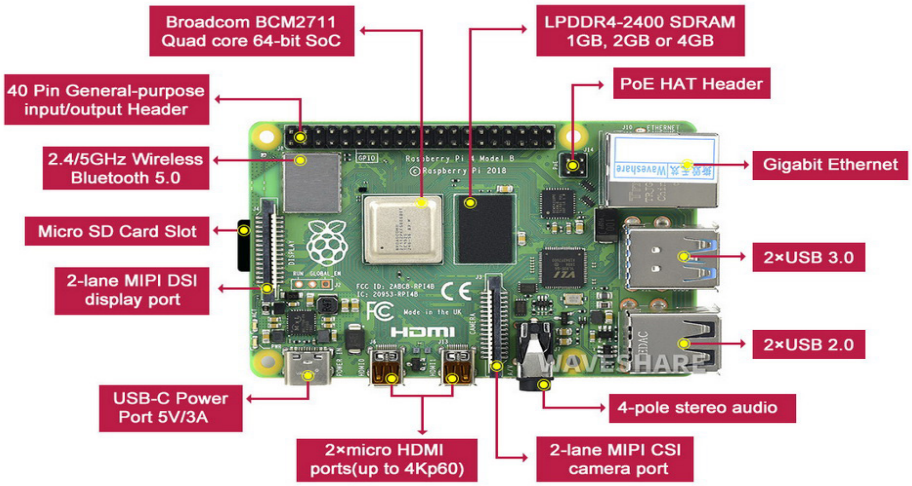

Figure 3: Raspberry Pi 4 Model B

3.5 Relay Module

Relays are electrically operated switches that open and close circuits.They permit a small amount of electrical current to control high current loads.They protect the control systems from high voltage and currents as they are used to provide electrical isolation between the control and load circuits.

Figure 4: 1-Channel Relay Module

3.6 Motor

Motors also known as actuators are fundamental components in robotics, providing the necessary actuation, precision, and control for robot movement and manipulation. With the help of motors, the system can be moved to different locations using a precise control system in the 2D axis.

Figure 5: Servo motor

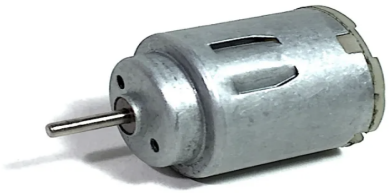

Figure 6: DC motor

3.7 Arduino IDE

The Arduino IDE is a software application designed specifically for programming Arduino boards. It provides users with a convenient and user-friendly interface for writing, editing, and uploading code to Arduino microcontrollers.

3.8 TorchAudio

TorchAudio is the audio and signal processing library with PyTorch, an open source ML framework of the Torch library. TorchAudio provides tools and utilities to work on audio and signals processing, and integrate them into PyTorch based ML workflows. It is useful in tasks like audio classification, speech/ voice recognition, audio transformation, and many more. It is widely used because it supports both CPU and CUDA devices. Its APIs can correctly backpropagate gradients and execute in non-Python environments with the assistance of the property Autograd and TorchScript, respectively [11].

3.9 React.js

React is a free and open-source front-end JavaScript library for building user interfaces based on components. It can be used to develop single-page, mobile, or server-rendered applications with frameworks like Next.js. It allows the creation of interactive and visually appealing user interfaces. Developers can design and implement components that display real-time status of the appliances [9].

3.10 Express.js

Express is a back-end web application framework for building RESTful APIs with Node.js, released as open-source software. Express is the back-end component of popular development stacks like the MEAN, MERN or MEVN stack together with the MongoDb database software and a JavaScript front-end framework [10].

CHAPTER 4: METHODOLOGY

4.1 System Block Diagram

In layman's terms, the home automation project with speech recognition is: you speak into a machine, and it takes that command to control whether household appliances turn on or off (in fig 7).

