© 2026 ADYPJIET, Volume 1(1), Online ISSN: 3049-1517, June 2026, pp. 14-23

Ajeenkya DY Patil Journal of Innovation in Engineering & Technology

Journal Homepage: https://www.adypsoe.in/adypjiet

Integrated Speech and Gesture-Based Human–Machine Interaction Framework for Robotic Vehicles

[1]Dr. Ankush Kadu,
[2]Tanuja Abhang,
[3]Adityaraj Deokar,
[4]Divyansh Singh,
[5]Shubham Dhavane

[1] Assistant Professor, Ajeenkya D Y Patil School of Engineering, Pune, INDIA
[1-5]UG student, Ajeenkya D Y Patil School of Engineering, Pune, INDIA

Article History:
Received: 22/05/2026
Revised: 12/06/2026
Accepted: 22/06/2026

Abstract:
Human–Machine Interaction (HMI) has significantly contributed to developing intelligent robotic systems capable of responding to multimodal commands. The paper presents an Integrated Speech and Gesture-Based HMI Framework designed for robotic vehicle control, enabling natural, intuitive, and contactless interaction between humans and machines. The proposed system combines speech recognition and hand gesture detection modules to enhance flexibility and operational efficiency in dynamic environments. Speech commands are processed using real-time voice recognition techniques, while gestures are captured through motion sensors or vision-based systems. The integration of both modalities ensures robust performance even under noisy or ambiguous conditions, offering improved accuracy and user convenience. Experimental results demonstrate that the framework achieves reliable command execution, low response latency, and adaptability across various scenarios. The approach holds potential for applications in assistive robotics, industrial automation, and defense operations, where hands-free and efficient communication with robotic systems is essential.

Keywords: Human–Machine Interaction (HMI), Speech Recognition, Gesture Control, Robotic Vehicle, Multimodal Interface, Real-Time System

© 2026 ADYPJIET, Volume 1(1), Online ISSN: 3049-1517, June 2026, pp. 14-23

1.0 Introduction

Robotic systems have become an integral part of modern technology and are widely used in industrial automation, healthcare, defense, surveillance, agriculture, and service applications. As robots continue to evolve from simple automated machines to intelligent systems, the demand for effective Human–Machine Interaction (HMI) has increased significantly. An efficient HMI system enables users to communicate with robots naturally and intuitively, improving operational efficiency, safety, and user experience. Researchers have explored various speech-processing techniques to enhance communication between humans and machines, including speech recognition systems based on Linear Predictive Coding (LPC), Artificial Neural Networks (ANN), and mel-scaled feature extraction methods for recognizing spoken commands and human speech patterns [1,2].

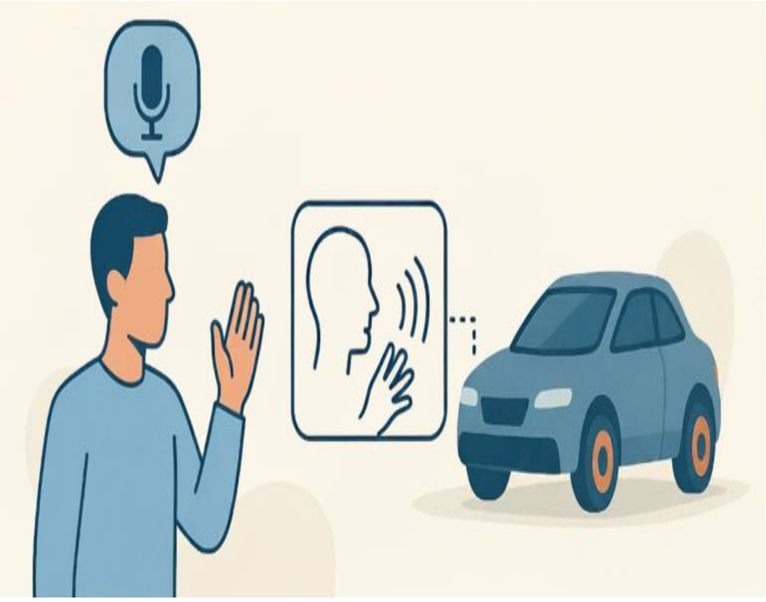


Figure1. Multimodal robotic vehicle

Speech recognition has emerged as one of the most convenient methods of robot control because it allows users to issue commands without physical contact. Several studies have demonstrated the successful implementation of voice-controlled robotic systems capable of interpreting spoken instructions and performing corresponding actions [3-5]. The advancement of low-cost embedded computing platforms, such as Raspberry Pi and microcontroller-based systems, has further accelerated the development of intelligent speech-enabled robots by providing sufficient computational power for real-time processing while maintaining affordability [4-5].

