**INTRODUCTION**

**Background**

Since evolution of mankind, human are using speech as the means of communication between each other. It is noticeable that it is the easiest mode of communication between two humans. Human brain can process the sound in the environment sensing it through the ears and can easily differentiate between language and speech. Language is the system/symbol of representation of information that multiple people/objects/living beings understand and can decipher it. We propose to develop a system that learns the human means of speech communication and recognize the patterns of the acoustic models with the help of deep neural network. Furthermore the classified output is applied for controlling the Robot Car.

**Importance**

Machine learning is fast growing engineering field. Though many machine learning algorithms were previously in existence, the ability to automatically apply complex mathematical calculations to big data – over and over, faster and faster – is a recent development. It is widely used everywhere from self-driving Google car to suggesting friends and videos on Facebook and YouTube respectively.

**Objectives**

1. To develop a system which recognizes speech
2. To develop algorithms that iteratively learns from data
3. To train a machine to react to certain input patterns (speech)

**PROCEDURE**

**Approach**

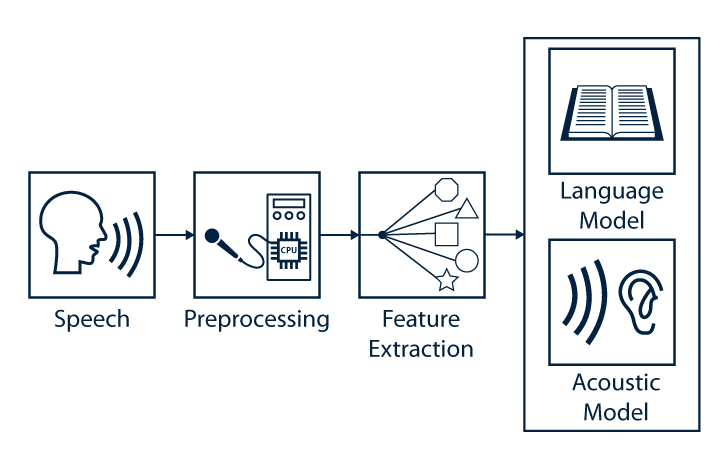


Fig 1: Speech Recognition Engine

**Speech**

Speech refers to the auditory and vocal medium typical used by humans to convey language. Furthermore, speech contains much more information than the spoken message, the linguistic message. Speech also includes information about speaker such as gender, age or emotional state.

**Preprocessing**

The preprocessing stage in speech recognition systems is used in order to increase the efficiency of subsequent feature extraction and classification stages and therefore to improve the overall recognition performance. Commonly the preprocessing includes the sampling step, a windowing and a denoising step. At the end of the preprocessing the compressed and filtered speech frames are forwarded to the feature extraction stage. The general preprocessing pipeline is depicted in the following figure.

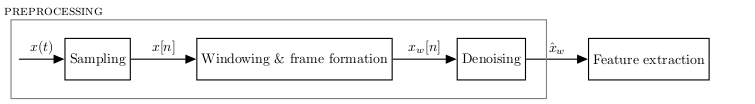


Fig 2: Preprocessing

**Feature extraction**

After the preprocessing step, feature extraction is the second component of automatic speech recognition (ASR) systems. This component should derive descriptive features from the windowed and enhanced speech signal to enable a classification of sounds. The feature extraction is needed because the raw speech signal contains information besides the linguistic message and has a high dimensionality. Both characteristics of the raw speech signal would be unfeasible for the classification of sounds and result in a high word error rate. Therefore, the feature extraction algorithm derives a characteristic feature vector with a lower dimensionality, which is used for the classification of sounds.

**Acoustic models**

At this point in our speech recognition system we assume that the acoustic waveform has been transformed into a feature vector by one of the mentioned feature extraction techniques. Now the task of the decoder is to find the sequence of words which are most likely to have generated feature vector. This means that we can estimate the word sequence by choosing, one after the other, and the word which maximizes the conditional probability

**Application**

We plan to apply the output from the speech recognition engine to control the movement of the Robot Car driven by the motor driver L293D

Audio Input

(Voice Command)

Microphone

Speech Processing and Recognition

Raspberry PI

Robot Car

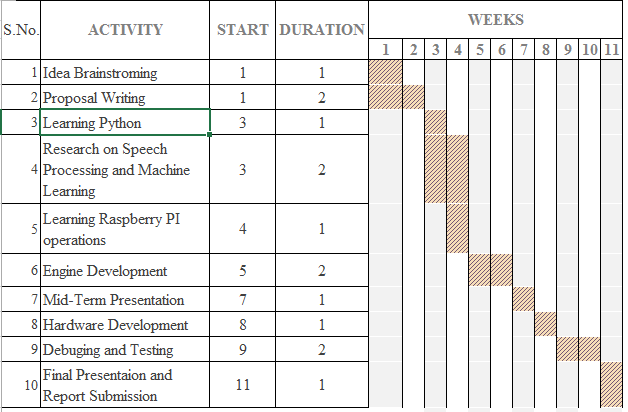
L293D IC

DC Motors

Fig 3: Data flow Diagram

**Gantt Chart**

Table 1:



**Cost Estimation**

Table 2:

|  |  |  |  |
| --- | --- | --- | --- |
| SN | Component | IC | Price(Nrs) |
| 1 | Raspberry Pi 3 |  | 6,000 |
| 2 | Motor Driver | L293D | 200-300 |
| 3 | Car Chassis |  | 1,000 |
| 4 | USB Microphone |  | 1,000 |
| 5 | Adapter(5v,2A),cable |  | 1,500 |
| 6 | LCD Display(16\*2) |  | 800 |
| 7 | Miscellaneous |  | 2,000 |
|  | Total |  | 11,800 |

**Design**

Audio Input

USB

LCD DISPLAY

G

P

I

O

P

I

N

s

L293D

Motor Driver

Raspberry PI MODULE

5V DC

Motors

Fig 4: System Block Diagram

**Expected Outcome**

We expect our system to recognize speech. System learns iteratively from the data through algorithms and trains the machine to react to certain input patterns. The machine consists of a robot car that reacts according to speech command. Speech command consists of START, STOP, RIGHT, LEFT, and BACK.

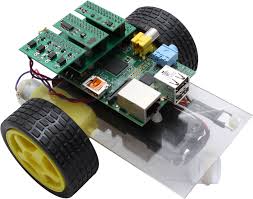


Fig 5: Expected Robot Car

**Conclusion**

In this project we will learn Machine learning using deep neural network. We plan to design a system that processes speech, recognizes pattern and reacts accordingly. The proposed research will achieve the following goals:

1. Learn Machine learning.
2. Implement modular approach in system development.
3. Develop teamwork spirit.
4. Learn hardware application of machine learning.