Speech Recognition based Industrial Cloud Robot for Service-Oriented Sustainable Manufacturing

Kalaiarassan G1, Franklin Prashanth C1, Prakash M1 and Shreya Phirke1

1Vellore Institute of Technology, Katpadi, Vellore, Tamil Nadu, India

E-mail: kalaiarassan.g@vit.ac.in

Abstract. Industrial Cloud Robotics is an amalgamation of cloud computing and an industrial robot to establish a service-oriented manufacturing system. Teleoperation of an industrial robot – the ability to control a robot from a remote location, is facilitated through this system. In this paper, a framework is proposed by utilizing speech recognition and industrial cloud robots for the applications of sustainable manufacturing. We have used Google’s speech recognition services to control the robot manipulator. We have employed Android Speech API in a custom android application that receives the speech signal, transcribes, and forwards it to a server. The application can be viewed or accessed by any host computer, which primarily serves as a user. The monitoring unit and the data is fetched by the Robot Speech Interface unit via the internet and web sockets. The interface triggers the required action of the robot through the relay board actuation of the digital input facility of the robot. Through this work, it is realized that, despite the disturbances and noise interferences, speed and reliability are not compromised.

1. Introduction

ISO 8372:2012 defines Industrial Robots as “an automatically controlled, reprogrammable, multipurpose manipulator, programmable in three or more axes, which can be either fixed in place or mobile for use in industrial automation applications” [13]. An archetypal robot manipulator is generally mounted on a fixed base and is connected to links that move in various directions about its axes, allowing the robot to work. The end-effector can be selected and customized based on the application of the robot. Ever since the discovery of the first industrial robot in 1954, industrial robots have found use in applications such as welding, painting, assembling, palletizing, pick and place, product testing, and more.

The demand and use of robots have dramatically burgeoned in the last few decades. Robots are being used for various applications that range from medical to manufacturing, to improve productivity and product quality. Due to the surge in demand, Industrial Robots are designed to make them user friendly in terms of the control features, decentralized accessibility aspects (remote access), and flexibility of operations. The rapid shift from manual labor to automation has given rise to a new category of robots, the Intelligent Robots. Intelligent Robots perform several complex robotic tasks that involve perceiving and reasoning the environmental factors by making use of sensor technology and control algorithms. Human-Machine Interaction (HMI) is a major feature that promotes effortless and safe interaction experience with robots. Machine Learning, HMI, and Cyber-Physical Systems are increasingly incorporated in industrial robots as they facilitate ease of functionality, flexibility, accessibility, and other control features.
In this paper, the domain of teleoperation is explored for convenient and sustainable manufacturing. Teleoperation is the ability to control a robot from a remote location. It is useful in addressing the issues related to inspection, programming, and servicing of the robot. The remote-control signals are sent through a wire via a local wireless system, over the internet. Teleoperated robots currently are extensively used in the field of medical treatment and surgeries, where the instructions given by the doctor are interpreted by the master and slave robots who treat the patients. Furthermore, these teleoperated robots are also used in industries that possess a threat to human health such as highly toxic areas and nuclear reactors.

Attributing to the convenience of operation that it offers, there have been expeditious advancements in the field of speech recognition and transcription by major Artificial Intelligence researchers, developers, and providers. Natural Language Recognitions of over 80 languages including its most common dialects is now possible. This inspires a new field of research and provides a broader scope for controlling various machines by implementing speech recognition technology. Moreover, the use of voice as input to control various devices and processes has been extensively developed as an effective tool for the human-machine interface (HMI). Siri, Cortana, Google Assistant are the pioneers of the field and have reached the very hands of the masses. Especially, Google’s ‘Google Home’ and Amazon’s ‘Echo’ have become part of many households. These systems offer AI integrated Natural Language Recognition features that understand the commands and respond using intelligent reasoning and algorithms. This paper attempts an integration of two features namely, IoT system and voice or speech control in the domain of robotics.

2. Overview of related works
L. Z. L. Zhizeng et.al [1] in 2004 successfully developed a speech recognition system to be applied for robots in their research. But the speech recognition was limited to only 50 words maximum. Commands which may be considered basic now, like the ‘left’, ‘right’, ‘front’ and ‘back’ were incorporated in the robot system and tested for its efficiency and reliability of the system. They implemented the DTW method of speech recognition. The time varying digital filter represented the shape of vocal cords and the Gain factor represented intensity of the voice. Endpoint-detect, signal-strengthen, windowing, framing of the signal were the signal pre-processing steps executed before feature extraction and DTW. The reciprocal spectrum coefficient of linear prediction co-efficient was used for efficient in the recognition of isolated word.

