Voice Controlled Robot: Using Speech Recognization

Nadimpalli Harshith Varma
Department of Computer Science
Amrita Vishwa Vidyapeetham
Chennai, India
harshithvarmanadimpalli27@gmail.com

Natra Charan
Department of Computer Science
Amrita Vishwa Vidyapeetham
Chennai, India
natracharan@gmail.com

Muli Likith Reddy
Department of Computer Science
Amrita Vishwa Vidyapeetham
Chennai, India
likhithreddymuli@gmail.com

C.Mohith Reddy
Department of Computer Science
Amrita Vishwa Vidyapeetham
Chennai, India
cheennepallimohithreddy@gmail.com

***Abstract*—This project consists of a speech-controlled robot system with hands-free operation via voice command for increased convenience and user-friendliness. The system involves a web interface by means of which the user provides voice commands that are translated into text through a speech-to-text converter module. The text is authenticated and transmitted to a microcontroller via a Bluetooth module. Upon receiving the command, the microcontroller decodes it and transmits corresponding control signals to the motor driver. The motor driver propels the motors to deliver smooth actions such as going forward, going backward or turning around the opposite way. All of which is performed in real-time for the dynamic and smooth motion of the robot.**

***Index Terms*—Voice Control, Speech Recognition, Bluetooth Communication, Microcontroller, Motor Driver, Robotics, Automation, Assistive Technology, Remote Navigation, Hands-Free Operation, Wireless Control.**

I. Introduction

Merger of the Internet of Things (IoT) and robotics has changed the traditional automa- tion through real-time communication, remote control, and autonomous decision-making. IoT provides the means by which robots can respond to the environment by a set of sensors and actuators to promote efficiency and flexibility. Intercon-nectedness makes the robotic system remotely monitored and controlled through web-based or mobile applications to reduce human interface. Robs can improve in performance by making them suitable for application in industrial automation, health, and smart houses. Real-time data collection and processing enhance robotic ability to be smart and accurate [1]

The advent of high-speed internet technologies like 4G LTE has enabled robotics based on IoT with new features, including enhanced speed and reliability of communication between the devices. Voice-controlled robotics is one of the most prominent applications of this technology, where robots are operated by voice commands transmitted through wireless networks. Enhanced interconnectivity allows robotic systems to be responded to in real-time by inputs, enhancing efficiency in assistive robotics, industrial control, and home automation. Robotics and IoT allow real-time monitoring, remote diagnosis, and predictive maintenance, enhancing round-the-clock operation of autonomous robotic systems. Latency is minimized and data trans-mission rate enhanced, hence enhancing scalability of robots in automation. [2]

IoT robotics goes beyond connectivity because it enables sensor-based autonomous decision-making. An up-to-date robotic system has a network of sensors, receives envi- ronmental information, and responds based on such information processed by microcontrollers. This is particularly handy in smart home robot applications in the sense that robots are able to sense the environment and manipulate devices such as lights, fans, and security systems. Furthermore, IoT-driven robotic platforms take advantage of cloud-based processing of information to enhance their decision-making capability even further so that they are capable of performing better in adaptive environments. AI empowered with IoT-based robotics further enhances the automation process by enabling self-learning systems, which enable robots to be more adaptive in new environments. [3]

A novel method has been proposed to optimize the control and monitoring experience by associating sensor information with user activity and optimized remote control. With this method, users are capable of remote control of domestic appliances provided there is ample ambient monitoring and processing. The sensor information are processed in a microcontroller-based system and transmitted via a web server to a mobile application, ensuring seamless communication and real-time responsiveness. With the integration of these technologies, IoT is also enhancing its abilities, offering smarter and more efficient automation solutions for daily use. Increased development of IoT infrastructure and control systems is expected to accelerate further innovations across various industries, making intelligent automation more feasible and cost-effective for a large user base. [4]

