Voice controlled robotic arm

by Sahana G

Submission date: 05-Dec-2022 03:03PM (UTC+0530)
Submission ID: 1971889155
File name: Voice_controlled_robotic_arm.pdf (1.04M)
Word count: 6986
Character count: 36248

ABSTRACT

In this developing world where everything is automated , robotics play an important role . Despite the goal of making lives easier for everyone, some of the robots make everyone's job easy . Autonomous robots in industrial companies, on the other hand, are more likely to perform complex jobs . Not all robotics take the form of a human person. A good example is a robotic arm. They are tough, adaptable, and accurate enough to handle a wide range of jobs.

In our project, We focus on speech recognition and its application to robotic arms.The robotic arm we constructed is lightweight & easily usable and can be controlled using voice input in various regional languages.The arm will be able to move in any desired direction based on the human speech commands given. All the robotic arms can be used to replace the human work and in performing repetitive works. The arm is controlled using live voice instruction where the image model predicts the audio label based on the spectrogram and decodes the prediction to transfer it to the microcontroller using serial port connection.The microcontroller has defined functions for every input it receives from serial connection which will have the amount degrees arm should to move. All the existing work was based on English voice instructions and to make it useful for the wider population we have included Kannada instructions also. Our model enables native language speakers to use robotics to make their life easier irrespective of their education background, age and gender.

CHAPTER 1

INTRODUCTION

The first term to consider in voice control is Speech Reconnaissance, i.e. making the system to understand human vocalization. Speech recognition is a technology where the system understands the words given by oration.

Speech is a great way to control and communicate with robots. The speech recognition circuit we'll describe is a separate component of the robot's main intelligence. This is beneficial because word recognition does not consume any of the robot's main CPU processing capacity. To check if a command has been issued to the robot, the CPU only needs to poll the voice circuit's recognition lines every now and then. This can be further improved by connecting the recognition line to one of the robot's CPU interrupt lines. A recognised phrase would produce an interrupt, alerting the CPU to the fact that a recognised word had been pronounced. The advantage of employing an interrupt is that it eliminates the need to poll the circuit's recognition line on a regular basis, thereby decreasing CPU overhead.

Another advantage is the programmability of this autonomous speech recognition system. You may programme and train the speech recognition circuit to recognise the specific words you want. The Speech recognition circuit is simple to connect to the robot's CPU. It will be easier to control and command an appliance (VCR, computer, TV security system, and so on) by speaking to it. At its most basic level, speech recognition allows the user to complete many tasks simultaneously while working on the computer or appliance (i.e., hands and eyes are occupied elsewhere).

Robotics is a subject in rapid development. There are many ways of making robots, and no one knows which method or technology is going to be used in the next 100 years .Robotics is evolving in accordance with Darwin's theory of evolution of the fittest. A robot is an electromechanical device .Robotic Arm that can sense its environment and make decisions (command). In general, a robot must be able to move (mechanically), feel (via a transducer), and make decisions (through remote control or artificial intelligence). A robotic arm is a manipulator that can execute tasks comparable to those performed by a human arm. The industrial application of robotic arms is crucial.

The ability of robotic arms to operate in hostile situations and places that are inaccessible to people is their biggest benefit. There are many versions, including Keypad Controlled, Voice Controlled, Gesture Controlled, and more. However, the majority of industrial robots are still programmed the old-fashioned way, which is still a laborious and time-consuming process that demands technical expertise. As a result, there is a need for new and simpler robot programming techniques.

Humans are replaced by robots and automation while doing mundane, dangerous, challenging, or hazardous work. In the high-tech era of today, automation significantly increases production capacities, enhances product quality, and lowers production costs. To programme, monitor, and perform normal maintenance on the computer, just a small number of people are required.

Robotic arms are manipulators that have the appearance of a human arm and are capable of carrying out challenging things like welding, trimming, picking, and painting. Furthermore, their ability to operate in hazardous conditions that are inaccessible to human operators is their greatest advantage. This initiative's main goal is to help companies cut labor costs and employee errors while raising productivity. However, it can help the differently abled with daily duties with small structural alterations.

While there are several ways to achieve wireless control, including keypad control, gesture control, and vision-based techniques, doing so takes time and requires a high level of technical proficiency and topic expertise. As a result of a high level of abstraction from the robot programming language, the voice-based method, in contrast, is not only simple but also helps the user manage the arm, making it simple to operate these engineering marvels.

CHAPTER 2

PROBLEM STATEMENT

Robots are utilized in a variety of settings, particularly when they can do difficult jobs or execute missions that would be risky for humans. Aside from being exact and consistent, robots may operate in any setting, increasing their versatility. Because robots can operate in hazardous areas, they eliminate risky employment for people. They are capable of lifting big loads, handling dangerous chemicals, and doing repeated jobs. This has helped businesses avoid several accidents while also saving time and money. Robots are employed in the medical profession for complex procedures such as prostate cancer surgery. Robots can reach and fit in places where human hands can't, allowing for more precision.

Robots are now used everywhere to make people's jobs easier . Making these robots work in various fields to make all of the work automated, which will have less human power is in the recent research . Speech controlled robots will be more accurate and faster when compared, as they can process the inputs quickly . Robots are now used instead of humans to automatize the work. It can work in hazardous situations and when it comes to doing repetitive work robots will be able to do it with the same efficiency every time unlike humans where they will be bored of doing the same tasks. These robots can help people like farmers where it could be used for agricultural work , and people who are physically challenged .

Making the robots process not only English but also regional languages is very helpful as many people will be able to use it . Many of them who don't know languages now will be able to use the robotic arm in their native language . This will help many of the farmers who want to cultivate , harvest the crops . And in the military where the soldiers can use their native language which will help from rivalries as they won't be knowing the actions of the robot. Physically challenged people like deaf , dumb and blind people will able to use for their everyday needs instead of depending on someone else .

