International Journal of Engineering Sciences 2024 16(1) 8 - 14

INTERNATIONAL JOURNAL
OF
ENGINEERING SCIENCES

Speech-Driven Control for Collaborative Robots to Enhance Human Robot Collaboration in Industrial Environment

Pratiksha P. Jawale and Dr. Shantipal S. Ohol
COEP Technological University, Pune, India

(Received 22 February 2024; Accepted 2 August 2024)

https://doi.org/10.36224/ijes.160102

Abstract
Speech recognition technology has emerged as a pivotal tool across various industries, enabling seamless and natural interactions between humans and robots. In the field of Human-Robot Interaction (HRI), it offers an intuitive and efficient means of controlling robots with straightforward implementation and evident usability, making it a valuable asset for improving communication in collaborative settings. This research focuses on developing a real-time, speech-driven external control system for collaborative robots (cobots) by integrating Python with the Google Speech-to-Text API, achieving precise control of robot movements along three axial directions and gripper operation. The study evaluates key performance metrics,including latency, speech recognition accuracy and user experience for highlighting the system's potential for enhancing human-robot collaboration

Keywords: Speech Recognition, Collaborative Robot, Human-Robot Collaboration

1. Introduction

Collaborative robots, commonly known as cobots, are industrial robots designed to work safely alongside humans in shared environments, revolutionizing industrial automation and reshaping traditional manufacturing methods [1]. Unlike conventional industrial robots, cobots are engineered for close, secure collaboration with people, ushering in a new phase of human-robot collaboration [9]. Their rapid rise in industries, particularly manufacturing, stems from their versatility in performing various tasks, improving worker comfort, and enhancing productivity [1]. However, this progress brings challenges, especially the need for simpler, more intuitive ways to control them. Traditional programming approaches often demand specialized skills, making it tough for a wider ange of users to adapt to cobot systems [8]. To overcome this, experts are developing easier communication methods between humans and robots, with voice control emerging as a promising solution [3]. Speech, being a natural way for people to interact, can make managing cobots more straightforward and hands- free, allowing workers to multitask more effectively [8]. Advances in speech recognition and natural language processing further support this, potentially reducing the learning curve and making cobots accessible to non-programmers and diverse users [2]. Still, adopting voice control in industrial settings isn't without hurdles. Issues like handling background noise, accommodating multiple languages, and

*Corresponding author
Email address: pratikshaphddocuments@gmail.com (Pratiksha P. Jawale)
ISSN 0976 – 6693.

meeting strict safety standards must be addressed [7]. Adding speech control to existing cobot setups requires thoughtful planning and innovative approaches to ensure it's both reliable and safe [9]. Building upon previous research, this paper further assesses the usability of a speech recognition system for the control of collaborative robots in an industrial environment [11]

2 Related Work

2.1 Process of Speech Recognition

Automatic speech recognition (ASR), often called speech or voice recognition technology, falls within the interdisciplinary realm of computational linguistics. This area concentrates on devising techniques and systems that empower computers to detect and transcribe spoken words into text form [6]. Today's ASR solutions predominantly harness machine learning strategies, emphasizing advanced deep learning frameworks such as recurrent

neural networks (RNNs) and transformers to enhance performance and precision [2]. These advancements have significantly improved the accuracy and applicability of speech recognition systems in various domains, including robotics [10]