In addition to speech-based interaction, gesture recognition has gained considerable attention as an alternative communication method. Human gestures provide a natural and intuitive means of conveying intentions, particularly in environments where speech recognition may be affected by background noise or communication constraints. Accelerometer-based gesture control techniques have been successfully applied in robotic arm manipulation and industrial robotic systems, enabling machines to interpret human movements and convert them into control commands [6-8]. These systems improve accessibility and allow users to interact with robotic devices through simple hand or body movements.

Recent developments in embedded systems and Internet of Things (IoT) technologies have enabled the integration of multiple communication modalities within a single robotic platform. Raspberry Pi-based robotic monitoring and control systems have demonstrated the feasibility of combining wireless communication, web interfaces, and intelligent control mechanisms for enhanced robot operation [9]. Similarly, Arduino-based robotic platforms have been utilized to implement voice-controlled robotic arms and embedded control applications due to their simplicity, flexibility, and cost-effectiveness [10].

Although speech and gesture recognition techniques have individually shown promising results, each approach has certain limitations. Speech-based systems may experience reduced accuracy in noisy environments, while gesture-based systems can be influenced by sensor errors, environmental conditions, or user movement variations. To address these challenges, researchers have increasingly focused on multimodal interaction frameworks that combine speech and gesture inputs into a unified control system [3,6,7]. By utilizing complementary input methods, the overall reliability, responsiveness, and adaptability of the robotic system can be significantly improved.

Therefore, the development of an integrated speech- and gesture-based human–machine interaction framework offers a practical solution for controlling robotic vehicles in real-time applications. Such a system enables users to communicate with robots using natural human actions while improving operational flexibility and reducing dependence on traditional control devices. The integration of speech recognition, gesture sensing, embedded controllers, and intelligent communication technologies provides a foundation for the next generation of interactive robotic systems capable of operating efficiently in industrial, assistive, and autonomous environments [3,6,10].

Muslima Umme and M. Babul Islam [1] presented a Bangla speech recognition framework using Mel-scaled Linear Prediction techniques. Their research focused on improving speech feature extraction and recognition accuracy for human–computer interaction systems. The study demonstrated that speech-processing algorithms can effectively identify spoken commands and form the basis for voice-controlled robotic applications.

Anup Kumar Paul, Dipankar Das, and Md. Mustafa Kamal [2] proposed a Bangla speech recognition system based on Linear Predictive Coding (LPC) and Artificial Neural Networks (ANN). Their work highlighted the importance of machine-learning techniques in enhancing speech recognition accuracy and robustness, making voice-based communication more reliable.

2.0 Literature Survey

Authors	Year	Technology Used	Findings
Adwait P. Naik and Annu Abraham [6]	2018	Arduino, Voice Recognition	Developed a voice-controlled robotic arm using Arduino. The system successfully converted voice commands into robotic movements, demonstrating a low-cost and user-friendly human–robot interaction approach.
P. Adeeb Ahammed and K. Edison Prabhu [7]	2016	Accelerometer-Based Gesture Control (similar to MPU6050 applications)	Proposed a robotic arm controlled through human hand movements using accelerometer sensors. The work showed that gesture-based control can improve intuitive robot operation and reduce dependence on conventional controllers.
Widodo Budiharto and Derwin Suhartono[3]	2015	Voice Recognition Robot	Designed an intelligent service robot capable of voice recognition and telepresence functions. The robot responded to spoken commands and demonstrated improved human–robot communication capabilities.
Pandapotan Siagian and Kisno Shinoda[10]	2015	Raspberry Pi	Developed a web-based monitoring and robotic arm control system using Raspberry Pi. The study demonstrated remote robot operation through internet-based communication and embedded processing.
Luis Fernando D. Haro et al.[4]	2014	Raspberry Pi, Speech Recognition	Implemented low-cost speaker and language recognition systems on Raspberry Pi, proving that embedded platforms can efficiently perform speech-processing tasks at low cost.