The problem statement focuses on controlling the robotic arm by voice commands , where the user will be able to give inputs in both English and Kannada languages. The Robotic arm will have two degrees of freedom(DOF), where two DOF means the number of movable joints the robot will have. The arm will have three different joints: shoulder, elbow and wrist so the arm will be able to rotate along these three axes. The Robotic arm is constructed using the microcontroller Arduino and we will have three servo motors as the robotic arm will have two DOF. The robotic arm will be constructed using cardboard which will be light weight, easy to construct and anyone with basic knowledge about the sensors and microcontroller will be able to construct the robotic arm. The live voice notes will be passed by integrating a microphone to the model.

The input to the robotic arm can be either pre-recorded or live voice instructions . As of now, we are building a robotic arm that will be able to process a single word input and do specific actions in both Kannada and English , which we will extend later to process statements where a robot will perform multiple tasks. Robotic arm can process live voice notes too . When a user sends a live voice input , he records it and the audio will be saved in a wav form . The audio will be then processed the same way as pre recorded voices.

This specific project is implemented in native language , so everyone will be able to use this in everyday life to make their job easier . To pick and place objects just through the voice commands without even doing any kind of work . After looking into ongoing research we found that there is not much research about the controlling robotic arm in the native language so if we build an arm that performs tasks based on the inputs given in native language. It will be helpful in various fields such as agriculture where farmers can control the robotic arm in their local language. This specific arm can be extended to do difficult jobs like weed removal , pesticide spraying.

CHAPTER 3

LITERATURE SURVEY

In this chapter, we present the current knowledge of the area and review substantial findings that help shape, inform and reform our study.

Background on Robotic Arm architecture and audio processing

This section details the papers read to gain information on background, data used, and the current methodologies being used in the field of robotics and audio processing.

Module 1:Construction of Robotic arm

[1] Rajesh Kannan Megalingam, Gedela Vamsy Vivek, Shiva Bandyopadhyay, Mohammed Juned Rahi(2017)"Robotic Arm Design, Development and Control for Agriculture Applications"Dol:10.1109/ICACCS.2017.8014623 August 2017.

In this paper the author examines numerous strategies and techniques for designing and developing a robotic arm for agricultural use.

Arm architecture:

A 3 degree-of-freedom robotic arm with joints that resemble those of a human hand. To give precise movement and control over the arm, servo systems were used at each joint. External 5:1 gears are installed on the servo motors to increase torque and hence improve their performance under load. A portable PC with Kinect SDK and an Arduino UNO are used to control the system.

working:

The study proposes a method for controlling a robotic arm with a depth sensor. A depth sensor captures human input, and an inverse kinematics technique is used to describe the robotic arms' movements.To solve inverse kinematics problem Jacobian Iteration method is used as a solution.Using inverse kinematics, the robotic arm captures, processes, and replicates the movement of the human arm in 3D space.First and foremost, a Kinect sensor takes the user's input and provides coordinates for the user's shoulder, elbow, and wrist joints. An inverse kinematics technique is used to calculate the angles at each joint required to completely characterize the location of the robot's joints. The robotic arm end effector moves to the desired location using a correct microcontroller board, which can be used to pluck a fruit or prune a branch, for example.

[2] Priyambada Mishra, Riki Patel,Trushit Upadhyaya, Arpan Desai “DEVELOPMENT OF ROBOTIC ARM USING ARDUINO UNO ”2017

Architecture:

Four potentiometers were utilized to control the servo motors in this work.They constructed the robotic arm out of cardboard. They are recommended for applications requiring low speed, medium torque, and precise positioning. This model appears to be a robotic crane.

Working:

The use of a servo motor to create robotic arm joints and regulate them with a potentiometer. The servo motors are controlled by an Arduino UNO board, and the Arduino's analogue input is delivered to a potentiometer.The user can manage the digital values of Arduino, which are used to alter the servo motor position. By rotating these pots, we can move the joints of the Robotic arm and pick, grab, or place any object.

[3] Miss. Musmade Gayatri, Miss. More Poonam, Miss. Lambe Deepti, Mr. Bansode Bipin"Voice and Gesture Controlled Robotic ARM"IJSRD - International Journal for Scientific Research & Development| Vol. 3, Issue 11, 2016 | ISSN (online): 2321-0613

System design:

With the help of the microphone, speech commands are given to the voice recognition module; the user can give commands into the microphone.The voice recognition module receives the analogue output of a microphone (EasyVR).It is capable of processing that signal and producing digital output. The signal is then passed to the ARM 7 controller, which passes it to the wireless transmission module (CC2500) for transmission. Another transreceiver in the robotic arm part receives the command/signal. The command or signal is subsequently sent to the ARM 7 controller. The ARM 7 controller sends a signal to the motor drive. We know that the ARM 7 controller's output is insufficient to drive a dc motor, so we utilize a motor drive to boost the signal's strength (L298). The motor driving signal is sent to several motors that govern the robotic arm's mobility.Hence robotic arm gives the movement according to the signal

[4] Md. Abdullah Al Ahasan, Md. Abdul Awal *and* Sheikh Shanawaz Mostaf"Implementation of Speech Recognition Based Robotic System"2011

Architecture:

In this paper,the authors used 2 motors.The ups and downs are controlled by motor one. The movement of the opening and closing is controlled by motor two. It can be halted here at any time the user wishes. Two of the robotic arm's joints can be moved. These joints are made by a 4 pole EM 546 stepper motor (found in an old printer). The gripper joint can move from 0 to 100 degrees, while motor 1 can move from 0 to 120 degrees.