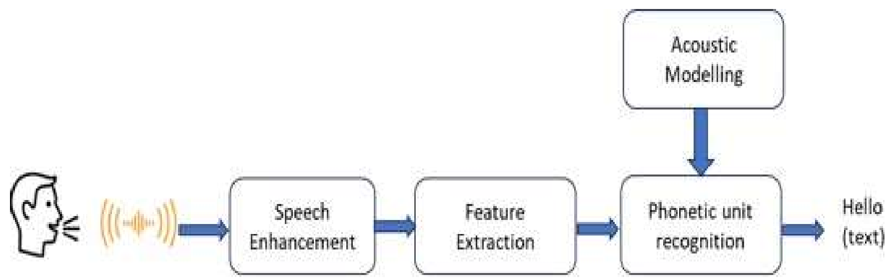

Figure 1: Speech Recognition Process

2.2 Literature Survey

Speech-based control systems provide numerous benefits in human-robot collaboration, particularly with cobots. They allow for hands-free operation, enabling workers to manage robots while simultaneously handling other tasks [3]. As a natural mode of human communication, speech can accelerate the learning process for newcomers and improve accessibility for individuals with physical impairments that limit their use of standard interfaces [8]. In certain cases, issuing voice commands proves quicker and more efficient than interacting with graphical displays [10]. Additionally, speech interfaces can process a diverse array of instructions, offering greater adaptability than rigid button-based controls. By delivering an intuitive and straightforward interaction method, these systems can lower the entry barrier for non-technical users, potentially enhancing both productivity and satisfaction in collaborative robotic environments [4]. However, deploying speech control in industrial contexts introduces several obstacles. Ambient noise from machinery and other sources can disrupt voice recognition accuracy [7]. Global manufacturing operations necessitate systems capable of interpreting commands in various languages, a task complicated by differences in accents, dialects, and specialized technical terms [5]. Privacy and security are also critical issues, as speech systems must protect against data breaches, unauthorized access, and vulnerabilities to voice-based exploits [7]. Furthermore, integrating voice control with outdated legacy systems or a variety of robotic platforms in industrial settings can pose significant technical challenges [9].

3 Experimentation

3.1 System Architecture

The system leverages a Mitsubishi RV-5AS-D, a 6-degree-of-freedom (DOF) cobots with a specified payload capacity and reach, designed for safe human-robot collaboration. It is paired with a computer running Windows 10, powered by an Intel i7 processor and 16GB of RAM, ensuring robust processing capabilities. Audio input is captured via a KinectV2 Microphone array optimized for noise reduction, recorded at a 16 kHz sampling rate in WAV format. The software stack includes Python, C++, and the Google-cloud-speech library for seamless integration with the Google Speech-to-Text API, configured with an English language model. The system architecture begins with audio capture, followed by transcription through the API, parsing of text into actionable commands like "forward," "backward," "left," "right," "upward," "downward," and "stop" (mapped to X, Y, Z axis movements), and finally, execution of corresponding robot actions, as illustrated in an architecture diagram (Figure 1).

Figure 2: System Architecture

3.2. Methods and system

The system begins by capturing the user's voice with an array of microphone, designed to isolate the speaker's voice and suppress background noise, making it ideal for loud industrial settings. A Python script then processes this audio using the Google-cloud- speech library to access the Google Speech-to-Text API, set to an English language model, which transcribes spoken phrases like " forward" into text. This text is parsed by a Python middleware that identifies specific commands such as "forward," "backward," "left," "right," "upward," "downward," mapped to the robot's X, Y, and Z axes discarding any unrelated speech and "Open" and "Close" for operating the gripper. The validated command is sent to a C++ control module through efficient inter-process communication methods like sockets or shared memory, chosen for their speed in real-time systems.

Once the C++ module receives the command, it converts it into precise movement instructions for the Mitsubishi RV-5AS-D collaborative robot (Cobot) using kinematic algorithms, such as inverse kinematics, to determine the joint angles needed for accurate motion. These instructions are transmitted to the cobot via a TCP/IP connection, enabling near-instantaneous response to voice commands. The combination of Python for speech processing and C++ for robot control optimizes the system: Python manages high-level tasks, while C++ handles performance-critical operations. This creates an intuitive, voice- driven interface that simplifies robot operation, eliminates the need for programming expertise, and enhances productivity by freeing users to multitask, making the cobot more accessible and valuable in industrial applications like manufacturing expertise, and enhances productivity by freeing users to multitask, making the cobot more accessible and valuable in industrial applications like manufacturing.

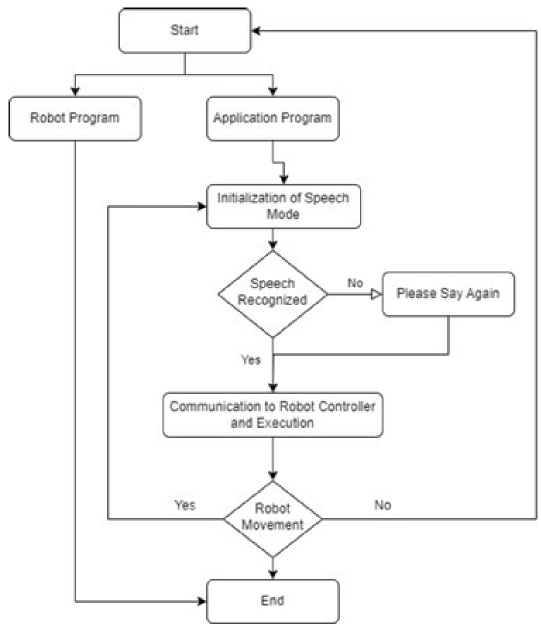

Figure 3: Flowchart of System

3.3 Result and Discussion

Certain verbal commands were mapped to matching robot motions along the X, Y, and Z axes to provide intuitive operation of the collaborative robot. These commands facilitate smooth and user-friendly interaction by enabling the operator to control the robot's motion merely by speaking.