II. Literature Review

Speech Emotion Recognition (SER) is also a critical component of human-robot interaction (HRI) since it offers the ability of emotion recognition through speech. Standard SER models in laboratory settings have good performance and can perform poorly with farfield speech, background noise, and reverberations, leading to degradations of accuracy. With increasing distance between the speaker and microphone, speech signals degrade, while noise and reverberations dominate. Against these, algorithms like Delay-and-Sum (D and S) and Minimum Variance Distortionless Response (MVDR) enhance speech intelligibility by focusing the speaker's voice and removing interference. Under dynamic HRI with user and robot movement, the changing acoustic environments also influence SER performance. Researchers have applied adaptive signal processing and simulated acoustic environment data augmentation to make the model strong. By training SER models under different levels of noise and reverberation, they are rendered robust against real-world challenges. With this synergy of cutting-edge signal processing and acoustically rich training, SER systems are rendered more accurate and dependable. Improved SER in HRI translates to more natural and effective human-robot interaction, enabling robots to respond well to human emotions. Such improvements render speech-based interfaces more feasible for real-world use. [5]

Speech recognition has departed from traditional machine learning methods like Hidden Markov Models (HMMs) and Gaussian Mixture Models (GMMs) to deep learning-based systems for better performance and accuracy in noisy environments. Deep Neural Networks (DNNs), Convolutional Neural Networks (CNNs), and Recurrent Neural Networks (RNNs) have enabled speech recognition to be enhanced, particularly for car driving control, reducing driver distraction and improving safety. Alternatively, there are challenges with managing environmental noise, command discrimination, and speed of processing in a way to keep it in real time. Hybrid techniques such as RNN-CNN hybrids have already shown to improve performance, and sensor data fusion may have the potential to bring even more accurate results. While this is okay, bias, generalizability, and large data sensitivity remain issues, and more needs to be done for deep models to successfully transfer to actual driving conditions. [6]

The article ”Robot Voice: A Voice Controlled Robot using Arduino” reports the development of a voice-controlled robot personal assistant capable of implementing human voice commands with the use of the in-built microphone and rendering associated operations and vocal responses. The robot is able to implement various functions such as motion, relocation of objects, and simple dialogues. The system translates voice commands in real-time with the assistance of an offline server, speech signals being transferred via a USB connection. Operating on a microcontroller-based system, the robot uses voice recognition technology to convert speech into actionable commands. The paper determines the potential uses of such robotic assistants in different areas like home automation, healthcare, automotive systems, and industrial processes for the reduction of manual labor in daily tasks. [7]

Speech-controlled robot systems are gaining popularity in the healthcare industry, mainly to use for cognitive training and rehabilitation. Deep learning and artificial intelligence, particularly Convolutional Neural Networks (CNNs) such as VGG-16, have enhanced speech recognition. Such systems are used for medical interventions, including orthopedic surgery and minimally invasive interventions. Studies show that robot systems with cognitive tasks, such as board games, increase cognition stimulation in people with disability. Voice-controlled robots increase independence for mobility-impaired people to use them on their own. Noise-cancelling technology and adaptive artificial intelligence systems increase the effectiveness of such technology. Successful speech recognition makes communication by care robots a seamless process. Robotics and artificial intelligence have promising applications for the prevention of dementia and cognitive rehabilitation. Multimodal interaction makes therapy robots more responsive. In general, the technologies enhance people's quality of life and accessibility in case of severe disabilities. [8]

Recent developments in robotics and artificial intelligence have significantly changed early childhood education through more natural human-robotic interaction. Previous learning robots have traditionally used programming languages, which are difficult for children and restrict their experimentation and interaction. Studies recognize natural and intuitive interfaces, such as voice interaction, as being essential in promoting direct interaction between children and robots. Experiments have shown that the combination of speech-to-text systems, large language models, and vision-based object detection improves the recognition and responsiveness of robots to child-directed commands. Though collaborative robots (cobots) have long been applied in manufacturing environments, their application in education is becoming even more prized in the area of developing STEM-related skills. Yet, there are issues in this sector too about how to preserve robust and enduring voice recognition programs that can tackle the linguistic abilities and mental capability of young children. With the overcoming of these defects, voice-controlled robotics sets a hopeful path to the enhancement of early schooling through even more engaging and inclusive lessons of learning. [9]

