International Journal For Advanced Research In Science & Technology
A peer reviewed international journal
www.ijarst.in
ISSN: 2457-0362

Speech Recognition for Secured Robot Operation

Dr.GSK Gayatri Devi[1], Theleru Surender Reddy[2]
Associate professor[1, 2]
Department of ECE
Malla Reddy Engineering College(MREC)

Abstract: Now a day's every system is automated in order to face new challenges. In the present days Automated systems have less manual operations, flexibility, reliability and accurate. Due to this demand every field prefers automated control systems. Especially in the field of electronics automated systems are giving good performance. Here in this project the robot follows a particular path in such a way that this path is always depends on the voice of ours. In this paper "Speech recognition for secured robot operation", like the title indicates that controlling action of Robot is done through the voice. In this project, there are two sections (transmitter & receiver). The instructions such as Go, Back, Left, Right etc. are processed and stored in the PC for authorized person's voice using MATLAB. In Transmitter Section, the instructions are delivered through Microphone in authorized person's voice which is given to the PC, transmitter encodes the signal and transmit voice signal with the help of wire antenna. The voice instruction is processed and is compared with the database. The data resembling the action for the instruction is sent by the transmitter if and only if the voice is proved to be authorized. In Receiver Section, the signals from the transmitter section are received by the RF receiver through the receiving antenna. The Controller will control the Robot direction according to the instruction which is being given at the transmitter section. Then the Robot will move in that particular direction for the given instruction with the help of DC motors. The microcontroller used in our project is ATMEL 89C51 which is an 8-bit microcontroller.

Index Terms- Microcontroller, Robot, DC Motor, RF transmitter, RF Receiver

I. Introduction

Robotic control approaches are extensively studied and developed in recent years, especially in industrial manufacturing. Being intelligent a human find's it very easy to understand and recognize anything said by another person. It does not need any mechanisms or procedures to follow for a machine recognize the speech is very difficult task. The main aim of this project is try to achieve this difficult task through various techniques with a small application of controlling a robot.In this paper, the human speech is inputted to pc through a Mic, which under goes a number of processes to make it understand by the machine. Once the speech is recognized the pc sends signals to a transmitter section which communicates with the robot through RF communication (robot is embedded to receive the signal. The receiver which is a robot receives these signals or better understandable commands upon which the robot makes its move. To achieve the goal speech processing is involved. Speaker recognition is the process of automatically recognizing who is speaking on the basis of individual information included in speech waves. This technique makes it possible to use the speaker's voice to verify their identity and control access to services such as voice dialing, banking by telephone, telephone shopping, database access services, information services, voice mail, security control for confidential information areas, and remote access to computers. Today, robots take an important role in the science and technology development. Indeed, robots have been widely applied in military, industrial and agricultural techniques to replace people's role in dangerous environments. Many technologies are developed for controlling robot in recent years. Among them, Voice – recognition technique is proved as an intelligent method that was increasing widely role in this context such as Google Assistant, Alexa, Cortana [1]. Following, a large number of smart/intelligent robot was designed by using wireless communication. They express the human-robot interaction (HRI) term with more productivity, less energy resource waste and more convenience than old control methods by physical communication (keyboard, computer, LAN, etc.). For example, the researches [2,3] uses Bluetooth to communicate/control to robots from the Bluetooth transmitter to command robot to perform operations. These studies use mainly circuit systems with old sensors such as RSC364 (used in radio transmission) to analyse the voice frequency to command. This approach faces disadvantages such as controlling is possible in low distance depends system, high cost for circuit construction and its quality depends on the recording quality of the microphone. Furthermore, [4,5] propose a new Bluetooth transmitter approach via smartphone. The control design in these works was simpler than the old type. By using WIFI, 3G, 4G, or LAN network, communication with the Robot is significantly improved in distance terms, more flexibility and more convenience [6]. However, they are still limited at storage capacity, direct / indirect connection and poor system security. In our part, we

International Journal For Advanced Research In Science & Technology
A peer reviewed international journal
www.ijarst.in
IJARST
ISSN: 2457-0362

propose an advanced approach to control robot by using IoT communication. Voice recognition command through Google Assistant, transmit/feedback IoT by intermediates servers. Bythis way, this approach has many important ad-vantages that are more secure and safely; controlling robot is not depend on physical distance; easy to access for monitoring by historical data on server. This article is structured as following: the next section proposes the control/monitor robot process. This section presents also the used techniques in this work. Based on this, we show our results that tested on an own developed robotic arm. The conclusion and the perspective will close the article.