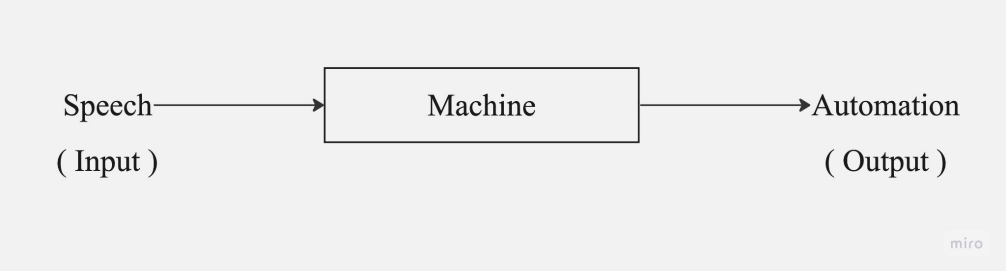

Figure 7: Block diagram of home automation

Delving deeper into the workings of the system, the process begins with the user providing speech as input through a portable microphone. This input is then transmitted to a Raspberry Pi via a combination of Arduino Nano and Bluetooth module. Within the Raspberry Pi, a speech-to-text model interprets the spoken words. Upon successful interpretation, the Raspberry Pi generates a signal that commands the control of the appliance(s) through the relay module. This comprehensive system involves multiple components working in tandem to seamlessly translate spoken commands into tangible actions for home automation. The block diagram of the system hardware is shown in figure 8.

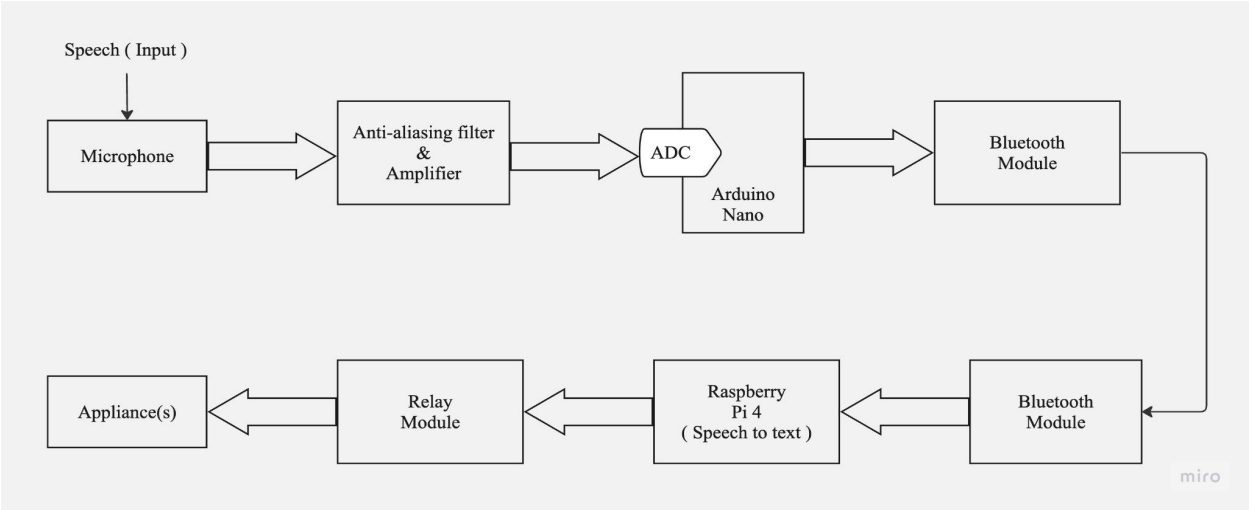

Figure 8: Block diagram of a system hardware

The process unfolds with the speech serving as input for the speech-to-text model. The converted text undergoes comparison with a wake word in the wake word detector model. If the text does not match a wake word, the system remains inactive until a wake word is detected, prompting the process to restart. Upon recognizing the wake word, the converted text that follows is subjected to a set of conditions, and signals are then dispatched to control the appliances. This iterative sequence ensures that appliance control is contingent on the recognition of a specific wake word, enhancing the system's precision and responsiveness. The flowchart is shown in figure 9.