Voice recognition technology has improved considerably, providing real-world automation with voice commands, especially in robotics. This work proposes a voice-controlled robot using Bluetooth communication, an Android app, and an Arduino-based controller to perform pre-defined commands like movement, lighting, and horn blinking. The system utilizes real-time monitoring with an ultrasonic sensor for detecting obstacles, thereby providing autonomous navigation. In comparison to GSM-based communication, Bluetooth is faster and more economical. The control system of the robot is powered by an L293D motor driver and relay module for increased functionality. Verification of circuit connections prior to hardware is accomplished through Proteus simulation software. The voice command system, implemented with MIT App Inventor, interprets user commands and carries them out through Arduino. Future improvements involve improving speech recognition accuracy through vector embeddings for adaptive command interpretation, increasing the system's robustness for multiple-user interfaces and extended automation applications. [10]

Voice-controlled robots have also been created as new technology that extends the human-to-robot communication by providing simplicity and smooth execution of commands by speech recognition. Traditional robotic control systems are generally manual, and this makes it difficult for users to access them and is inconvenient. Recent advancements in speech recognition, artificial intelligence, and microcontroller-based technologies have helped develop robots that can comprehend and respond to voice commands with great accuracy. Addition of deep learning models such as Singular Value Decomposition (SVD) and Principal Component Analysis (PCA) has also increased speech processing to allow real-time execution of commands. These technologies have enabled various applications in fields such as automation, assistive technology, and interactive robotics. Issues such as noise interference, multilingual recognition, and efficiency of real-time processing remain vital research areas in spite of advances. Continuous advancement of machine learning and natural language processing will optimize voice-controlled robot systems for sensitivity, adaptability, and performance for innumerable applications. [11]

The human action recognition and speech recognition literature highlights growing artificial intelligence (AI) use to enable natural interaction with smart home systems. Current research aims at using machine learning algorithm approaches to categorize human actions on the basis of verbal commands to achieve maximum automation and ease of use. Speech-to-numeric feature conversion using feature extraction techniques like CountVectorizer enables the numeric features to allow for proper classification of categories, subcategories, and actions. Classifiers such as Naive Bayes and Random Forest are highly accurate, confirming the viability of AI in user intent detection. Previous research sets up issues with emotion recognition, interference by noise, and context comprehension influencing the performance of identification. The union of Internet of Things (IoT) and AI introduces additional automation through devices that can respond dynamically when called upon to do so by voice commands. Smart personal assistants apply deep learning to enhance command interpretation, facilitating home automation, security, and assistive technology. Multilingual capabilities and adaptive learning are also investigated to make interactions tailored. As research into AI continues, the optimization of recognition models for different user needs remains an area of research interest, offering consistency in real-world applications. [12]

Voice-controlled robot cars have garnered significant attention in automation and assistive technology as they have the potential to enhance human-machine interaction. Traditional robot systems employ manual or remote control, whereas voice control is a more natural and easier method, particularly for the disabled. Advances in speech recognition and Bluetooth communication have enabled voice commands to be combined with microcontroller-based platforms like Arduino. There have been a variety of research works on deploying speech-controlled robot systems, focusing on their applications in surveillance, search and rescue operations, automation of industries, and healthcare. Machine learning algorithms have also helped increase the voice recognition performance to a greater extent, enabling robots to execute orders without errors even in noisy environments. Interference of background noises, restricted range of Bluetooth communications, and varying speech patterns remain to be research issues. With the advancement of speech recognition algorithms and incorporation of advanced communication technology, voice-controlled robotic cars can be further optimized for useful application, enhancing efficiency and ease of use in a variety of industries. [13]

