International Journal of Engineering Science Invention Research & Development; Vol. 12, Issue 10, April 2026
www.ijesird.com, E-ISSN: 2349-6185

DESIGN OF A SPEECH-DRIVEN ROBOTIC ASSISTANT FOR REAL-TIME HUMAN–ROBOT COMMUNICATION

L. Ganesh[1], D. Sai Shanmukh Ram[2], K. Radha Krishna[3], M. Ganesh[4], B. Lokesh[5], Dr. B. Siva Prasad[6]
[1,2,3,4,5] *B.Tech Students, Department of Electronics and Communication Engineering*
[6] *Professor, Department of Electronics and Communication Engineering,*
Nadimpalli Satyanarayana Raju Institute of Technology

***Abstract*—This paper presents the design and implementation of a Speech-Driven Robotic Assistant aimed at enabling real-time human–robot communication using voice commands. The proposed system integrates speech recognition, embedded processing, and robotic control to facilitate intuitive interaction between humans and machines. A microphone module captures voice input, which is processed using speech recognition algorithms implemented on a Raspberry Pi platform. The recognized commands are interpreted and transmitted to a microcontroller (Arduino), which controls the robot's movement and actions accordingly.**

The system supports real-time execution of commands such as navigation, object monitoring, and environment interaction, thereby reducing the need for manual control interfaces. Speech recognition has become a key enabling technology for human–robot interaction (HRI), allowing natural communication between humans and machines. Additionally, voice-controlled robotic systems have demonstrated effectiveness in assistive and automation applications.

The proposed solution is cost-effective, scalable, and suitable for applications in assistive robotics, smart surveillance, healthcare support, and industrial automation.

***Index Terms*—Speech Recognition, Human–Robot Interaction (HRI), Raspberry Pi, Arduino, Voice Control, Embedded Systems, Real-Time Processing, Assistive Robotics, Smart Surveillance, Industrial Automation, IoT**

I. Introduction

Human–Robot Interaction (HRI) has emerged as a significant research domain due to rapid advancements in robotics, artificial intelligence, and embedded systems. Modern robotic systems are increasingly designed to interact with humans in a natural and intuitive manner, with speech recognition playing a vital role in this evolution.

Traditional robotic systems rely on manual interfaces such as keyboards, joysticks, or mobile applications, which can be inefficient in dynamic environments. Speech-based interaction provides a more natural alternative, enabling users to communicate with robots through voice commands. Recent developments in speech-to-text technologies and artificial intelligence have significantly improved the accuracy and usability of such systems.

The proposed system utilizes a Raspberry Pi for high-level processing and an Arduino microcontroller for hardware control. Voice commands are captured through a microphone and converted into text using speech recognition algorithms. These commands are then processed and executed by the robot in real time.

IoT integration further enhances the system by enabling remote monitoring and communication. Such systems are widely used in applications like surveillance, assistive robotics, and smart environments, where real-time interaction and automation are essential.

II. Literature Survey

Speech-controlled robotic systems have gained considerable attention due to their ability to enable natural human–machine communication. Early systems relied on predefined voice commands with limited accuracy. However, advancements in speech recognition and artificial intelligence have significantly improved system performance. Several studies have explored the use of Raspberry Pi and Arduino platforms for implementing speech-driven robotic systems. These systems typically involve capturing voice commands, converting them into text, and executing corresponding actions through microcontrollers [2]. Recent research has also focused on integrating advanced technologies such as machine learning and natural language processing to improve interaction capabilities. For instance, voice-controlled robotic systems using AI models can interpret complex commands and provide intelligent responses [3]. In addition, IoT-based robotic assistants have been developed to provide real-time monitoring and remote control functionalities. These systems enhance flexibility and scalability, making them suitable for practical applications such as surveillance and smart home automation [5]. Despite these advancements, challenges such as noise interference, latency, and limited contextual understanding remain. Ongoing research aims to address these issues and improve the reliability of speech-driven robotic systems.

III. Proposed Methodology

The proposed system focuses on the design and implementation of a Speech-Driven Robotic Assistant that enables real-time human–robot communication through voice commands. The methodology consists of four main stages: speech acquisition, signal processing, command interpretation, and robotic actuation.

Initially, the user's voice input is captured using a high-sensitivity microphone module. The analog audio signal is converted into a digital format and transmitted to the Raspberry Pi, which acts as the primary processing unit. The Raspberry Pi employs speech recognition algorithms to convert the audio input into textual data using Python-based libraries or offline/online recognition engines.