B. M. Chinnathurai et.al [10] have worked on building a robot for the sorting work, in a waste segregation process, in order to aid and amplify the process of recycling. The robot system had a vision system which was programmed using a Raspberry pi on Linux OS for object detection. Once the object was detected the robot collected the required objects.

S. N. A. Mohamad et.al [8] developed word semantic models that signified the relation between semantic databases that represented the meaning of a word with the registered word from the speech recognizer, using MATLAB. The noise signals were filtered using the MFCC algorithm by extracting the sequence of features for every short-time frame of the input signal.

3. Problem Definition
Increased manufacturing requirements and technological advancements have led to the expansion of the Industrial Robots industry. It is estimated that the market size Industrial Robot industry will be around 24500 Million USD by the end of 2025, with a CAGR of 10.2% between 2019 to 2025. Industrial robots are being deployed in automotive as well as several non-automotive sectors. While installing a robot may increase productivity by reducing the cycle time, it manifests several financial and logistical challenges like a large upfront investment, employee skillset training, implementation of safety measures, among others. Furthermore, the scant number of expert robot programmers or service engineers in this field can lead to delayed response time against service requests, high service charges, increased downtime, and accumulation of work that affects the quality of the product.
Moreover, every industrial robot is copyrighted by their respective manufacturers and provide little or no scope for customization. Additional financial investments have to be made to customize the robot as per the requirement. Different types of equipment have to be purchased from the Original Equipment Manufacturer (OEM) for every customization required. The functionality of the robot is restricted due to factors such as controller limitation: manufacturers have dedicated controllers for each model, and they are non-interchangeable. The compatibility of various features is pre-fixed within the controller. For instance, if an industrial robot does not facilitate speech recognition if the controller has no provisions for that feature. Additionally, if speech-recognition is required, a new robot has to be purchased.

Existing HMI techniques like touch and gesture require higher human effort. All other HMIs require an embedded system, usually wired to the robot or sometimes are wireless till a small range. However, provisions for distant remote control and command of the robot are not available.

4. Methodology
The proposed methodology for the project is given below:

5. The System
5.1. System Layout
The basic system layout is as given in the below figure 2. Here, the major three components are the Input Unit, which in our project is an android mobile phone, the intermediate cloud and the robot system. The robot in return is continuously monitored by an IP camera system which again transfers the data to the cloud.
Figure 2. The System Layout

5.2. System Framework

The remote-control unit is the first subsystem of the framework, it is accessed and used by the programmer or operator of the robot service company. The primary element of the system is internet-enabled applications compatible with devices like mobile phones, computers, or any PDAs. For this particular project, we focus on the mobile device as our preferred mode of the user interface for the system. An android application is developed for the same, which is powered by the Android Speech recognition API. This application acts as our core engine for speech recognition since we focus on speech as our desired mode of input with other methods like text and UI commands available.

The cloud is a remote server that receives the data emitted from the mobile phone and processes it for keyword detection, which specifies the nature of the job. (like pick, take, write, move, etc). The cloud is programmed to identify the keywords and separate them from the objects in the command. The cloud passes the command or data to the robot via web sockets. The system is monitored by a camera that uploads the data to the cloud. This data viewed by the operator to verify the robot's performance according to his commands.

Figure 3. Proposed System Framework
6. The Interface Module
The interface module is constructed using NodeMCU as the base brain. The microcontroller is programmed to receive the data signals from the server. A relay board is actuated by the triggering signals from the microcontroller. Relay is used as the Robot takes input as only 24V through its digital input ports. Hence, the MCU triggers the relays to give the input to the robot. The interface setup is as shown in Figure. 4. This setup is fixed on the robot's peripheral mount head on J3.

![Interface Module](image)

**Figure 4.** The interface module

7. Experimentation and Result
The system was subjected to various tests below to check the performance and the reliability against different variables.