Speech-based robot platforms have revolutionized automation through natural computer-human conversation in the process of speech recognition. Combining microcontroller boards like Arduino with advanced methods like Mel Frequency Cepstrum Coefficients (MFCC) and Dynamic Time Warping (DTW) has elevated speech recognition levels. Utilization of Bluetooth communication, wireless cameras, and sensors for obstacles has expanded areas of application in surveillance, medical practice, and harmful environments. Despite advances, problems of interference due to noise and narrow communication range remain. Present research is focused on AI-driven innovations and sophisticated real-time processing towards making voice-commanded robots faster and more efficient. [14]

Speech-enabled robot systems combine speech recognition and automation to facilitate effective remote control. Arduino microcontrollers, RF modules, and motor drivers allow for wireless control of robots, with Microsoft Visual Studio being used to implement a GUI for voice command processing. RF communication provides low-cost and reliable short-range control. Robots like these are engineered to be used in dangerous environments, industrial automation, and assistive technologies. Speech recognition is impaired by noise, and signal processing must be improved. DC gear motors improve torque efficiency for smooth motion. Future developments in AI and IoT will enhance adaptability, accuracy, and responsiveness in voice-controlled robots. [15]

Recent studies on human-drone interaction have been directed toward the creation of multilingual speech recognition systems for improving drone control. Conventional approaches with RF controllers are cumbersome, leading to the adoption of natural user interfaces such as voice commands. This paper presents a deep learning-based automatic speech recognition system for English, Arabic, and Amazigh for UAV control. The model, having learned from more than 38,000 samples, is highly accurate, at 93.76% for English, 88.55%for Arabic, and 82.31% for Amazigh. The system is resilient to background noise, particularly that of drone propellers, making command execution trustworthy. A graphical user interface is also used to ease interactions, showing identified commands and UAV status. Hardware deployment on a quadrotor UAV proved real-time efficiency, verifying the model's applicability. Comparison with current approaches shows enhanced accuracy and usability. Ongoing research plans to increase the command dictionary and improve adaptability for different languages and environments. [16]

Smart home technology is progressing with voice commands through NLP techniques, making it more accessible and user-friendly. Current platforms such as Google DialogFlow and Microsoft LUIS are not without limitations, so an adaptable IoT-Fog-Cloud (IoTFC) architecture is called for. In this paper, a system is presented that can handle voice commands locally or in the cloud to minimize latency and maximize privacy. It can handle under-resourced languages and provide a low-cost alternative to industrial NLP systems. Real-world testing validates its dependability for smart home automation. Future development will improve AI-powered intent detection and multimodal interaction for hands-free automation. [17]

Discusses different types of robot control, contrasting autonomous, manual, and voice-based systems. Refers to the challenges of decision-making in real time and wired control inefficiency. Speech-to-text and Bluetooth-based robotic control are supported by research, while smartphone integration is justified. Research on motion control, PID controllers, and microcontrollers embedded as a feasibility confirmatory measure reaffirms feasibility. Voice assistants are prioritized in enhancing human-robot interaction. RF-based robots and joystick-operated robots are contrasted regarding usability. The review emphasizes the need for hands-free operation in dangerous environments. In general, it warrants the necessity of a user-friendly, voice-controlled robotic system. [18]