working:

The audio is transmitted to the computer interface via the wireless technology.The audio is captured by the computer system's receiver driver.The program's sound to text converter subsystem subsequently converted the captured audio to a text string.For the exaction command, this text string was compared to the database.If a command is located in the text, the robot is given the

implementation command to complete the task.This project's software is written in the C# programming language. For speech rearrangement, Microsoft Speech SDK 5.1 is used.
Research Gap:sometimes it can't recognize command because of different pronunciation and different tone of different people

[5] Dino Dominic Ligutan, Alexander C. Abad, Elmer P. Dadios, Gokongwei College of Engineering, De La Salle University, Manila, "Adaptive robotic arm control using Artificial Network", 2018 IEEE 10th International Conference on Humanoid, Nanotechnology, Information technology, Communication Control,Environment and Management(HNICEM), 2018, Pages: 1 – 6.

This paper showcases an Artificial Neural Network-based Adaptive Control technique for training a 4-DOF robotic arm (ANN). At regular intervals, the technique comprises uniform sampling of the coordinates of the robot's joints and end effector in joint space. These coordinates are then preprocessed to create a jacobian matrix, which is subsequently sent into a neural network to produce the upper and lower joint limits that are utilized to maneuver the arm. The architecture of this robotic arm's controller is unique in that it properly compensates for loading effects with an absolute inaccuracy of less than 1%. (5 millimeters).

Module 2: Audio processing

[6] Lajish V.L., Vivek P. and R.K. Sunil Kumar"Malayalam speech controlled multipurpose Robotic arm"26th Kerala Science Congress Pookode, Wayanad: 28-31 Jan. 2014

The authors created a robot arm that uses Malayalam Speech commands to control it.

Methodologies:

Mel Frequency Cepstral Coefficients(MFCC) are used to extract the commands' features. Dynamic Time Warping (DTW) techniques are used to recognise commands.

Working:

The word units are captured at 8 kHz, quantized at 16 bits, and then processed at 10 ms frames per second with a 25-millisecond overlapping Hamming window. The speech units are then parameterized using 12 Mel Frequency Cepstral Coefficients (MFCC), normalized log energy, and first and second order differences, providing a total of 39 components. As features matching approaches in the next phase, Sakoe Chiba's non linear sequence alignment, known as DTW, was applied. This algorithm compares two time series that differ in time or speed, as well as finding the best alignment between them if one of the time series is "warped" non-linearly by expanding or contracting it along its time axis.The code for the spoken instruction is sent to the microcontroller for the appropriate arm motion, which is recognised from the SR unit. The motor driver IC in the microcontroller runs the motors in SCMRA in response to the received instructions.

[7] Gupta D, Bansal P, Choudhary K (2018) The state of the art of feature extraction techniques in speech recognition. In: Agrawal S, Devi A, Wason R, Bansal P (eds) Speech and language processing for human-machine communications, vol 664. Advances in intelligent systems and computing. Springer, Singapore, pp 197–207

Loudness, pitch, and quality are all characteristics of sound. Different sorts of parametric representations based on these features can be used to represent the voice signal. Mel frequency cepstral coefficients (MFCC) and Bark frequency cepstral coefficients are two available techniques for this (BFCC). The Mel scale is a perceptual scale based on signal "pitch" measurement. Mel scale cutoff frequencies represent nonlinear frequency perception in the human auditory system.

Mel scale suggests a maximum sampling frequency of 14,000 Hz.The crucial bandwidth of the human hearing spectrum is divided into 24 nonoverlapping critical bands. In the Bark scale, each crucial bandwidth corresponds to one bark. The Bark scale's 24 center frequencies are fixed. The bark scale is based on the "loudness" of a voice signal . The Bark scale recommends a sample frequency of 15,500 Hz.

[8] Himgauri Kondhalkar and Prachi Mukherji"Speech Recognition Using Novel Diatonic Frequency Cepstral Coefficients and Hybrid Neuro Fuzzy Classifier"

The authors of this work proposed the DFCC feature extraction approach, which is based on diatonic scale frequencies. The overall classification accuracy of this approach is 95.20 percent, compared to 93.50 percent for the previous MFCC technique. Because diatonic scale cut off frequencies can be generated for any sampling frequency value, DFCC pattern synthesis can be employed for a variety of applications such as voice recognition and music classification. We created a hybrid NF classifier in the second stage of the proposed study to recognise patterns generated by the DFCC algorithm. The accuracy of the DFCC and NF classifier combination is 99.23%, which is much greater than the accuracy of existing SVM and NN classification approaches. In the suggested work, the NF classifier is implemented in a fresh and unique way.

[9] Dr.Punitha.P and Hemakumar"Speaker Dependent Continuous Kannada Speech Recognition Using HMM"2014 International Conference on Intelligent Computing Applications

The problem of Kannada speech recognition is addressed in this study. Using the HMM Method, the suggested system detects continuous Kannada speech in speaker-dependent mode. The proposed method preprocesses the original Kannada speech signal before framing it every 20 millisecond with a 6.5 millisecond overlapping interval. Second, the voiced section is identified by establishing a dynamic threshold based on signal size and short time energy. Finally, in the voiced part of the signal, LinearPredictive Coding (LPC) coefficients are extracted and converted into Real Cepstrum Coefficients. Fourth, the Real Cepstrum Coefficients are processed through a k-means clustering algorithm with k = 3 and then through the Baum-Welch Algorithm, which creates a three-state HMM model for each syllable, subword, or sentence.This experiment employed 20 different words that can be used to control small mobile devices.Each of these phrases was recorded ten times for training purposes and three times for testing purposes on one male speaker. In studies, the command success rate of individually stated words was good, with an accuracy rate of 87.76 percent and a miss rate of roughly 12.24 percent, precision of 0.56, recall rate of 0.68, and F1 measure of 0.61. The Math lab is used for computations.