7.1. Environmental Conditions
The working environment is the first test, where the system was subjected to ideal quiet and noisy conditions. 10 samples were taken in a trial for each condition and the observation in as given in Table 1.

| Conditions                  | Trial 1 | Trial 2 |
|-----------------------------|---------|---------|
|                            | Number of correct detections | %Correct detection | Number of correct detections | %Correct detection |
| Normal inbuilt phone mic    | 18      | 90      | 19      | 95      |
| Ideal condition             |         |         |         |         |
| With noise disturbances     | 15      | 75      | 16      | 80      |
| Dedicated mic (headsets)    | 19      | 95      | 20      | 100     |
| Ideal condition             |         |         |         |         |
| With noise disturbances     | 17      | 85      | 18      | 90      |

Table 1. Reliability for speech recognition at ideal and noisy conditions
7.2. Internet Band Width

The maximum data transfer rate of a connection or network is described by Bandwidth. It measures the amount of data that can be sent over a particular network in a given time period.

| Internet upload speed | Trial 1 Time in Sec | Trial 2 Time in Sec |
|-----------------------|---------------------|---------------------|
| 3.5 MBPS (Jio 4G net) | 3                   | 4                   |
| 6.8 MBPS (Idea 4G net)| 3                   | 3                   |

As the data transfer required is very small (the data is only audio and text), the changes in internet bandwidth do not affect the performance of the system, unless the bandwidth falls very low. When the upload speed is less than 0.8 MBPS, the speech recognition is slow.

8. Conclusion and Future Scope

The proposed framework using speech as the input has shown desirable performance under the various experimental conditions and thus is reliable. Resources spent on service engineers can be minimized by implementing this system. This system further allows easy energy optimization, resource sharing, and convenience of manufacturing by enabling remote robot access. It can save time, monetary resources, labour costs, and provide the client with the best possible service. Hence, this tele-commanding of the robot is a better alternative to the conventional system of programming debugging, inspection, and service.

The system is to be developed as a cyber-physical system, where the robot controller communicates with the cloud server directly. The aim is to eliminate the need to program the local robot. Expert programmers can program energy and memory-efficient algorithms capable of performing a predefined set of applications. The program is stored on the cloud server and the task of the controller is to select the program required and download it from the cloud.

The integration of Artificial Intelligence for the vision of the system, using ImageNet, will enable image recognition of various objects that are trained. Almost all real-time objects can be detected, and thus be acted upon.

References
[1] L. Z. L. Zhizeng and Z. J. Z. Jingbing 2004 Speech recognition and its application in voice based robot control system 2004 Int. Conf. Intell. Mechatronics Autom. 2004. Proceedings pp 960-963
[2] Q Hong, C Zhang, X Chen and Y Chen 2007 Embedded speech recognition system for
intelligent robot Proc. 14th Int. Conf. Mechatronics Mach. Vis. Pract. M2VIP2007 pp 35-38

[3] P Mayorga O, J M Olguín E, A M Hernández C and J I Flores M 2007 Gaussian components optimization for a robot controlled by speech commands in Mexican Spanish Electron. Robot. Automot. Mech. Conf. CERMA 2007 - Proc. pp 324–329

[4] Detlef RICHTER 1998 Robot Control by Digital Binocular Computer Vision and Digital Speech Recognition Optimization of Electrical and Electronic Equipments – Bravov. pp 617-622

[5] Rihab Chaâri, Fatma Ellouze, Anis Koubâa, Basit Qureshi, Nuno Pereira, Habib Youssef and Eduardo Tovar 2016 Cyber-physical systems clouds: A survey Comput. Networks 108 pp 260-278

[6] I V Blagouchine and E Moreau 2010 Control of a speech robot via an optimum neural-network-based internal model with constraints IEEE Trans. Robot 26 pp 142-159

[7] X Liu, S S Ge, R Jiang and C H Goh 2016 Intelligent speech control system for human-robot interaction Chinese Control Conf. CCC pp 6154-6159

[8] S N A Mohamad, A A Jamaludin and K Isa 2017 Speech semantic recognition: A control mechanism for assistive robot USYS 2016 - 2016 IEEE 6th Int. Conf. Underw. Syst. Technol. Theory Appl. pp 139-144

[9] O Rogalla, M Ehrenmann, R Zollner, R Becher and R Dillmann 2002 Using Gesture and Speech Control for Commanding a Robot Assistant Proceedings at the 2002 IEEE Int. Workshop on Robot and Human Interactive Communication vol. pp 454-459

[10] B M Chinnathurai, R Sivakumar, S Sadagopan and J M Conrad 2016 Design and implementation of a semi-autonomous waste segregation robot Conf. Proc. - IEEE SOUTHEASTCON pp 1-6

[11] R Baheti and H Gill 2011 Cyber-physical Systems Impact Control Technol. pp 161-166

[12] S N A Mohamad, A A Jamaludin and K Isa 2017 Speech semantic recognition: A control mechanism for assistive robot USYS 2016 - 2016 IEEE 6th Int. Conf. Underw. Syst. Technol. Theory Appl. pp 139-144

[13] ISO 2012 Robots and Robotic Devices – Vocabulary (Online: International Organization for Standardization)