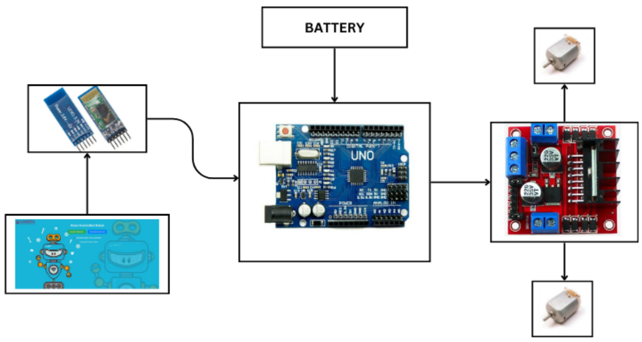

Fig. 1: Work Flow Diagram

III. Methodology

A. User Input via Web Page

The system begins with an interactive web page in which the user inputs voice commands. The page is programmed with HTML, CSS, and JavaScript for ease of use. A feature to read access to the microphone is added in the page so that speech can be recorded real-time. The interface also includes visual feedback, displaying the recognized command for verification of the user before processing.

B. Speech-to-Text Conversion and Validation

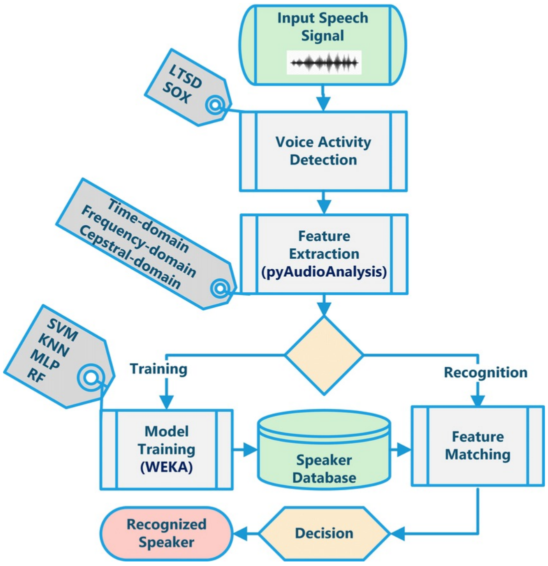

Fig. 2: Speech to Text Conversion

When input is received by the user in the form of spoken voice commands the speech is converted into text to increase accuracy.

C. Command Transmission via Bluetooth

After it is received for testing, the text command is sent remotely to the microcontroller via a Bluetooth module (HC-05). The wireless transmission takes place in the order:

*1) **Command Encoding**:* The text command is formatted into the transmission format using Bluetwist.

*2) **Building Bluetooth Communication**:* The website communicates with the Bluetooth module using Web Bluetooth API or an independent backend server.

*3) **Data Transmission**:* The instruction passes as a serial signal to the microcontroller and subsequently becomes real-time response.

D. Microcontroller Processing and Decision Making

After it passes the command to the microcontroller (Arduino UNO), it reads and performs the command. The step is as follows:

*1) **Parsing the Command**:* The incoming string is parsed for identifying the correct operation.

*2) **Run Control Logic**:* There is mapping of every command with a motor action and further to the control algorithm.

*3) **Error Handling and Redundancy**:* The system refuses the input when it finds an illegal command and stops commands later.

E. Robot Action Execution

The microcontroller provides control signals to a motor driver (HW095) to provide motion control. The motor driver regulates direction and motor speed so that the robot can:

*1) **Move Forward:*** Both the motors rotate in the same direction.

*2) **Move Backward:*** Both the motors rotate in the backwards direction.

*3) **Turn Left:*** One motor is slowed down and the other continues to rotate.

*4) **Turn Right:*** The other motor is halted and the other continues.

*5) **Stop:*** The two motors stop working.