CHAPTER 4

DATA

This chapter serves to describe the data under consideration. To check the proper working of any algorithm built, data is a key source required .

Overview:

The dataset we are using contains audio files in 16 bit pcm.The input to the constructed robotic arm can be either pre-recorded or live voice notes, which will either be in English or Kannada. The input should be a single-word voice command wherein the arm will perform the desired action irrespective of the language. All the pre recorded audio data inputs can be used to train data and live voice inputs and are given at instant. All the voice inputs are grouped under the label it belongs which is required to test the predictions. All voice data are unique no two voice inputs of the same user are present in data.

English single-Word voice data

- We found a dataset where in the voice data was used to recognize the words , we can use the same for controlling the robotic arm by just considering the words or commands the arm can perform .
- The dataset has specific words for which a robotic arm can do specific actions .
- Complex actions like picking will not be performed and robots will be able to perform only actions which will have only three DOF involved .
- The basic actions like moving up , down , left , right will be performed .

Kannada single-Word voice data

- Due to non-availability of the kannada dataset, we have created the dataset by recording the voice instructions manually
- Recorded voice was not in format to preprocess , so we had to convert it to 16 bit pcm.
- All the voice notes are unique.

English Single-word voice	Kannada	Action
UP	Mele	The robotic arm will be lifted up
DOWN	Kelage	The robotic arm will be moved down
LEFT	Yadake	The robotic arm will be moved to the left side
RIGHT	Balake	The robotic arm will be moved to the right side

The model would predict the output as the word which is then decoded to u , d , l , r which is useful for serial communication. The output of the model which is decoded will be sent to the microcontroller.

CHAPTER 5

PROJECT REQUIREMENTS SPECIFICATION

The following chapter annotates the requirement study of the project and the functionalities that need to be satisfied. The subsections in the chapter consist of all the features of the products comprising functional and non-functional requirements along with the bottleneck, expected risks, constraints, and the expected outcomes.

5.1 Product Features

The functionalities that are expected to be delivered by the project are as follows:

- Lightweight and easily usable robotic ARM construction
- Arm can be processed in both English and Kannada Language voice instructions.
- A pre-trained model will predict the output of the live voice instructions
- The output is decoded into letters , which is further used for serial communication with microcontroller.
- Microcontroller calls the specific functions in which the degree it should be moved is specified.
- Arm will move to a specific direction that is predicted by the model.

5.2 Functional Requirements

- The arm will receive the input in the form of pre-recorded voice notes or live voice notes.. The data set includes voice instructions in different voices, in both the considered languages. This is done to improve the efficiency of the arm.
- The user will have to select the language and the instruction he wishes to pass to the arm. The processed voice notes will be fed into the arm to get the desired result.
- There is nothing as error input , because the model would predict the audio which is maximum related to , so when the user gives a wrong input arm will function.
- The voice notes are the inputs to the arm. The instruction is processed and passed to the arm and the arm performs the desired action.

- The consequences of limiting the degree of freedom, torque and the languages are that the arm will not be able to perform heavy load tasks and can be controlled by only a small percentage of the population.

5.3 Non-Functional Requirements

5.3.1 Performance requirements

- The system will have high reliability .
- Robotic arms to be built with different levels of repeatability depending on their intended application. The degree of movement can be altered according to the user requirements.
- Ensure minimal response time during the prediction of the output from the model.
- The speed of the servo motors and the degree of rotation can be altered according to the user's specification.
- Simple and lightweight arm construction
- Environment is one of the most difficult hurdles in a speech recognition system. Background noise, room acoustics, and channel conditions can all be used to portray the environment. These settings increase signal noise and generate vocal noise.

5.3.2 Safety Requirements

- Preventive maintenance should be provided before running the machine, including self-diagnostics of the robotic arm control system, visual inspection of cables and connectors, battery or abnormal signals like noise or vibrations.
- The servo motors we use will require electricity so it is always necessary to take precautions.

CHAPTER 6

SYSTEM REQUIREMENTS SPECIFICATION

This chapter details the system requirements of the project.

6.1 Current System

There are several methods for establishing wireless control currently available, including keypad control, gesture control, and vision-based techniques; however, these methods are time consuming and need a high level of technical ability and topic knowledge. The voice-based method, on the other hand, is not only simple, but it also aids the user in operating the arm, thanks to a high level of abstraction from the robot programming language, making these engineering marvels simple to operate.

6.2 Design Considerations

6.2.1 Design Goals

- Design a robotic arm suitable to work with 2 degree of freedom and which is not too bulky and also compatible to use
- Basic components were used so it would cost very little.
- The serial connection is required to send input to microcontroller
- The Model is pre-trained with data so that it will be useful to predict the output.

6.2.2 Operating Environment

- The hardware components used for the construction of the robotic arm are Microcontroller , capacitors , servo motors , variable resistors , sensors and potentiometers .
- The built robotic arm receives input as pre -recorded voice and based on the voice the user inputs the arm will be performing the desired tasks .

6.3 Constraints, Assumptions, and Dependencies

- The medium of instruction is in English and Kannada only.
- The user is expected to give input clearly.
- Only English and Kannada language are accepted and it works only for these instructions.
- We restrict the robotic arm to work for only limited actions which will have 2 DOF(degree of freedom).
- We will be sending the recorded voices , and also the live recorded voices .