Dhanny Wijaya Thiang [5] developed a mobile robot control system using speech recognition implemented on an ATmega162 microcontroller. The system successfully interpreted limited voice commands and translated them into robot movements, demonstrating the feasibility of microcontroller-based voice-controlled robotics.

Mohammad Javed Ansari, Ali Amir, and Md. Ahsanul Hoque [8] designed a microcontroller-based robotic arm capable of operating in both gesture-controlled and automated modes. Their research demonstrated that combining human gestures with automated functionalities can increase operational flexibility and improve robot usability in industrial applications.

Pedro Neto, J. Norberto Pires, and A. Paulo Moreira [9] presented an accelerometer-based control mechanism for an industrial robotic arm. The study showed that human hand movements could be accurately translated into robotic actions, providing a natural and intuitive method for robot control. The findings established the foundation for modern gesture-recognition systems that commonly employ sensors such as the MPU6050.

3.0 Gap Analysis and Observations

The reviewed literature shows that speech recognition and gesture recognition have been widely used for robotic control. However, most existing systems rely on a single interaction method, either voice commands or gesture inputs. Voice-controlled systems may be affected by environmental noise, while gesture-based systems can experience sensor inaccuracies and user-dependent variations. Additionally, many studies focus on robotic arms rather than mobile robotic vehicles. Therefore, there is a need for an integrated framework that combines both speech and gesture recognition to improve reliability, flexibility, and overall human–machine interaction. The proposed system addresses this gap by providing multimodal control for robotic vehicle applications.
Most existing robotic control systems use either speech recognition or gesture recognition independently. Integrating both methods can enhance control accuracy, system robustness, and user convenience, especially in real-time robotic vehicle applications.

Table No. 1: **Gap Analysis**

Author(s)	Gap Identified	Analysis
Adwait P. Naik and Annu Abraham (2018)	System relies only on voice commands.	Voice-based control is effective in quiet environments, but performance may degrade due to noise, accent variations, or speech recognition errors. No alternative control mechanism is provided.
P. Adeeb Ahammed and K. Edison Prabhu (2016)	Uses only gesture-based control.	Gesture recognition improves intuitive interaction but may suffer from sensor drift, unintended movements, and user fatigue during prolonged operation.
Widodo Budiharto and Derwin Suhartono (2015)	Focuses primarily on voice recognition and telepresence.	The system lacks multimodal interaction, reducing reliability when speech commands cannot be recognized accurately.
Pandapotan	Emphasis on web-based	Remote control capabilities are provided, but

Siagian and Kisno Shinoda (2015)	monitoring rather than natural human interaction.	direct human–robot communication through speech or gestures is limited.
Luis Fernando D. Haro et al. (2014)	Concentrates on speech processing rather than robotic control.	Demonstrates Raspberry Pi-based speech recognition but does not integrate gesture recognition or vehicle navigation functions.
Muslima Umme and M. Babul Islam (2014)	Focuses only on speech recognition algorithms.	The work improves recognition accuracy but does not address real-time robotic applications or multimodal control.
Mohammad Javed Ansari et al. (2014)	Limited integration between gesture control and intelligent decision-making.	The robotic arm operates using predefined actions without combining speech-based commands for enhanced flexibility.
Dhanny Wijaya Thiang (2009)	Supports only a limited set of voice commands.	The system demonstrates feasibility but lacks scalability and adaptability for complex robotic operations.
Pedro Neto et al. (2009)	Uses accelerometer-based control only.	Human motion is effectively captured, but the absence of speech interaction reduces communication flexibility.
Anup Kumar Paul et al. (2009)	Research focuses on speech recognition accuracy.	The study contributes to speech-processing techniques but does not address robotic control or multimodal human–machine interaction.