II. Literature Survey

To convert voice to text or computer commands, it must be performed by a complex process with many steps. Basically, an Analog-to-Digital Converter (ADC) converts these analog (analog) waves (human voice) into data that system can understand, then filtering (separate) noise into different frequency bands to eliminate, see more details in [7]. Fortunately, all of these complex processes are supported by Google Assistant. Google Assistant system will have to receive sound from the human speaking on the phone, the phone will record the monologue through google, the data will be processed as text characters. The communication module used in this work was the advanced development protocol, which is a virtual interface platform that communicates with the ESP32 chipset: Adafruit IO. This is an IoT platform built around the Message Queue Telemetry Transport (MQTT) Protocol [8], that has extensive literature focused on IoT solutions, all developed by the Adafruit Industries. In other side, If-This-Then-That (IFTTT) is a widely used platform with over 11 million users [9] which is considered as a useful bridge between Google Assistant and Adafruit IO platform. IFTTT platform is designed as a trigger-Action platform which supports stitching together various online services APIs such that end-users may write simple conditional programs. These simple programs often take the form “IF triggering condition, THEN takes a specific action”. We designed the synchronization operation of these two servers. When a voice command is recorded on Google Assistant, it is encoded and transmits this encrypted signal into text form from Google’s API through ITFFF Server to Adafruit IO and sent directly to Arduino ESP 32.

III. Proposed Work

Speech Feature Extraction:

The purpose of this module is to convert the speech waveform to some type of parametric representation (at a considerably lower information rate) for further analysis and processing. This is often referred as the *signal-processing front end*. The speech signal is a slowly timed varying signal (it is called *quasi-stationary*). An example of speech signal is shown in Figure. When examined over a sufficiently short period of time (between 5 and 100 msec), its characteristics are fairly stationary. However, over long periods of time (on the order of 1/5 seconds or more) the signal characteristic change to reflect the different speech sounds being spoken. Therefore, short-time spectral analysis is the most common way to characterize the speech signal.

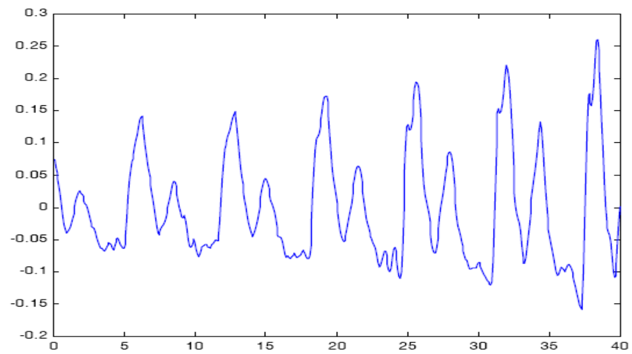


Fig: An example of speech signal

A wide range of possibilities exist for parametrically representing the speech signal for the speaker recognition task, such as Linear Prediction Coding (LPC), Mel-Frequency Cepstrum Coefficients (MFCC), and others. MFCC is perhaps the best known and most popular, and these will be used in this project.

The feature extraction techniques which can be used are explained below:

MFCC’s:

MFCC’s are based on the known and variation of the human ear’s critical bandwidths with frequency, filters spaced linearly at low frequencies logarithmically at high frequencies have been used to capture the phonetically important characteristics of speech. This is expressed in the *mel-frequency* scale, which is a linear frequency spacing below 1000 Hz and a logarithmic spacing above 1000 Hz. The process of computing MFCCs is described in more detail next.

Pattern Recognition:

This direct approach involves manipulating the speech signals directly without explicit feature extraction of the speech signals. There are two stages in this approach, mainly the training of speech patterns and recognition of patterns via pattern comparison. Several identical speech signals are collected and sent to the system via the training procedure. With adequate training, the system is able to characterize the acoustics properties of the pattern. This type of classification is known as the pattern classification. The recognition stage does a direct comparison between the unknown speech signal and the speech signal patterns learned in the training phase. It

International Journal For Advanced Research In Science & Technology
A peer reviewed international journal
www.ijarst.in
ISSN: 2457-0362

generates a "accept" or "reject" decision based on the similarity of the two patterns.

1. It is simple to use and the method is fairly easy to understand
2. It has robustness to different speech vocabularies, users, features sets, pattern comparison algorithms and decision rules.
3. It has been proven that this method generates the most accurate results.

Acoustic-Phonetic Approach:

The acoustic-phonetic approach has been studied in depth for more than 40 years. It is based on the theory of acoustics phonetics that suggest that there exist finite, distinctive phonetic units of spoken language and that the phonetic units are widely characterized by a set of properties that are manifest in the speech signal, or its spectrum, over time. The first step in this approach is to segment the speech signal into discrete time regions where the acoustics properties of the speech signal are represented by one phonetic unit. The next step is to attach one or more phonetic labels to each segmented region according to the acoustic properties. Finally the last step attempts to determine a valid word from the phonetic labels generated from the first step. This is consistent with the constraints of the speech recognition task.