6.4 Risks

- The hardware component may not be that efficiently working
- Creating the dataset as it contains regional language so it will be tedious task (suitable different voice notes)
- Compatibility of the considered devices with each other

6.5 External Interface Requirements

6.5.1 User Interfaces

- The inputs for the robotic arm are pre recorded voice notes wherein the user will be able to send voice inputs of his own choice .
- User will be able to select any voice note of his choice , the robotic arm will be able to do desired tasks based on the instruction decoded from the input.
- If the inputs processed cannot be decoded into arm instructions , the arm will not be able to do any kind of task . (We can indicate these kinds of in
- puts by making the robot do some action so that the user will know)
- We could use online platforms for simulation and error messages can be displayed as dialog boxes .

6.5.2 Hardware Requirements

- It requires hardware components like microcontroller , sensors , motors , resistors , capacitors , potentiometer actuators for building the arm .
- Number of motors required for the functioning of the arm depends on the degrees of freedom. We will be building an arm with 5 dof which will require 5 servo motors with specific amounts of torque .
- The microcontroller programming will be in C , python
- The inputs sent to arm are pre recorded voices which will be processed (like feature extraction) and decoded into arm instructions

6.5.3 Software Requirements

- Python
- Tensorflow
- py audio
- py Serial
- Arduino IDE

CHAPTER 7

SYSTEM DESIGN

This chapter describes the design aspects of the project. High-level design, system perspectives, design of classes, and UML diagrams that have helped better clarity of execution of the problem statement are presented in detail.

7.1 High-Level System Design

7.1.1 Logical user groups

The user of this application is:

a. User who sends the required voice input .

7.1.2 Application components

1.Data components

1. Different kinds of voice data , desired action , and robotic arm instruction .

2. Interfacing systems

Receiving pre recorded voice inputs and decoded arm instruction to the arm

7.2 System perspectives

This section describes the project design from different perspectives, thereby covering all relevant factors that need to be taken into consideration while designing and developing the product.

7.2.1 Conceptual or Logical Perspective

This section elaborates on the logical functional elements of the system.

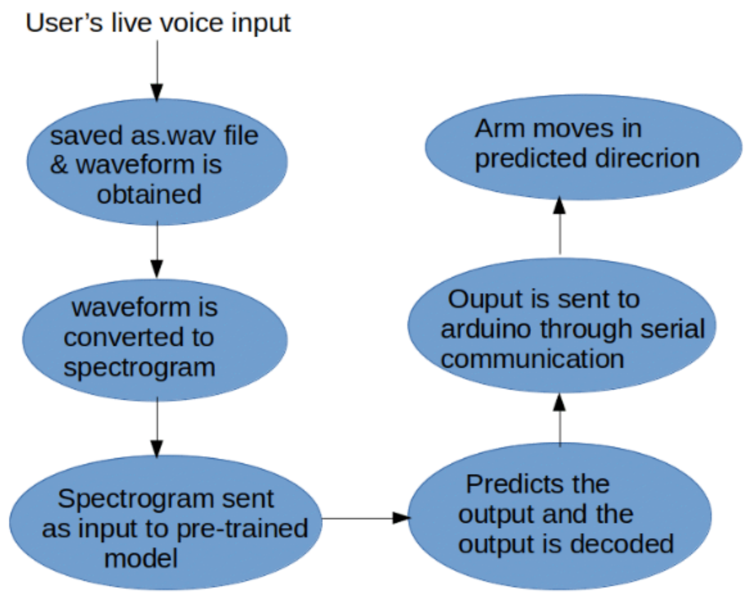


Figure 7.2.1 Control flow Diagram

This diagram is also called a control flow diagram or a data flow diagram. This high-level diagram is one of the basic and required deliverables in the design and architecture phase. It is used by a non-expert to understand the functionality of the system without much difficulty. It doesn't have many technical aspects involved and is simple to understand. It is a flowchart that shows all the branches and conditions which the system can take while operating in its course time.

7.2.2 Physical Perspective

The Physical view of the distributed system can be described as:

1. Devices - end-user robotic arm

2. Artifacts: independent modules - Input voice data , Pre-trained model, Decoding the output to send it to microcontroller, Serial connection between software and hardware.

1. **Project management -** The project is divided into distinct phases that must be developed individually and then integrated as part of the logical flow. The main phases are as follows:

   a. Construction of the hardware of the robotic arm which works for keyboard instructions

   b. Pre-trained model which predicts the output.

   c. Integration of software and hardware parts .

2. **Code Organization -** Each of the above modules will be developed individually, ensuring reusability wherever required, and later integrated together. Code will be documented and maintained in detail.

7.2.3 Reusability Considerations

The project design allows the reusability of various components with minor modifications.

1. Pre - trained Model - The application relies on the input given by the user by interacting with the system. All functionality that requires user interaction will make use of the same modules.
2. Robotic arm construction modules: Arm construction will remain the same , only the action it performs will differ for different input and any other mode of controlling .