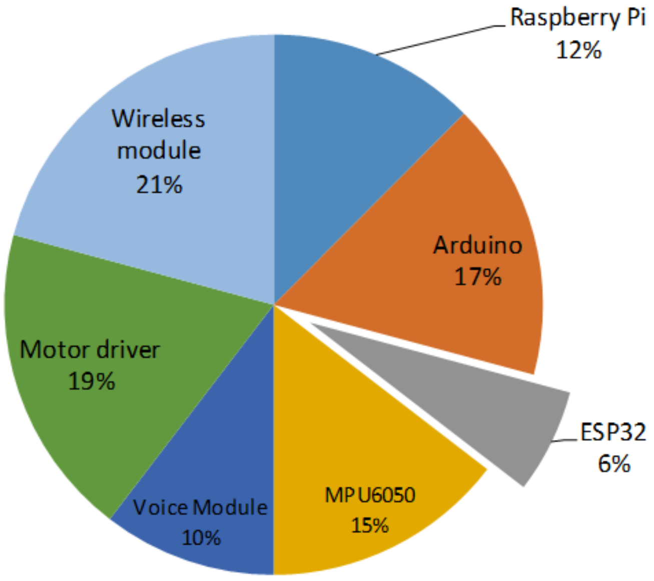


Fig 2: Component Usage Across 20 Journals

© 2026 ADYPJIET, Volume 1(1), Online ISSN: 3049-1517, June 2026, pp. 14-23

The pie chart highlights the usage of different components across 20 journals. It shows that wireless modules and motor drivers are the most widely used, forming the core of robotic vehicle systems. Arduino is used more often in simpler setups, while Raspberry Pi appears in projects needing higher processing power. Gesture recognition with MPU6050 is common, but voice modules are less frequently used. Overall, the chart suggests that combining gesture and speech control is still less explored, making our project unique and relevant

4.0 Methodology

A. System Design

The system uses an ESP32 microcontroller to process inputs from a microphone and IMU/accelerometer sensor, sending control signals to the motor driver to move the robotic vehicle. Initial testing is done on a breadboard, with final implementation on a PCB for compactness and reliable operation.

B. Hardware Components:

1. **ESP32-WROOM-32:**
   Central microcontroller for processing inputs from sensors and controlling the robotic vehicle.
2. **MPU6050:**
   IMU sensors used to detect gestures and motion for gesture-based control.
3. **ElecHouse V3:**
   Captures voice commands for speech-based interaction with the vehicle.
4. **Raspberry Pi 4:**
   Acts as a secondary processing unit, possibly for advanced tasks like image processing or IoT connectivity.
5. **TB6612FNG Motor Driver:**
   Controls the DC motors on the robotic chassis based on signals from the ESP32.
6. **4WD Robotic Chassis:**
   Provides locomotion platform for the robot, including wheels and mounting structure for components.
7. **Power Supply:**
   Supplies power to both the robotic chassis and the hand-held control unit, including ESP32, Raspberry Pi, and sensors.
8. **Switches & Connectors:**
   Used for power control, signal routing, and modular connections between components.

© 2026 ADYPJIET, Volume 1(1), Online ISSN: 3049-1517, June 2026, pp. 14-23

9. **Jumper Wires / PCB:**
   Facilitates connections during prototyping and final assembly on the PCB.

