International Journal of Student Research in Technology & Management
Vol 1 (1), pg 50-57

SPEECH CONTROLLED ROBO-CAR

Pradeep L. Yadav[1], Manali N. Tawde[2] Prof. Sanjay [3]B.Deshmukh, [4]Dimple M. Kothari
EXTC Department
Dwarkadas J. Sanghvi College of Engineering, Vile-Parle (West), Mumbai
[1]pradeepyadav28081989@gmail.com, [2]tawdemn@gmail.com, [3]s_deshmukh@rediff.com,
[4]dimpskothari_28@yahoo.com

Abstract

The main goal of this paper is to introduce "hearing" sensor and also the speech synthesis to the robotic car such that it is capable to interact with human through Spoken Natural Language (NL). Speech recognition (SR) is a prominent technology, which helps us to introduce "hearing" as well as Natural Language (NL) interface through Speech for the interaction. The most challenging part of the entire system is designing and interfacing various stages together. Our approach was to get the analog voice signal being digitized. The frequency and pitch of words be stored in a memory. These stored words will be used for matching with the words spoken. When the match is found, the system outputs the address of stored words. Hence we have to decode the address and according to the address sensed, the car will perform the required task. Since we wanted the car to be wireless, we used RF module. The address was decoded using microcontroller (DSPIC30F) and then applied to RF module. This together with driver circuit at receivers end made complete intelligent systems.

I. INTRODUCTION

Think about a creating a car which would be controlled by your voice. By giving a command, the car would drive you to your destination. The voice recognition algorithm we used could be applied to daily life; for example it would be most helpful to disabled people to perform their daily work [1]. We created a speech controlled car using various electrical and mechanical domains such as digital signal processing, analog circuit design, and interfacing the car.

When we say voice control, the first term to be considered is Speech Recognition i.e. making the system to understand human voice. Speech recognition is a technology where the system understands the words (not its meaning) given through speech.

The purpose of this project is to build a robotic car which could be controlled using voice commands.

Generally these kinds of systems are known as Speech Controlled Automation Systems (SCAS). Our system will be a prototype of the same. We are not aiming to build a robot which can recognize a lot of words. Our basic idea is to develop some sort of menu driven control for our robot, where the menu is going to be voice driven. What we are aiming at is to control the robot using following voice commands. Robot which can do these basic tasks:-

- move forward
- Move back
- Turn right
- Turn left
- Load
- Release
- Stop (stops doing the current job)
- Speed control
- Obstacle detection

II. RELATED WORK

Voice enabled devices basically use the principal of speech recognition. It is the process of electronically converting a speech waveform (as the realization of a linguistic expression) into words (as a best-decoded sequence of linguistic units).

Converting a speech waveform into a sequence of words involves several essential steps:

1. A microphone picks up the signal of the speech to be recognized and converts it into an electrical signal. A modern speech recognition system also requires that the electrical signal be represented digitally by means of an analog-to-digital (A/D) conversion process, so that it can be processed with a digital computer or a microprocessor.

2. This speech signal is then analyzed (in the analysis block) to produce a representation consisting of salient features of the speech. The most prevalent feature of speech is derived from its short-time spectrum, measured successively over short-time windows of length 20–30 milliseconds overlapping at intervals of 10–20ms.Each short-time spectrum is transformed into a feature vector, and the temporal sequence of such feature vectors thus forms a speech pattern.

3. The speech pattern is then compared to a store of phoneme patterns or models through a dynamic programming process in order to generate a hypothesis (or a number of hypotheses) of the phonemic unit sequence. (A phoneme is a basic unit of speech and a phoneme model is a succinct representation of the signal that corresponds to a phoneme, usually embedded in an utterance.) A speech signal inherently has substantial variations along many dimensions.

III. ALGORITHM

A HMM, in simple terms is a model which is used to model a system about which we know nothing except its input and output sequences. A number of HMM models are proposed in the literature [2] such as left right, cyclic etc. We assume the left right model (Bakis model) in this paper. We train the HMM so that it produces an output which closely matches the available output sequence. A HMM is characterized by

(i) N, the number of states in the model. Although the states are hidden, for many practical applications there is often some physical significance attached to the states or to sets of states of the model.
(ii) M, the number of distinct observation symbols per state, i.e., physical output of the system being modeled.
(iii) The state transition probability distribution A= {aij} where aij = P[q t+1 =j/qt =i], 1≤ i,j ≤N, aij can be greater than or equal to 0.
(iv) The observation symbol probability distribution in state j, B= {$b_j(k)$}, where b_j (k) = P[V_k at q_t = j] 1≤ j ≤N, 1≤ k≤N
(v) The initial state distribution S= {Si} where Si = P [q_t = i]1≤ i ≤N

The speech recognition problem is:

Given an observation sequence O= O0 O1 O2 . . . OT-1 where each Ot is data representing speech which has been sampled at fixed intervals, and a number of potential models M, each of which is a representation of a particular spoken utterance (e.g. word or sub-word unit), find the model M which best describes the observation sequence, in the sense that the probability P (M|O) is maximized (i.e. the probability that M is the best model given O).