The recognized text is then processed by a command interpretation module, where it is matched with predefined control instructions. These commands correspond to specific robotic actions such as forward, backward, left, right movement, stopping, and task execution. Only valid commands are executed to ensure system reliability and minimize errors.

The processed command is transmitted to the Arduino microcontroller via serial communication. The Arduino functions as the control unit and generates appropriate signals to drive the motor driver (L298N), which controls the speed and direction of the DC motors.

The robot executes the desired action in real time with minimal delay. A feedback mechanism using a speaker or display module provides confirmation messages or alerts to enhance user interaction.

The system operates in a continuous loop, ensuring real-time processing and seamless interaction between the user and the robotic system.

IV. Block Diagram

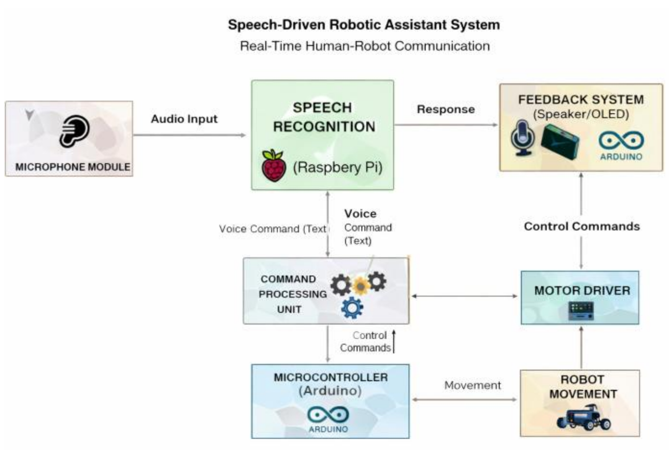

Fig. 1. Speech-driven robotic assistant block diagram

A. Overview

The proposed system is a speech-driven robotic assistant designed for real-time human–robot interaction. It integrates voice acquisition, speech processing, command interpretation, and embedded control into a unified architecture. The Raspberry Pi performs high-level tasks such as speech recognition and response generation, while the Arduino handles real-time hardware control.

The system captures user voice input through a microphone and converts it into text using speech-to-text techniques. The command is processed and mapped to control actions, which are executed by the Arduino through a motor driver to control robot movement. A feedback mechanism using a speaker and OLED display provides audio and visual responses.

The modular architecture ensures efficient performance and allows easy extension with advanced features such as IoT connectivity and intelligent decision-making.

B. Description of Block Diagram

The system architecture begins with the microphone module, which captures the user's voice input. This audio signal is sent to the speech recognition unit implemented on the Raspberry Pi, where it is converted into text using speech-to-text techniques. The recognized text is forwarded to the command processing unit, which interprets the user command and determines the appropriate action. Based on the processed command, control signals are generated and transmitted to the Arduino. The Arduino acts as a hardware controller and sends signals to the motor driver, which controls the movement of the robot. The motor driver enables operations such as forward movement, backward movement, and turning. Simultaneously, the system includes a feedback mechanism consisting of a speaker and OLED display. The Raspberry Pi generates responses using text-to-speech, which are delivered through the speaker.The OLED display can provide visual feedback. Thus, the system forms a closed-loop architecture where user input is processed and corresponding actions are executed in real time.

C. Data Flow Explanation

The data flow in the system follows a sequential pipeline:

1) Audio input is captured by the microphone
2) Speech is converted into text (STT)
3) Text is processed into control commands
4) Commands are sent to Arduino
5) Arduino controls motors via motor driver
6) Feedback is generated via speaker/display

D. Functional Modules

The system can be divided into the following modules:

1) Input Module: Microphone captures voice signals
2) Processing Module: Raspberry Pi performs speech recognition and command processing
3) Control Module: Arduino executes hardware-level control
4) Actuation Module: Motor driver controls robot movement
5) Feedback Module: Speaker and OLED provide output

E. Working Principle

The system operates by capturing voice commands through a microphone and processing them using the Raspberry Pi.

The processed commands are transmitted to the Arduino, which controls the robot's movement through a motor driver. Simultaneously, feedback is provided to the user through audio and visual outputs, enabling real-time interaction.

V. Experimental Results

The proposed system was tested to evaluate the performance of voice recognition, command execution, and real-time response. Various experiments were conducted to verify system accuracy and responsiveness.