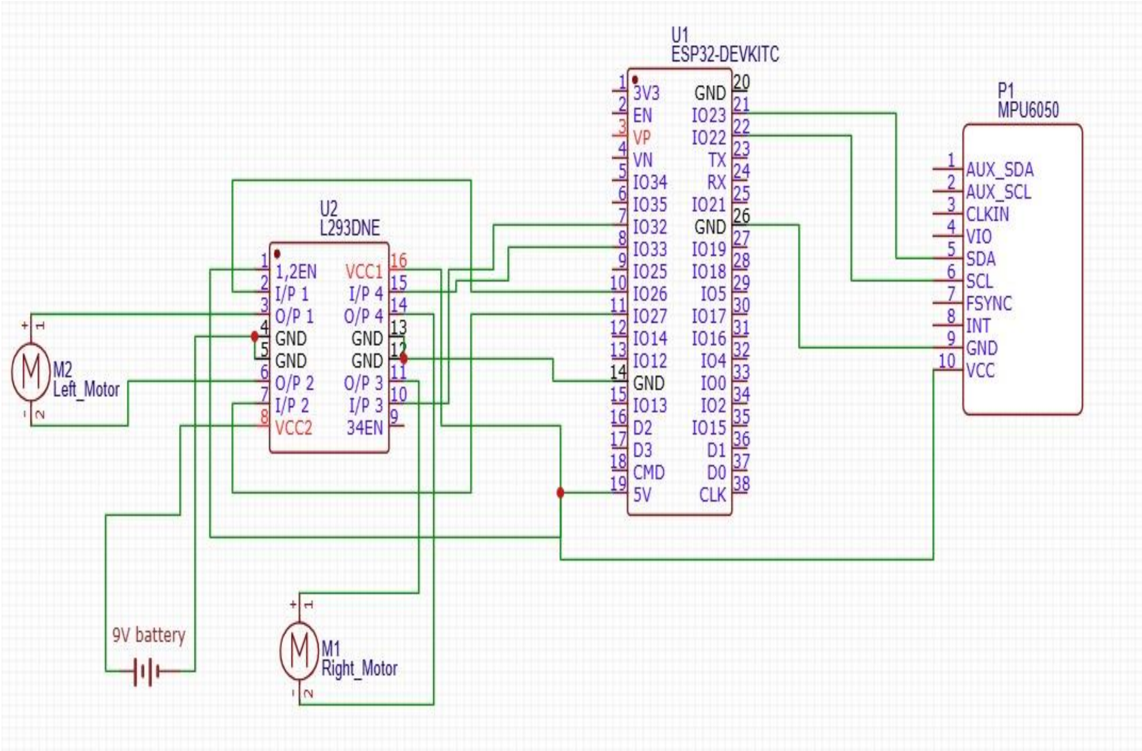
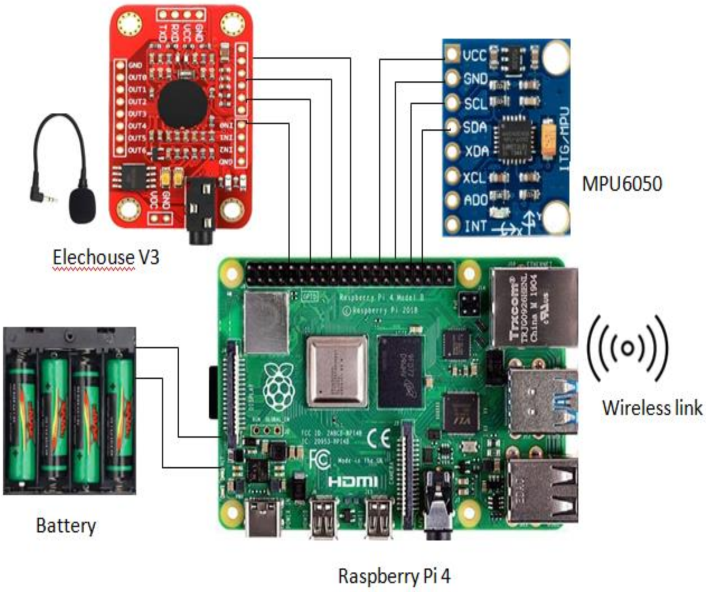