IV. Results

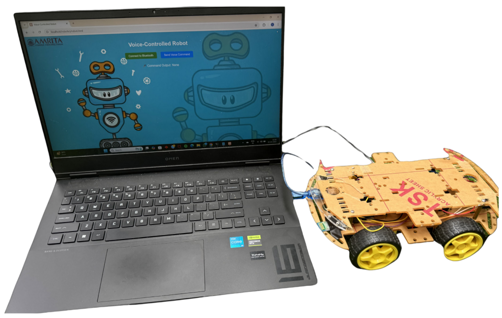

Fig. 3: Model Architecture Diagram

This speech-driven robot car is able to hear- voice instructions, textifying them and carrying out operations real-time in tandem. Hardware consists of a motorized four-wheel car with help from motors via a microcontroller and an inter- preter command device to textify users' command. On issue of voice instructions such as "move ahead," "backward," "right turn," "left turn," and "stop," there was smooth operation by means of translation in terms of motors moving accordingly by the system and, therefore, ensured that there was proper movement by the robot. It was subjected to different conditions so that it would test its responsiveness, accuracy, and movement under different conditions. With the integration of a Bluetooth module, wireless communication between the robot and the control system was obtained with added mobility. Power efficiency was achieved with the addition of a rechargeable battery for prolonged use. The device functioned as expected in word instruction, justified its application in smart automation and assist robots.

V. Discussions and Conclusion

Use of a voice-controlled robot system is a major milestone towards automation of human tasks in the absence of their physical presence, with immediate access and instantaneous response. Effective utilization of speech recognition, decision-making through a microcontroller, and Bluetooth communication offers easy interaction between humans and the robot system. With rigorous testing, the robot executed commands like "move forward," "move backward," "turn left," "turn right," and "stop" with satisfactory accuracy and fluidness of movement. There were some issues, however, such as speech recognition failure when there was background noise and accent variation among speakers, limited range of communication via Bluetooth, and slow processing in executing long or incorrect commands. Despite all these issues, the system can replace old manual and joystick-controlled robots with a superior and more efficient option. Unlike conventional control systems, the system in question is simpler for the user to manage, especially in home automation, assistive technology, and industrial automation.

Having a voice command input interface to convey the input over the internet does serve a purpose in this regard in the environment where remote control is facilitated. Bluetooth offers secure short-range communication, but at a scale larger than to Wi-Fi or IoT technology would be ideal for working distance and roaming. Addition of sensors to detect obstacles and adaptive speech algorithms with coupling would provide additional functionality to the robot, and it would be intelligent and autonomous. The current work testifies to the viability of speech-controlled robots in making things economic and efficient enough to automate them. Machine learning speech recognition technology will develop over the next several years and advanced connectivity features will more and more define human-robot interaction as robots become more capable and intuitive to interact with in various industries.

VI. Future Scope

The future direction of this work is to increase voice-controlled robots with enhanced artifi- cial intelligence, machine learning, and IoT feature. The future direction of one of the most advanced fields is speech recognition rate based on deep learning architectures like Recur- rent Neural Networks (RNNs) and Transformer-based models with focus on eliminating disparity based on accent, background noise, and context sensi- tivity. Scaling to 5G or GSM or Wi-Fi connectivity will make it larger operating range, allowing remote operation over long distances for industrial automation and assistive applications. Real-time obstacle avoidance and autonomous navigation through the use of LiDAR sensors and computer vision can also make the system more versatile and intelligent, even further making the system suitable in ever-changing environments such as warehouses, smart homes, and hospitals. Multimodal interaction based on voice commands and face or gesture recognition can give more human-like robot-human models of interaction. Also, with cloud computing and edge-based AI utilized for real-time processing and decision support, the system will be even more efficient and responsive. Since robotics is constantly being developed using AI, as is the case here, the described system can be engineered to be utilized for healthcare, defense, surveillance, disaster recovery, and personal care, for the advantage of the autonomous robotics and human-machine collaboration industry overall.

References

[1] Teeda, V., Sujatha, K. and Mutukuru, R., *Robot voice: A voice controlled robot using Arduino*. arXiv preprint arXiv:2402.03803, 2024.

[2] Luo, C., *[Retracted] A Voice Recognition Sensor and Voice Control System in an Intelligent Toy Robot System*. Journal of Sensors, 2023(1), p.4311745, 2023.

[3] Basyal, L., *Voice recognition robot with real-time surveillance and automation*. arXiv preprint arXiv:2312.04072, 2023.