A. Voice Recognition Testing

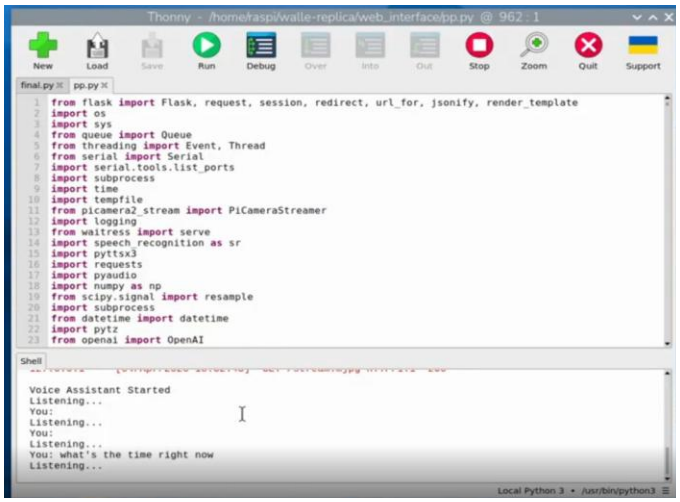

Fig. 2. Voice assistant interface and speech recognition execution on Rasp-berry Pi

The system successfully captures user voice input and converts it into text using speech recognition. The recognition accuracy was observed to be high under normal environmental conditions.

B. Text-To-Speech Output

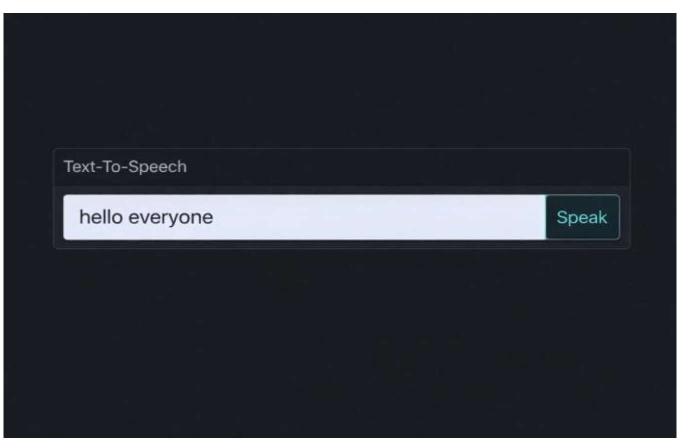

Fig. 3. Text-To-Speech Interface for Voice Output

Generation

The text-to-speech (TTS) module generates audio responses based on processed commands. The output speech is clear and audible, providing effective interaction between the user and the robot. The response time for speech generation is observed to be minimal.

C. Web-Based Control Interface

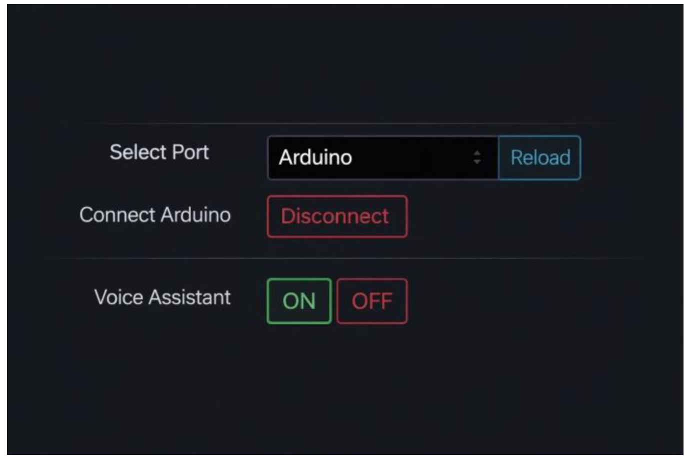

Fig. 4. Web-based control interface for Arduino communication and voice assistant control

A web-based interface was developed to control the Arduino and manage the voice assistant system. The interface allows users to connect to the Arduino, enable or disable the voice assistant, and monitor system status. The communication between the Raspberry Pi and Arduino is successfully established through this interface.