7.3 Use Case Diagram

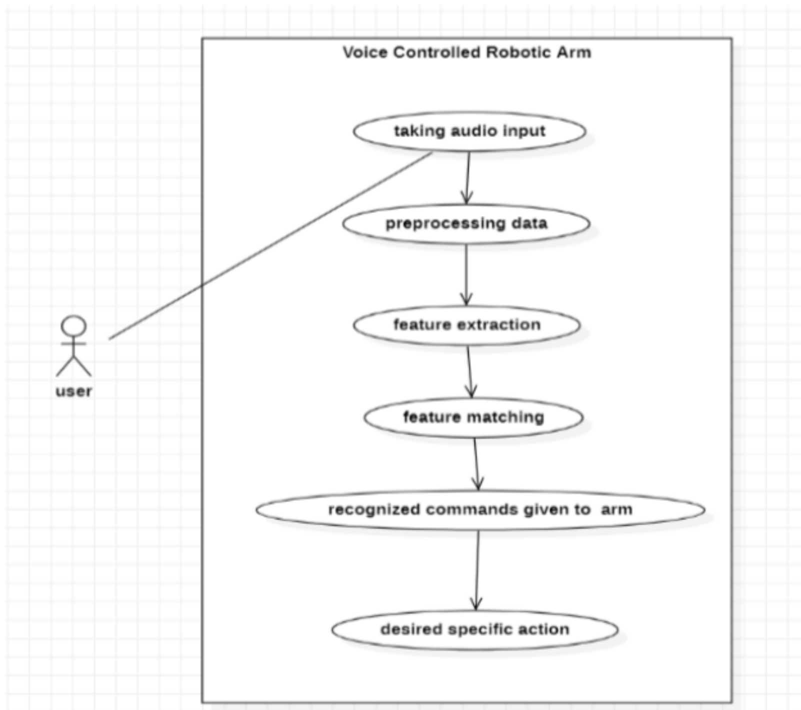


Figure 7.3.1 Use case Diagram

Use a case diagram in the design description to depict the overall functionalities of the application. It also shows how all users of an application from the external system interact with the application.

High Level Design Architecture:

A high-level design defines the system architecture at the project level. Architecture defines the subsystems to be built, the internal and external interfaces to be developed, and the interface standards to be defined.

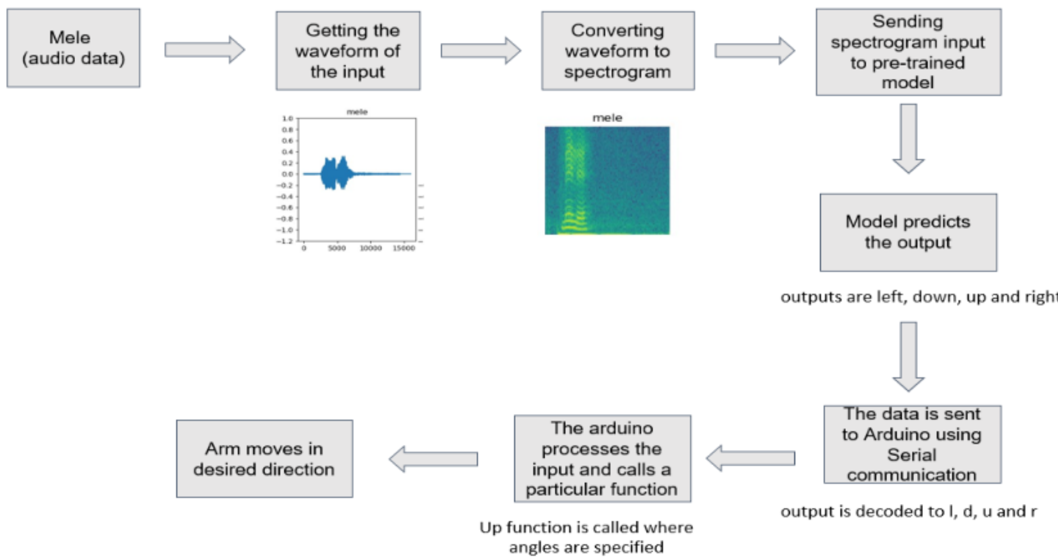


Figure 7.3.2 High level design architecture

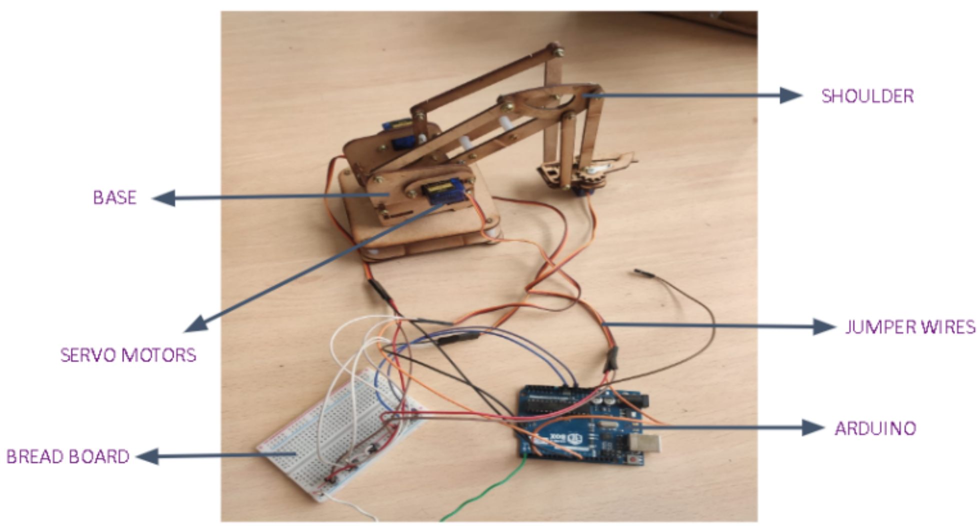
In this the user is given to choose his own choice of voice input in english or kannada . The chosen input is then given a high level design where the waveform of the input is converted to spectrogram, The spectrogram is then sent to the pre-trained model where the model predicts the output and the output is decoded to communicate with the microcontroller. The decoded output of the model is received as input from the microcontroller . The microcontroller has a function defined for every input which has the amount of degree the arm should move. Based on the input given function the particular function is called and the arm will move in the desired direction.

CHAPTER 8

PROPOSED METHODOLOGY

1. Robotic arm construction

- The electronic components used include Arduino UNO microcontroller, Servo motors and UART.
- The body of the arm is constructed using cardboard to make the model light weight and easy to construct
- The model mimics the human arm with the base acting as shoulder, an elbow to move up and down followed by the gripper which is the wrist.
- The Arduino board is dumped with the code which has different functions for different inputs (l,r,d,u). The servo motors enable movement of the arm. The arm has 2 DOF and can therefore move up, down, left or right.
- The communication between the Machine Learning model and the arm is established using UART. UART enables Serial Communication between the two entities.