Artificial Intelligence:

This approach is a combination of the acoustic-phonetic approach and the pattern recognition approach. It uses the concept and ideas of these two approaches. Artificial intelligence approach attempts to mechanize speech recognition process according to the way a person applies its intelligence in visualizing and analyzing. In particular among the techniques used within this class of methods are the use of an expert system for segmentation and labeling so that this crucial and most complicated step can be performed with more that just the acoustic information used by pure acoustic-phonetic methods. Neural Networks are often used in this approach to learn the relationship between the phonetic events and all the known inputs. It can also be used to differentiate similar sound classes.

Dynamic Time Warping:

Dynamic Time Warping is one of the pioneer approaches to speech recognition. It first operates by storing a prototypical version of each word in the vocabulary into the database, then compares incoming speech signals with each word and then takes the closest match. But this poses a problem because it is unlikely that the incoming signals will fall into the constant window spacing defined by the host. For example, the password to a verification system is Queensland. When a user utter "Queensland" the simple linear squeezing of this longer password will not match the one in the database,this is due to the longer constant window spacing of the speech "Queensland". Dynamic Time Warping solves this problem by computing a non-linear mapping of one signal onto another by minimizing the distances between the two. Thus Dynamic Time Warping (DTW) is a much more robust distance measure for time series, allowing similar shapes to match even if they are out of phase in the time domain.

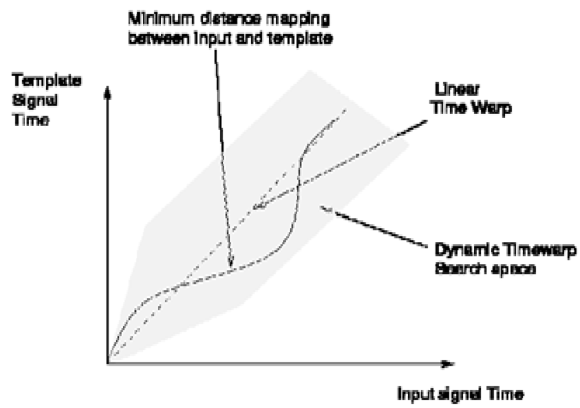


Fig: Dynamic Time Warping

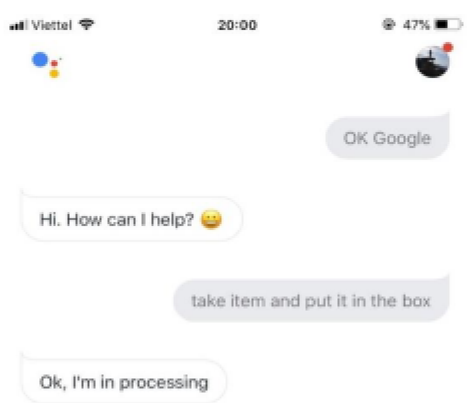
Figure above shows the graph on Dynamic Time Warping, where the horizontal axis represents the time sequence of the input stream, and the vertical axis represents the time sequence of the template stream. The path shown results in the minimum distance between the input and template streams. The shaded in area represents the search space for the input time to template time mapping function.

IV. Results

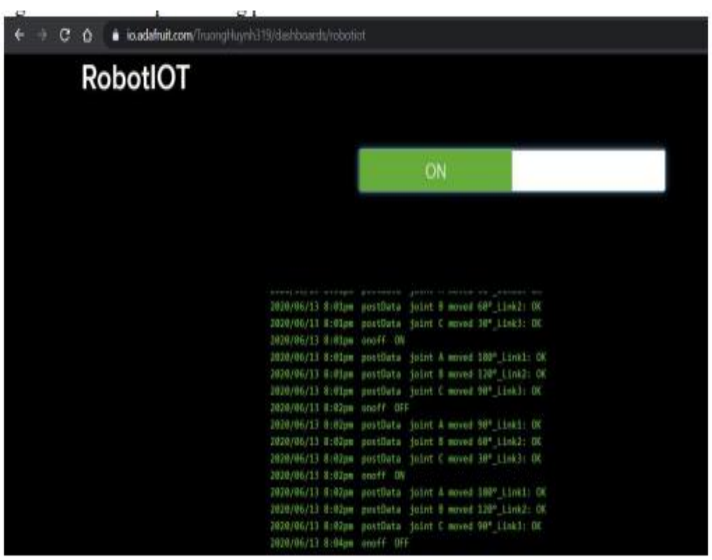
Results and discussions Following to proposed process, following steps describe our implementation and its results: 1) Create an applet editor (on ITFFF platform), where we choose triggers ("If This") and the subsequent actions ("Then That"). We choose in the next "Google Assistant" as the service for our trigger, and then set up its settings such as: the Google Assistant's response, language, etc. In our case, the phrase "Ok I'm in processing" to response command order. It implies that when the ESP processor is received the request and is performing actions. The final part of this applet is the Action, we choose "Adafruit", meaning to "Send data to Adafruit IO". Till this applet is completed, the converted voice-command to text from Google Assistant will be sent to Adafruit IO platform in which transmit to microcontroller. Figure below shows us an example of voice - command by Google Assistant to ESP processor and its response. 2) Create a new feed of Adafruit IO: In which we create threes "postData" variables correspond to threes joints of Robotic Arm and "onoff"