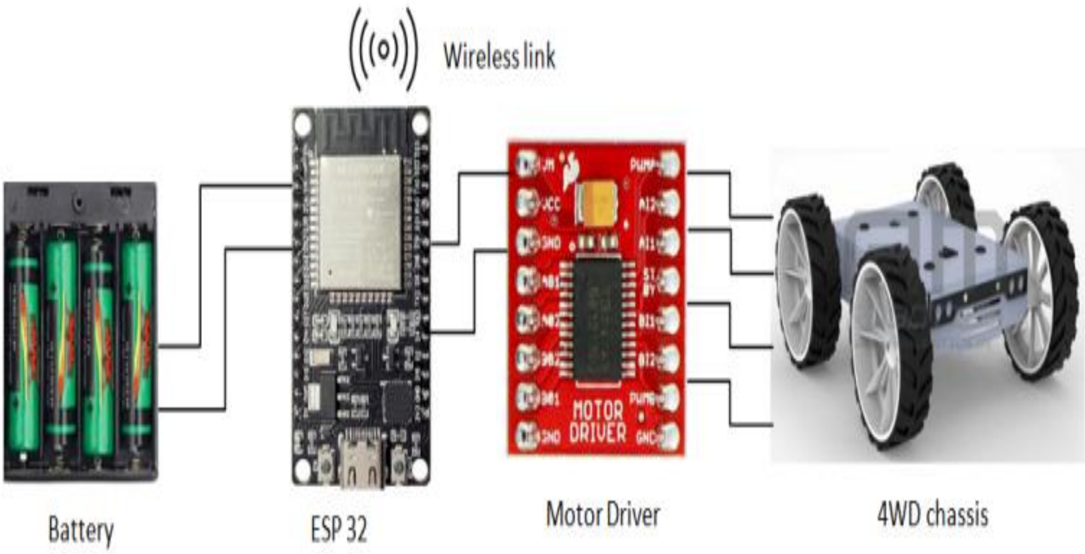
Figure 2. Schematic Digram

The Robot Unit schematic was designed and simulated in Proteus **t**o represent the complete control and actuation section of the multimodal robotic vehicle. It consists of the Raspberry Pi 4 as the main controller, which receives wireless signals from the Hand Unit through the nRF24L01 receiver module**.** The received data is processed and decoded into corresponding motion commands such as forward, backward, left, right, or stop. These control signals are then sent to the TB6612FNG motor driver**,** which drives the 4WD robotic chassis using four DC motors connected in differential configuration. A regulated Li-ion/Li-Po battery pack powers both the controller and the motor driver, ensuring stable voltage and current supply. The schematic also includes indicator LEDs and connectors to visualize signal transmission and motor direction during simulation. The Proteus design helps verify correct wiring, logic flow, and voltage levels before physical implementation, ensuring reliable operation of the robotic vehicle during testing and deployment.

5.0 Architectural View of the Proposed Voice and Gesture Controlled Robotic Vehicle

The overall system architecture integrates two major units Hand Unit and Robot Unit connected through a wireless communication link**.** The Hand Unit includes the ESP32 microcontroller**,** MPU6050 gesture sensor, and Elechouse V3 voice module. It processes both gesture and voice commands locally and transmits encoded signals to the robot using the nRF24L01 transceiver**.**





The Robot Unit is built around the Raspberry Pi 4**,** which acts as the main controller. It receives commands wirelessly, verifies them through a command decoder**,** and sends control signals to the TB6612FNG motor driver that operates the 4WD robotic chassis. Power is supplied through a Li-ion or Li-Po battery pack. Optional modules like LEDs or buzzers provide system feedback, and future versions may include sensors for obstacle detection and visual navigation. The architectural flow ensures modularity, scalability, and low-latency communication between both units, creating a compact and efficient human–machine interaction system.

6.0 Conclusion

The project focuses on building a low-cost and efficient system using Raspberry Pi 4, ESP32, MPU6050, Elechouse V3, TB6612FNG motor driver, and a 4WD robotic chassis to control a robot using voice and hand gestures. This kind of system can be useful for real-time embedded applications where it needs to work reliably without internet. The main idea is to create a robot that is easy to control and can be used for automation, assistive technology, and learning purposes.

References:

1. Adwait P. Naik, Annu Abraham, Arduino-Based Voice Controlled Robotic Arm, 2018.
2. P. Adeeb Ahammed, K. Edison Prabhu, Robotic Arm Control Through Human Movement Using Accelerometers, 2016.
3. Widodo Budiharto, Derwin Suhartono, Intelligent Service Robot with Voice Recognition and Telepresence Capabilities, 2015.
4. Pandapotan Siagian, Kisno Shinoda, Web-Based Monitoring and Control of Robotic Arm Using Raspberry Pi, 2015.
5. Muslima Umme, M. Babul Islam, Experimental Framework for Mel-Scaled LP-Based Bangla Speech Recognition, 2014.
6. Mohammad Javed Ansari, Ali Amir, Md. Ahsanul Hoque, Microcontroller-Based Robotic Arm Operational to Gesture and Automated Mode, 2014.
7. Anup Kumar Paul, Dipankar Das, Md. Mustafa Kamal, Bangla Speech Recognition System Using LPC and ANN, 2009.
8. Dhanny Wijaya Thiang, Limited Speech Recognition for Controlling Movement of Mobile Robot Implemented on ATmega162 Microcontroller, 2009.
9. Pedro Neto, J. Norberto Pires, A. Paulo Moreira, Accelerometer-Based Control of an Industrial Robotic Arm, 2009.
10. Luis Fernando D. Haro et al., Low-Cost Speaker and Language Recognition Systems Running on a Raspberry Pi, 2014.