2. Pre-processing

- Using tf.audio, we convert the audio to a waveform during the preprocessing stage. Since Dataset and decode wav only contain single channel audio, the excess axis is removed using the tf.squeeze function.
- We generate the label and waveform for each audio file, then transform the audio dataset into a waveform dataset.

3. Feature extraction

During feature extraction, waveforms are transformed into spectrograms.

- The dataset's waveforms are displayed in the time domain.By computing the short-time Fourier transform (STFT), which shows frequency variations over time and may be represented as 2D pictures, we convert the waveforms from the time-domain signals into the time-frequency-domain signals.
- A signal is broken down into its component frequencies using the Fourier transform (tf.signal.fft), but all temporal information is lost.STFT (tf.signal.stft) on the other hand divides the signal into time windows, performs a Fourier transform on each window, keeping some time information, and returns a 2D tensor.
- The waveforms must be the same length in order for the resulting spectrograms to have equal dimensions. Simply zero- padding the audio segments that are under a second in length (using tf.zeros) will accomplish this.
- Choose the frame length and frame step arguments when calling tf.signal.stft so that the resulting spectrogram "image" is nearly square.
- An array of complex integers that represent magnitude and phase are generated by the STFT. We only use the magnitude, which we can determine by applying tf.abs to the tf.signal.stft output, though.

4. Feature Recognition

Since we converted the audio data into spectrogram images, we built a simple convolution neural network (CNN) model for feature recognition.

- tf.keras.Sequential model will use the following Keras preprocessing layers
- Resizing: downsampling the input to make the model train more quickly.

- Normalization: To normalize an image's pixels according to their mean and standard deviation.To compute aggregate statistics (that is, the mean and the standard deviation) for the Normalization layer, its adapt method would first need to be invoked on the training data.

Convolution neural networks perform far better than other artificial neural networks due to three extremely unique processes: convolution, pooling, and flattening.

Convolution layer

2 convolutional layers with relu activation function and 64 and 32 kernels and kernel size of 3

- Convolution is just searching an image by moving a kernel-based filter across it to find various aspects of the picture.Kernels are merely 2D matrices with various weights.Basically, as this kernel passes across the image, the pixel values are replaced with the average of the weighted sum of their weight for that particular section of the image.These kernels are an incredible tool for locating the image's key details.

Pooling layer

- We must pool the features after adding the convolution layer to our model.Pooling merely shrinks the image without erasing the information we discovered during convolution. We employ a single MaxPooling layer, whose method returns the larger value within a given range and takes the form of a matrix.By doing this, we may compress the image without sacrificing

Flattening layer

- The final stage in processing the image is to apply a flattening layer, which is nothing more than turning a 3D or 2D matrix into a 1D input for the model.
- Fully connected Dense layer: used for classification.

Additionally, we employ the Adam optimizer for compilation, SparseCategoricalCrossentropy() for the loss, and accuracy and data fitting for the metrics. Also, train the model across ten iterations.

For feature recognition, we construct a simple artificial neural network (ANN) model.

ANN:

For feature recognition, we construct a simple artificial neural network (ANN) model.

tf.keras.Sequential model will use the following Keras preprocessing layers:

- **Resizing:** downsampling the input to make the model train more quickly.
- **Normalization** : Each pixel in the image should be normalized depending on its mean and standard deviation.
- 2D and 3D data are first flattened into a single layer, and then input is sent to dense, fully connected layers for further categorization.

CHAPTER 9

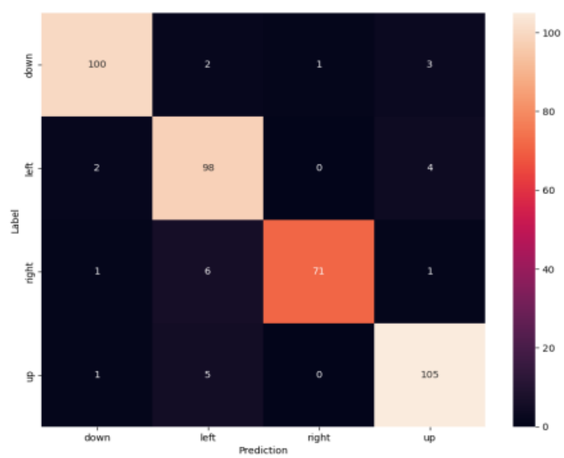
RESULTS AND DISCUSSION

In this section, we discuss the results obtained from all the test cases we have conducted and have inferred results from the same.

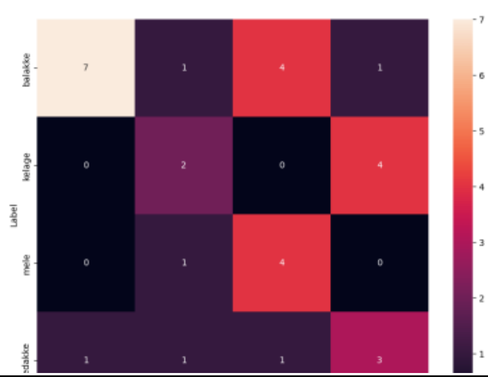
We obtained the following confusion matrix when we tested the model with the voice dataset of both English and Kannada.