International Journal For Advanced Research In Science & Technology
A peer reviewed international journal
www.ijarst.in
IJARST
ISSN: 2457-0362

variable to represent feedback values from ESP processor while it is performing



In addition, ESP32 processor connect to Adafruit IO by MQTT Protocol. It's performed by to install the Adafruit MQTT Client library, which can be found under the Arduino Library Manager [10] by including at the start of our program: #include #include "Adafruit_MQTT.h" #include "Adafruit_MQTT_Client.h" Once the connection is completed, ESP processor will receive the voice-command from Google Assistant through above step 1 and Robotic Arm perform actions corresponds. Sensor from mini-servo feedbacks to Adafruit Server to monitor Robotic Arm's activities, Figure below prove us controlling and monitoring Robotic Arm's performing process.



In above figure , information feedback from joint A, B, C and onoff correspond to the components of Robotic Arm described in figure 2. They are updated automatically and continuously on Adafruit IO server to monitor in real-time. 3) Finally, these feedbacks are synthesized in our Adafruit server and send a notification on our smartphone when Robotic Arm has finished its activities, as described in figure 5. The results show us that the control and monitoring of the robot's execution are successful. The feedback from the robot's sensors (~5-10ms/cycle) almost no-delay.

I. Conclusion

Our contribution in this article to provide an advanced approach of controlling and monitoring robot's operations, Voice commands are successfully transmitted from Google Assistant on Smartphone through IoT cloud to Arduino processor. User will get notification while desired activities have done successful. This research reduces human activities in dangerous places or situations: such as radioactive mineral exploration, geological measurement. In perspective, an extension monitoring protocol with camera system is necessary to developer on integrating with our methodology. In addition, robot positioning with build self-propelled algorithm make improve supervision system..

References

1. López, Gustavo, Luis Quesada, and Luis A. Guerrero. 2017. Alexa vs. Siri vs. Cortana vs. Google Assistant: a comparison of speech-based natural user interfaces. International Conference on Applied Human Factors and Ergonomics. Springer, Cham.
2. Abdullahi, Zakariyya Hassan, et al. 2015. Mobile robot voice recognition in control movements. International Journal of Computer Science and Electronics Engineering (IJCSEE) 3.1: 11-16.
3. Fezari, Mohamed, and Mounir Bousbia-Salah. 2006. A voice command system for autonomous robots' guidance. 9th IEEE International Workshop on Advanced Motion Control, IEEE.
4. Kahar, Saliyah, et al. 2011. Data transferring technique for mobile robot controller via mobile technology. International Conference on Pattern Analysis and Intelligence Robotics. Vol. 1. IEEE.
5. Ersahin, Gonca, and Herman Sedef. 2015. Wireless mobile robot control with tablet computer. Procedia-Social and Behavioral Sciences 195: 2874-2882.
6. Hassan, Mohammad Ashri Abu. 2012. A review of wireless technology usage for mobile robot controller. Proceeding of the International Conference on System Engineering and Modelling (ICSEM).
7. Andics, Attila, et al. 2010. Neural mechanisms for voice recognition. Neuroimage 52.4: 1528- 1540.
8. Kodali, R. K., and Kopulwar S. M. 2016. A low-cost implementation of MQTT using ESP8266. 2nd International Conference on Contemporary Computing and Informatics (IC3I). IEEE.
9. IFTTT, 2011. https://platform.ifttt.com.
10. Maguluri, Lakshmana Phaneendra, et al. 2018. Efficient Smart Emergency Response System for Fire Hazards using IoT. International Journal of Advanced Computer Science and Applications (IJACSA) 9.1: 314-320.