Table 1: Speech Command and action of Robot

Command	Direction
Forward/Backward	Robot Moves in +X/-X Direction
Move Left/Move Right	Robot moves in +Y/-Y Direction
Upward/Downward	Robot Moves in +Z/-Z Direction
Open/Close	To open and close the Gripper
Stop	Robot Stop the movement

Table 2: Latency Recorded

Command	Latency Received(s) (trials)										
	1	2	3	4	5	6	7	8	9	10	Avg.
Forward	1.37	1.39	1.25	1.49	1.39	1.63	1.42	1.04	1.58	1.39	1.395
Backward	1.4	1.35	1.39	1.32	1.35	1.36	1.55	1.54	1.25	1.08	1.359
left	1.5	1.46	1.36	1.25	1.33	1.64	1.35	1.3	1.33	1.25	1.377
right	1.45	1.35	1.36	1.39	1.24	1.34	1.6	1.25	1.54	1.3	1.382

The performance of the deployed speech recognition system for robot control was evaluated through three critical metrics: accuracy, latency, and user experience. The system exhibited high reliability, with minimal errors in command interpretation and execution across diverse testing conditions. Latency defined as the time from speech input to the initiation of robot movement was measured in an experiment where ten readings were recorded for each directional operation, yielding an average of 1.378 seconds. This prompt response time supports a natural flow in human-robot interaction (HRI), vital for precision tasks and real-time feedback, though slight millisecond variations were noted due to human error in stopwatch measurements. User feedback further affirmed the system's success, praising the intuitive and natural control afforded by speech recognition, which significantly enhanced the robot-user interface. These results highlight the system's dependability and its potential for integration with other control modalities, such as gesture- based commands.

The speech recognition accuracy for the speech-controlled robot was assessed using a set of commands: 'forward,' backward, Upward, downward, 'left,' 'right'. Each command was tested five times by speaking into a microphone, and the system's output was compared to the expected command. Accuracy was determined as the percentage of correctly recognized commands out of the total attempts, offering a measure of the system's performance for robot control. Accuracy to be found in this case is 93.33%. Command recognition matrix is shown in figure 4.

