

Raktim Gautam Goswami<sup>1</sup>, Abhishek Bairagi<sup>2</sup> & G V V Sharma<sup>3</sup>

## CONTENTS

<b>1</b>	<b>Hardware Setup</b>	<b>1</b>
<b>2</b>	<b>Implementation</b>	<b>1</b>

**Abstract**—This manual shows how to develop a voice recognition algorithm and use it to control a toycar.

### 1 HARDWARE SETUP

- 1.1 Assemble the motors, chassis and wheels to build the toycar.
- 1.2 Stick the breadboard to the chassis of the toycar.
- 1.3 Stick a 9V battery to the breadboard and connect the positive and negative terminals to extreme ends of the breadboard.
- 1.4 Stick a 9V battery to the breadboard and connect the positive and negative terminals to extreme ends of the breadboard.
- 1.5 Provide 9V to the supply pin of the Arduino.
- 1.6 Plug the L293D motor driver IC in Fig. ?? on the breadboard.
- 1.7 Connect the L293D pins according to Table ??.
- 1.8 Connect the HC05 pins according to Table ??.

### 2 IMPLEMENTATION

- 2.1 Dump the following code in Arduino using its IDE.

```
wget https://raw.githubusercontent.com/gadepall/EE1390/master/bot/codes/vcb.cpp
```

- 2.2 Install Google API "Arduino Bluetooth Controller" using google play-store
- 2.3 Open the app and connect to HC-05.

The authors are with the Department of Electrical Engineering, Indian Institute of Technology, Hyderabad 502285 India . e-mail: 1. ee17btech11004@iith.ac.in, 2. ee17btech11051@iith.ac.in, 3. gade-pall@iith.ac.in

- 2.4 Open voice control section in the app and tap to give following commands.  
*Left, Right, Forward, Back & Stop*

## 3 BUILDING THE NEURAL NETWORK

### 3.1 Theory

We have used linear regression in our model. Here, all the features are tried to be approximated using an n-dimensional straight line (n being the number of features). The equation used for this is

$$\sum_i W_i * x_i + b \quad (3.1)$$

In matrix form it is

$$out = W.X + B$$

The output(out) is then put as input to the sigmoid function and the output of it is a number scaled between 0 and 1. This is the actual output(Y') we are interested in . The sigmoid function is defined as

$$sigmoid(x) = 1/(1 + exp(-x))$$

The cost function is then calculated using mean squared error as

$$J = 0.5 * (Y - Y')^2$$

Gradient descent algorithm is used to get minimum error using the derivative of the error(J) with respect to weight (W).This process is carried on for a number of times to get the best accuracy.

a) *How is the descent algorithm obtained from the cost function?*

: We initialized the parameters W1 and b . Now we want Mean Square Error function to be minimum.The way we do this is by taking the derivative (the tangential line to a function) of our cost function with respect to each parameter. Derivative at that point and it will give us a direction to move

towards. And then we update the value of all the parameters according to the derivative obtained. And then we iterate the process (number of iterations are decided by us). We make steps down the cost function in the direction with the steepest descent. The size of each step is determined by the parameter  $\alpha$ , which is called the learning rate. The gradient descent algorithm is repeated until convergence:

$$M_j := M_j - (\text{learningrate}) * (\text{deltaLoss}) * \text{input}$$

### 3.2 Python code

<https://github.com/rakimgg/ML-algorithm-for-speech-recognition/blob/master/code.py>

This is the full code that is used for training. The accuracy we are getting is around 98 percent.

### 3.3 Dataset

We have made our own dataset by recording 25 samples of each word. Each of these samples are recreated by adding empty elements in the front and back in many different combinations to create a dataset of 6250 samples for each word. All the audio files are imported to an array in the code and converted to mfcc format before training. For creating training dataset we recorded 25 audio files of each of the following words -

- 1) Forward
- 2) Left
- 3) Right
- 4) Back
- 5) Stop

The code for generating 6250 samples for each word from 25 samples can be found in the github link attached.

<https://github.com/abhishekbairagi/Making-Dataset-for-ML/blob/master/250files.py>

## 4 TRANSFERING THE WEIGHTS TO RASPBERRY PI (YET TO BE DONE)

The weights ( $W1$  and  $B$ ) are saved in a file at the end of the code. These weights will be transferred to the raspberry pi and a simple program written, will record audio on the raspberry pi, do the calculations using the weights and predict the text output. This output will be sent, using bluetooth, to the toy car, which will move accordingly.