For CNN model:

Only English voices were sent .
input voice instructions are shuffled and passed to the model.
Validation accuracy: 94.66%
Test accuracy: 94%

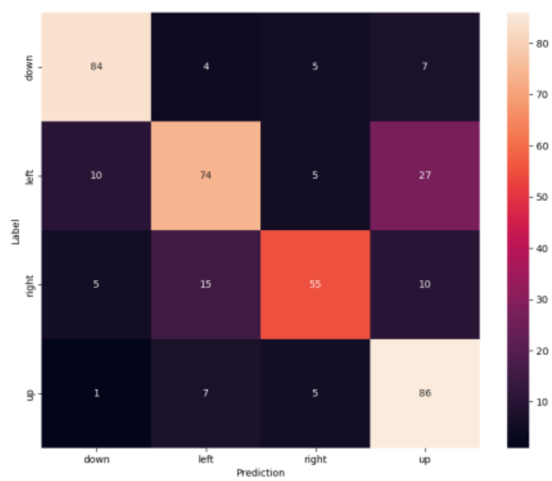


Only Kannada dataset
Validation accuracy:81.67%
Test accuracy:78%

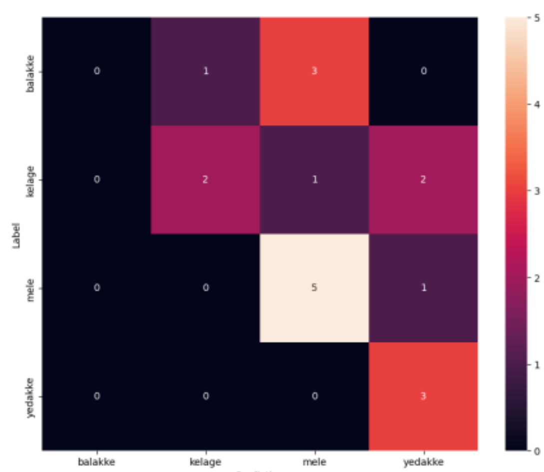


For ANN model:

English Dataset
Validation accuracy:78.31%
Test accuracy: 75%



Kannada Dataset
Validation accuracy:78.31%
Test accuracy: 75%



Comparing both the models we obtain maximum accuracy for the English dataset using CNN model. CNN models have better accuracy than ANN models. Higher accuracy was obtained in English due to a larger dataset of 4000 voice instructions.

CHAPTER 10

CONCLUSION OF CAPSTONE PROJECT PHASE-II

It began with shortlisting a few problem statements in our area of interest, selecting one out of them, and then developing a problem statement. We expressed our interest in agriculture and robotics, which led to the formulation of a problem statement. The developed robotic arm will be able to process the user's instruction in both Kannada and English and perform the desired action.

During the phase-1 of our capstone project, we were focused on extensive research about the architecture of the robotic arm along with the collection data set. We have decided on the methodologies to build the robotic arm and to process the voice notes. The deliverables of capstone phase-1 were data set collection in both Kannada and English along with the finalized methodologies. The data sets collected in English have instructions in single words like "Up, Down" whereas the dataset had to be created manually for Kannada.

The main goal of our project is to process voice notes and ensure that the robotic arm moves in the specified direction. The project can be broken down into 4 parts. The first part includes building of the robotic arm which is able to process voice instructions. The robotic arm comprises the hardware part of the project. The second part involves audio processing of the voice instructions which needs to be fed into the robotic arm. The third part involves passing the voice notes to the robotic arm in a way that the machine will be able to understand it. The fourth part involves ensuring that the arm moves in the desired direction which is done using Machine Learning.

The phase-2 of the project include building of the arm and implementing machine learning models to pass the voice instructions to the arm using Serial communication. We have met all the requirements of Phase-2 of our project. We have built the robotic arm using cardboard and servo motors as specified in the Phase-1 of our project. We have implemented a couple of machine learning models to ensure that highest accuracy is attained during feature recognition and extraction. We have improvised the project by including UART which enables serial communication between the system and Arduino/arm. The original plan included only pre-recorded voice notes; we have now integrated a microphone into the model which enables us to pass live voice instructions to the system.

The constraints of our proposed model include: the robotic arm has limited degrees of freedom which limits the movement of the robotic arm and the arm will be built using cardboard which makes it difficult for the arm to pick and place heavy load objects. The assumptions made are that all the considered components are compatible with each other and that the components can be reused as their configuration will not be changed.

Our model doesn't require any personal and physiological biometrics of the user which makes it safe to use. The model will be able to process instructions in both Kannada and English and makes use of pre-processed and live voice notes which makes it accessible to a wider population.

CHAPTER 11

PLAN OF WORK FOR CAPSTONE PROJECT PHASE-3

Phase 3 deals with the writing and publication of the paper. Also extending the word instructions to speech instructions, such that the robotic arm interprets the specific command from that sentence.

Voice controlled robotic arm

ORIGINALITY REPORT

13%	9%	6%	5%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	dokumen.pub Internet Source	2%
2	www.researchgate.net Internet Source	2%
3	Punitha P., Hemakumar G.. "Speaker Dependent Continuous Kannada Speech Recognition Using HMM", 2014 International Conference on Intelligent Computing Applications, 2014 Publication	2%
4	Submitted to PES University Student Paper	2%
5	www.tensorflow.org Internet Source	2%
6	Rajesh Kannan Megalingam, Gedela Vamsy Vivek, Shiva Bandyopadhyay, Mohammed Juned Rahi. "Robotic arm design, development and control for agriculture applications", 2017 4th International Conference on Advanced	1%

Computing and Communication Systems (ICACCS), 2017
Publication

7	www.slideshare.net Internet Source	1%
8	Submitted to A'Sharqiyah University, Oman Student Paper	1%
9	Submitted to Nelson Mandela Metropolitan University Student Paper	1%

Exclude quotes	On	Exclude matches	< 50 words
Exclude bibliography	On		