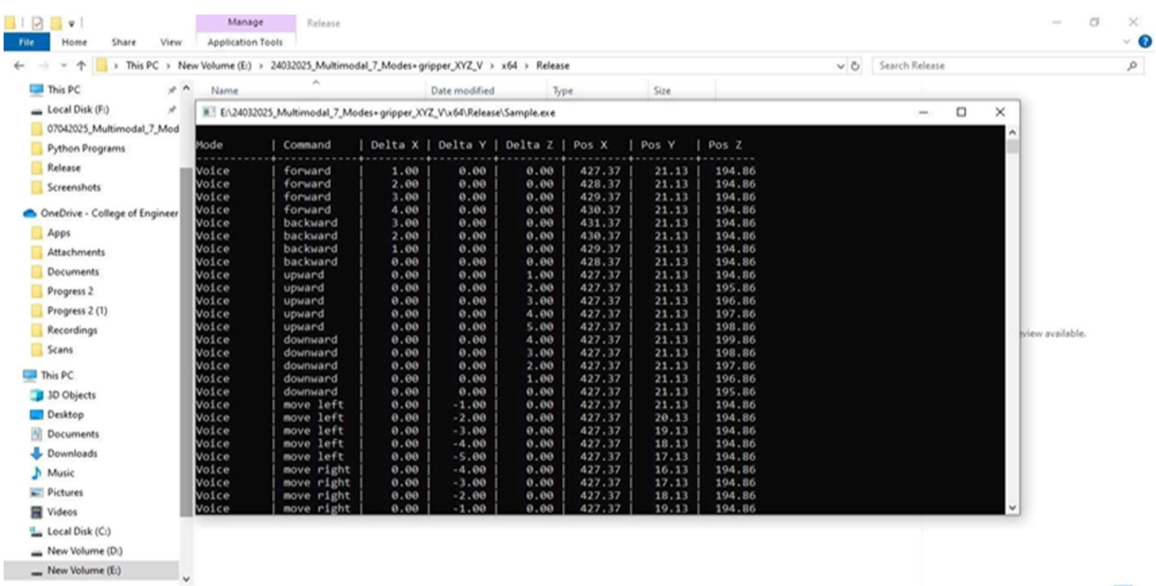

Figure 4: Command Recognition and Robot movement

The user experience for the speech-controlled robot was assessed through a blend of task performance and subjective feedback. Participants were tasked with completing a series of navigation exercises using simple voice commands as listed in table 1, The evaluation focused on the system's effectiveness and ease of use, with metrics like task success and response times considered. Observations and user input emphasized the robot's consistent command recognition as a key strength, making it straightforward and intuitive to operate. While the system demonstrated satisfactory performance, some users noted slight response delays as a potential area for refinement. Overall, the findings indicate that the speech- controlled robot provides a user-friendly and reliable experience, with opportunities to further optimize its responsiveness for an even smoother interaction. Integrating speech recognition into the robot control system presented challenges related to speech input accuracy and system responsiveness, both of which were effectively addressed. Background noise, a common issue in speech recognition, was mitigated using a directional microphone, while misinterpretations from similar-sounding phrases or ambiguous instructions were resolved by implementing advanced parsing logic and a context-aware command parsing system within the Python middleware. To ensure swift and precise command execution, communication between Python and C++ components was optimized using efficient Inter Process Communication methods. Additionally, refinements to the C++ control algorithms enhanced overall system performance, boosting the precision and responsiveness of speech-derived commands. Together, these solutions not only overcame the integration hurdles but also elevated the system's reliability and the seamlessness of the human-robot interaction experience.

4. Conclusion

In this study, we implemented and evaluated a speech-controlled system for collaborative robots leveraging the Google Speech-to-Text API, demonstrating its potential as an intuitive interface for human-robot interaction. The system achieved a speech recognition accuracy of 93.33% single-word commands such as 'forward,' 'backward' 'upward' and 'downward,' indicating reliable performance within a controlled environment. User experience accounted with participants appreciating the simplicity of voice commands, though minor response delays were noted. These results affirm the feasibility of using off- the-shelf speech recognition for robotic control, particularly in collaborative settings. Future work could focus on enhancing real-time responsiveness, adapting to diverse accents and noisy conditions, and expanding the command set to further improve functionality and user satisfaction.

References

1. Mohd J., Abid H., Ravi S., Shanay R., Rajiv S.: Significant applications of Cobots in the field of manufacturing. Cognitive Robotics. Volume 2, Pages 222-233, ISSN 2667- 2413(2022).
2. Zhang, Y., Pang, J., & Jiang, B.: Recent advances in speech recognition using transformer- based models. *IEEE/ACM Transactions on Audio, Speech, and Language Processing* 29, 2809–2824 (2021).
3. Mavridis, N.: A review of verbal and non-verbal human–robot interactive communication. Robotics and Autonomous Systems 63, 22–35 (2015).
4. P. P. Jawale and S. S. Ohol, Improvisation in human–robot interaction using optimized multimodal operational techniques, in Innovations in Electrical and Electronic Engineering: Proceedings of ICEEE 2023, R. N. Shaw, P. Siano, S. Makhilef, A. Ghosh, and S. L. Shimi, Eds. Singapore: Springer, 2024, vol. 1109, pp. 403-413.
5. Jani, M., Panchal, S., Patel, Raiyani, A.: Multilingual Speech Recognition: An In-Depth Review of Applications, Challenges, and Future Directions. (2024).
6. Juang, B. & Rabiner, Lawrence. Automatic Speech Recognition - A Brief History of the Technology Development. (2005).

7. Li, J., Chen, C., Pan, L., Rahimi Azghadi, M., Ghodosi, H., Zhang: Security and Privacy Problems in Voice Assistant Applications: A Survey. (2023).
8. Khayrallah, H., Trott, S., & Feldman, J.: Natural language for human robot interaction. International Journal of Human-Computer Interaction 34(4), 295–305 (2015).
9. S. Robla-Gómez, V. M. Becerra, J. R. Llata, E. González-Sarabia, C. Torre Ferrero and J. Pérez-Oria, Working Together: A Review on Safe Human-Robot Collaboration in Industrial Environments, in IEEE Access, vol. 5, pp. 26754 26773. (2017)
10. Tellex, S., Gopalan, N., Kress-Gazit, H., & Matuszek, C.: Robots that use language. Annual Review of Control, Robotics, and Autonomous Systems 3, 25–55 (2020).
11. P. P. Jawale and S. S. Ohol, Improvisation in human–robot interaction for real time speech operation using speech for collaborative Robot, 9rh International Conference on Intelligent Technologies, pp 217-223. (202