D. Notification System

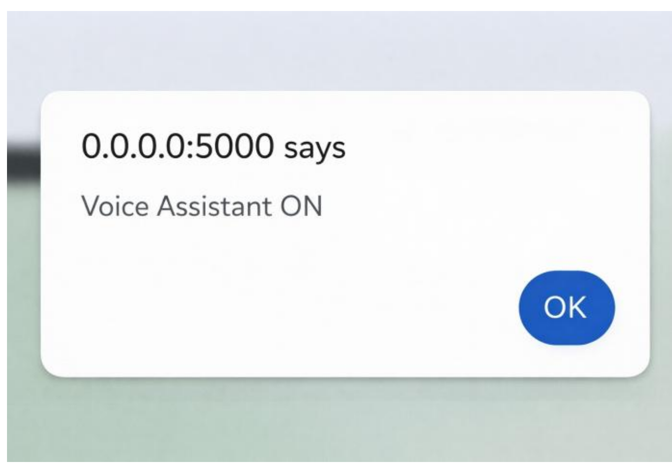

Fig. 5. Notification System

The system incorporates a browser-based notification mechanism to inform the user about the status of the voice assistant. When the voice assistant is activated or a command is successfully executed, a notification is displayed on the web interface. This provides real-time feedback to the user and enhances system usability.

The notification system ensures that users are aware of system responses without relying solely on audio output. It improves interaction by confirming successful command execution and system connectivity. Experimental results show that notifications are generated instantly with minimal delay, ensuring effective communication between the user and the system.

VI. CONCLUSION

In this paper, a Speech-Driven Robotic Assistant for real-time human–robot communication has been successfully designed and implemented. The system integrates speech recognition, embedded processing, and robotic control to enable intuitive and hands-free interaction between humans and machines. By utilizing a Raspberry Pi for speech processing and an Arduino for hardware control, the system achieves efficient command execution with minimal delay.

The developed system demonstrates reliable performance in recognizing predefined voice commands and executing corresponding robotic actions such as navigation and task control. The inclusion of feedback mechanisms further enhances user interaction, making the system more responsive and user-friendly.

Overall, the proposed solution provides a cost-effective and scalable approach for real-time voice-controlled robotic systems. It can be effectively applied in areas such as assistive robotics, smart surveillance, and automation. Future enhancements involving artificial intelligence and advanced speech processing techniques can further improve system accuracy, adaptability, and functionality.

ACKNOWLEDGMENT

We would like to thank Design Space lab, Department of Electronics and Communication Engineering at Nadimpalli Satyanarayana Raju Institute of Technology (NSRIT) for providing the resources and support required for this research work.

REFERENCES

[1] J. R. Deller, J. G. Proakis, and J. H. L. Hansen, *Discrete-Time Processing of Speech Signals*. IEEE Press, 2000.
[2] A. Sehgal et al., “Speech Recognition Using Raspberry Pi,” *International Journal of Computer Applications (IJCA)*, 2017.
[3] D. Jurafsky and J. H. Martin, *Speech and Language Processing*. Pearson, 2019.
[4] X. Huang et al., *Spoken Language Processing*. Prentice Hall, 2001.
[5] A. Bahga and V. Madisetti, *Internet of Things: A Hands-On Approach*. 2015.
[6] T. Young et al., “Recent Trends in Deep Learning for NLP,” *IEEE Computational Intelligence*, 2018.
[7] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*. MIT Press, 2016.
[8] O. Vermesan and P. Friess, *IoT Technologies*. 2013.
[9] D. Evans, “The Internet of Things,” Cisco, 2011.
[10] W. Shi et al., “Edge Computing: Vision and Challenges,” *IEEE Internet*
[19]

of Things Journal, 2016.
[11] M. Satyanarayanan, “The Emergence of Edge Computing,” *IEEE Computer*, 2017.
[12] M. Quigley et al., “ROS: Robot Operating System,” 2009.
[13] B. Siciliano and O. Khatib, *Springer Handbook of Robotics*. Springer, 2016.
[14] R. Szeliski, *Computer Vision: Algorithms and Applications*. 2010.
[15] Y. Chen et al., “Sensor Fusion in Robotics,” IEEE, 2015.
[16] S. Boll, “Suppression of Acoustic Noise in Speech,” IEEE, 1979.
[17] H. Hermansky, “Perceptual Linear Prediction (PLP),” *Journal of the Acoustical Society of America*, 1990.
[18] A. Graves et al., “Speech Recognition with Deep Neural Networks,” IEEE, 2013.
[19] G. Hinton et al., “Deep Neural Networks for Acoustic Modeling,” *IEEE Signal Processing*, 2012.
[20] Y. LeCun, Y. Bengio, and G. Hinton, “Deep Learning,” *Nature*, 2015.