[4] Pavani, V., Rohit, P., Kumari, T.J. and Leela, N.S., *Voice controlled personal assistant robot for smart appliances by using Raspberry-Pi*, 2024.

[5] Grágeda, N., Busso, C., Alvarado, E., García, R., Mahu, R., Huenupan, F. and Yoma, N.B., *Speech emotion recognition in real static and dynamic human-robot interaction scenarios*. Computer Speech and Language, 89, p.101666, 2025.

[6] Hussein, H.H., Karan, O. and Kurnaz, S., *Enhancing Driving Control via Speech Recognition Utilizing Influential Parameters in Deep Learning Techniques*. Electronics, 14(3), p.496, 2025.

[7] Teeda, V., Sujatha, K. and Mutukuru, R., *Robot voice: A voice controlled robot using Arduino*. arXiv preprint arXiv:2402.03803, 2024.

[8] Schewior, J., Grefen, R., Verde, R., Ergardt, A., Zhao, Y. and Kullmann, W.H., *Speech-Controlled Robot Enabling Cognitive Training and Stimulation in Dementia Prevention for Severely Disabled People*. In Current Directions in Biomedical Engineering (Vol. 10, No. 4, pp. 543-546). De Gruyter, 2024.

[9] Aguilera, C.A., Castro, A., Aguilera, C. and Raducanu, B., *Voice-Controlled Robotics in Early Education: Implementing and Validating Child-Directed Interactions Using a Collaborative Robot and Artificial Intelligence*. Applied Sciences, 14(6), p.2408, 2024.

[10] Basyal, L., *Voice recognition robot with real-time surveillance and automation*. arXiv preprint arXiv:2312.04072, 2023.

[11] Mawaddah, N.I., Tamba, C.R.A., Hutagalung, D.P., Nainggolan, F.D., Lubis, E.M. and Jamzuri, E.R., *Automatic speech recognition for human-robot interaction on the humanoid robot Barelang 7*. In ICAE 2023: Proceedings of the 6th International Conference on Applied Engineering, ICAE 2023, 7 November 2023, Batam, Riau Islands, Indonesia (p. 15). European Alliance for Innovation, 2024.

[12] Peña-Cáceres, O., Silva-Marchan, H., Albert, M. and Gil, M., *Recognition of Human Actions through Speech or Voice Using Machine Learning Techniques*. Computers, Materials Continua, 77(2), pp.1873-1891, 2023.

[13] Srivastava, S. and Singh, R., *Voice controlled robot car using Arduino*. International Research Journal of Engineering and Technology (IRJET), 7(05), pp.4033-4037, 2020.

[14] Chaudhry, A., Batra, M., Gupta, P., Lamba, S. and Gupta, S., *Arduino based voice controlled robot*. In 2019 International Conference on Computing, Communication, and Intelligent Systems (ICCCIS) (pp. 415-417). IEEE, 2019.

[15] Manwani, A., Dhirani, J., Baloch, A. and Wasan, R.H., *Design of voice-controlled robot vehicle*. Journal of Applied Engineering Technology (JAET), 6(1), pp.40-47, 2022.

[16] Choutri, K., Lagha, M., Meshoul, S., Batouche, M., Kacel, Y. and Mebarkia, N., *A multi-lingual speech recognition-based framework to human-drone interaction*. Electronics, 11(12), p.1829, 2022.

[17] Iliev, Y. and Ilieva, G., *A framework for smart home system with voice control using NLP methods*. Electronics, 12(1), p.116, 2022.

[18] Kuriakose, S., M M, H., Keerthana, D.G., Adarsh, S. and H. K., *Wireless Voice Controlled Robot*. In 2023 9th International Conference on Advanced Computing and Communication Systems (ICACCS), Coimbatore, India, pp. 189-194, 2023. doi: 10.1109/ICACCS57279.2023.10113022.