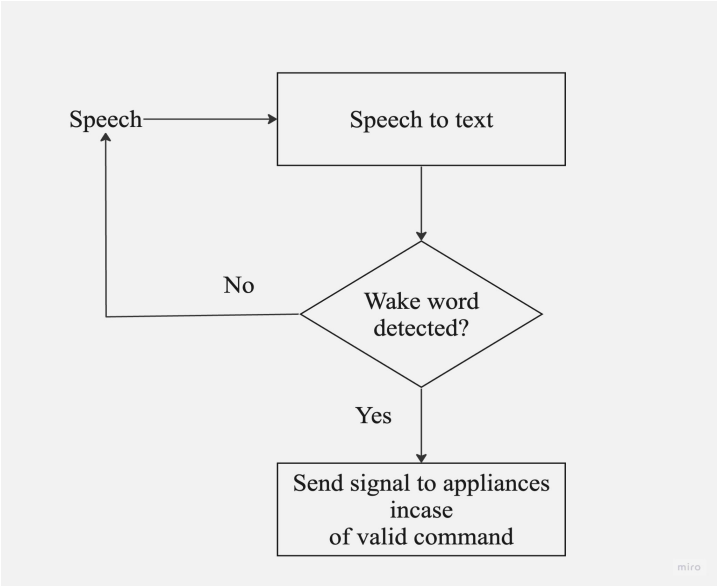

Figure 9: Flowchart of Speech Recognition

CHAPTER 5: CONCLUSION AND ANALYSIS

5.1 Expected Output

The expected output at the end of the development of this system is given below:

- There will be a home autonomous system where different home appliances can be operated with the help of microphone and voice command eliminating the need for manual control.
- This system can recognize speech and operates on electrical appliances and its status can be seen on a web application.
- The incorporation of voice commands, the system caters to individuals with disability or mobility challenges, making them less dependent on people for simple tasks of operating switches or appliances.

5.2 Cost Estimation

After Research, the cost of the project is estimated as below:

S.N.	Budget Heading	Cost (Nrs.)
1.	Condenser Microphone	99
2.	Raspberry Pi 4 Model B	20,000
3.	Arduino Nano 3.0	800
4.	Bluetooth Module HC-06	820
5.	4 Channel Relay Module	480
6.	Servo Motor	350
7.	DC Motor	100
	Total	22,649

Table 1 : Cost Estimation

5.3 Conclusion

In a nutshell, this proposal outlines a comprehensive and innovative approach to home automation using speech recognition. Traditional systems relied on manual input methods and complex interfaces, so the need for seamless interaction for disabled individuals and those seeking a more user-friendly experience was evident. The integration of speech recognition offers a hands-free and intuitive solution that transcends the limitations of conventional control interfaces. It is a future-proof investment.

REFERENCES

[1] Ayorinde Oduroye, et al. “Voice Recognition Door Access Control System.” *ResearchGate*, unknown, 4 Oct. 2019, www.researchgate.net/publication/337012927_Voice_Recognition_Door_Access_Control_System.

[2] Humaid AlShu'eili, et al. “Voice Recognition Based Wireless Home Automation System.” *ResearchGate*, unknown, May 2011, www.researchgate.net/publication/252013629_Voice_recognition_based_wireless_home_automation_system.

[3] Das, Sukhen, et al. *A Bluetooth Based Sophisticated Home Automation System Using Smartphone*. 1 Oct. 2016, ieeexplore.ieee.org/abstract/document/7859709, https://doi.org/10.1109/icicpi.2016.7859709.

[4] Mirco Ravanelli, et al. “The Pytorch-Kaldi Speech Recognition Toolkit.” *HAL (Le Centre Pour La Communication Scientifique Directe)*, 1 May 2019, ieeexplore.ieee.org/abstract/document/8683713, https://doi.org/10.1109/icassp.2019.8683713.

[5] Hannun, Awni, et al. “Deep Speech: Scaling up End-To-End Speech Recognition.” *ArXiv.org*, 2014, arxiv.org/abs/1412.5567.

[6] Gondi, Santosh, and Vineel Pratap. “Performance Evaluation of Offline Speech Recognition on Edge Devices.” *Electronics*, vol. 10, no. 21, 4 Nov. 2021, p. 2697, https://doi.org/10.3390/electronics10212697.

[7] Benzeghiba, M., et al. “Automatic Speech Recognition and Speech Variability: A Review.” *Speech Communication*, vol. 49, no. 10-11, Oct. 2007, pp. 763–786, www.sciencedirect.com/science/article/abs/pii/S0167639307000404, https://doi.org/10.1016/j.specom.2007.02.006.

[8] Li, Jinyu. “Recent Advances in End-To-End Automatic Speech Recognition.” *APSIPA Transactions on Signal and Information Processing*, vol. 11, no. 1, 2022, https://doi.org/10.1561/116.00000050.

[9] “Describing the UI – React.” *React.dev*, 2023, react.dev/learn/describing-the-ui.

[10] “Express - Node.js Web Application Framework.” *Expressjs.com*, 2017, expressjs.com/.

[11] “Torchaudio Documentation — Torchaudio 2.1.1 Documentation.” *Pytorch.org*, 2021, pytorch.org/audio/stable/index.html.