IV. CIRCUIT DIAGARAM

Fig.1 resembles the simple block diagram of the system. A voice recognition is used as the transmitter. The receiver is the robotic car that includes a, microcontroller and motor drivers, RF module**.**

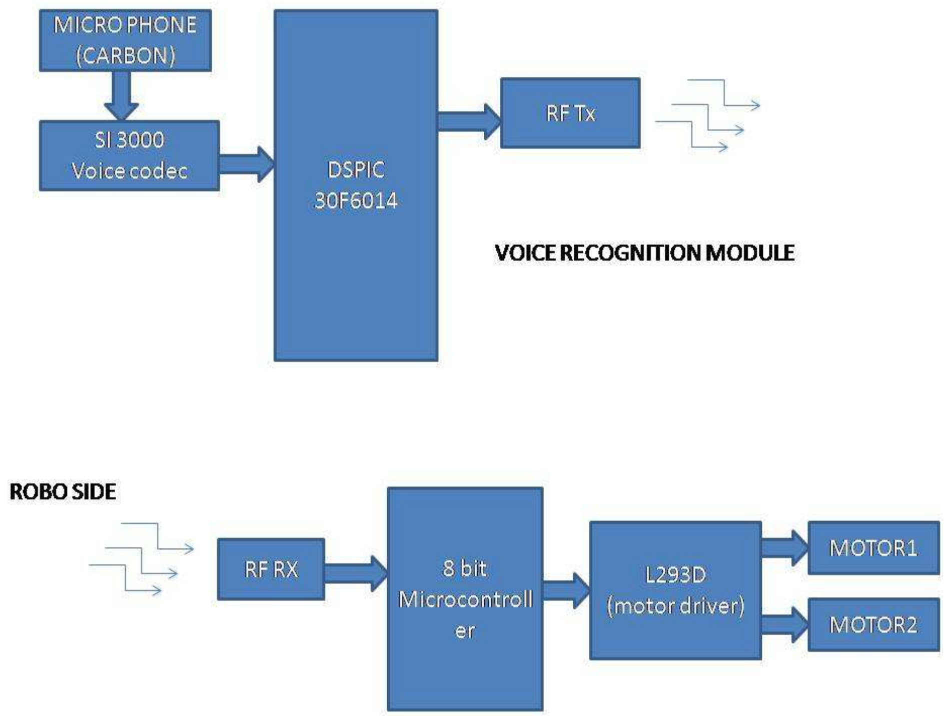


Fig1: Block Diagram

V. MOTOR DRIVER CIRCUIT

The L293D (Fig. 3.2) is a quad, high-current, half-H driver designed to provide bidirectional drive currents of up to 600 mA at voltages from 4.5V to 36V. It makes it easier to drive the DC motors

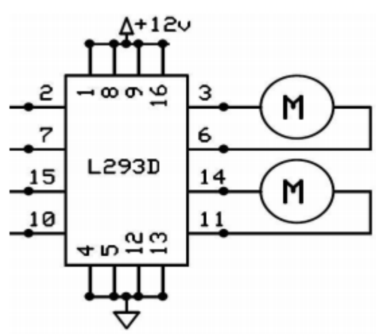


Fig2: Motor Diver

The L293D consists of four drivers. Pins IN1 through IN4 and OUT1 through OUT4 are input and output pins, respectively, of driver 1 through driver 4. Drivers 1 and 2, and drivers 3 and 4 are enabled by enable pin 1 (EN1) and pin 9 (EN2), respectively. When enable input EN1 (pin 1) is high, drivers 1 and 2 are enabled and the outputs corresponding to their inputs are active. Similarly, enable input EN2 (pin 9) enables drivers 3 and 4 [2], [3].

VI. dsPIC30F SPEECH RECOGNITION

The dsPIC30F Speech Recognition Library provides voice control of embedded applications that require an alternative user interface. With a vocabulary of up to 100 words, the Speech Recognition Library allows users to control their application vocally. The Speech Recognition Library is an ideal front end for hands-free products such as modern appliances, security panels and cell phones. The Speech Recognition Library has very modest memory and processing requirements and is targeted for the dsPIC30F5011, dsPIC30F5013, dsPIC30F5012 and dsPIC30F5014 processor. Items discussed in this paper are:

• Overview of the dsPIC30F Speech Recognition Library

• Speech Recognition Process Flow

Overview of the dsPIC30F Speech Recognition Library:

The dsPIC30F Speech Recognition Library provides an audio interface to a user's application program, allowing the user to control the application by uttering discrete words that are contained in a predefined word library. The words chosen for the library are specifically relevant to the interaction between the application program and the user. Upon recognition of a word, the application program takes an appropriate action, as shown in Figure

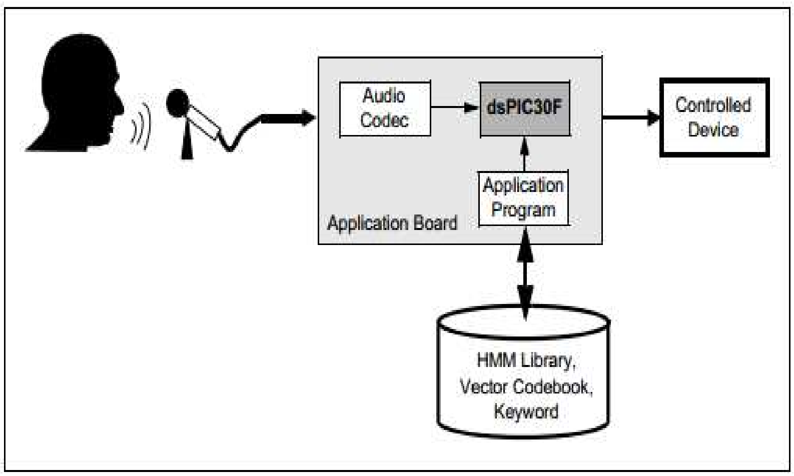
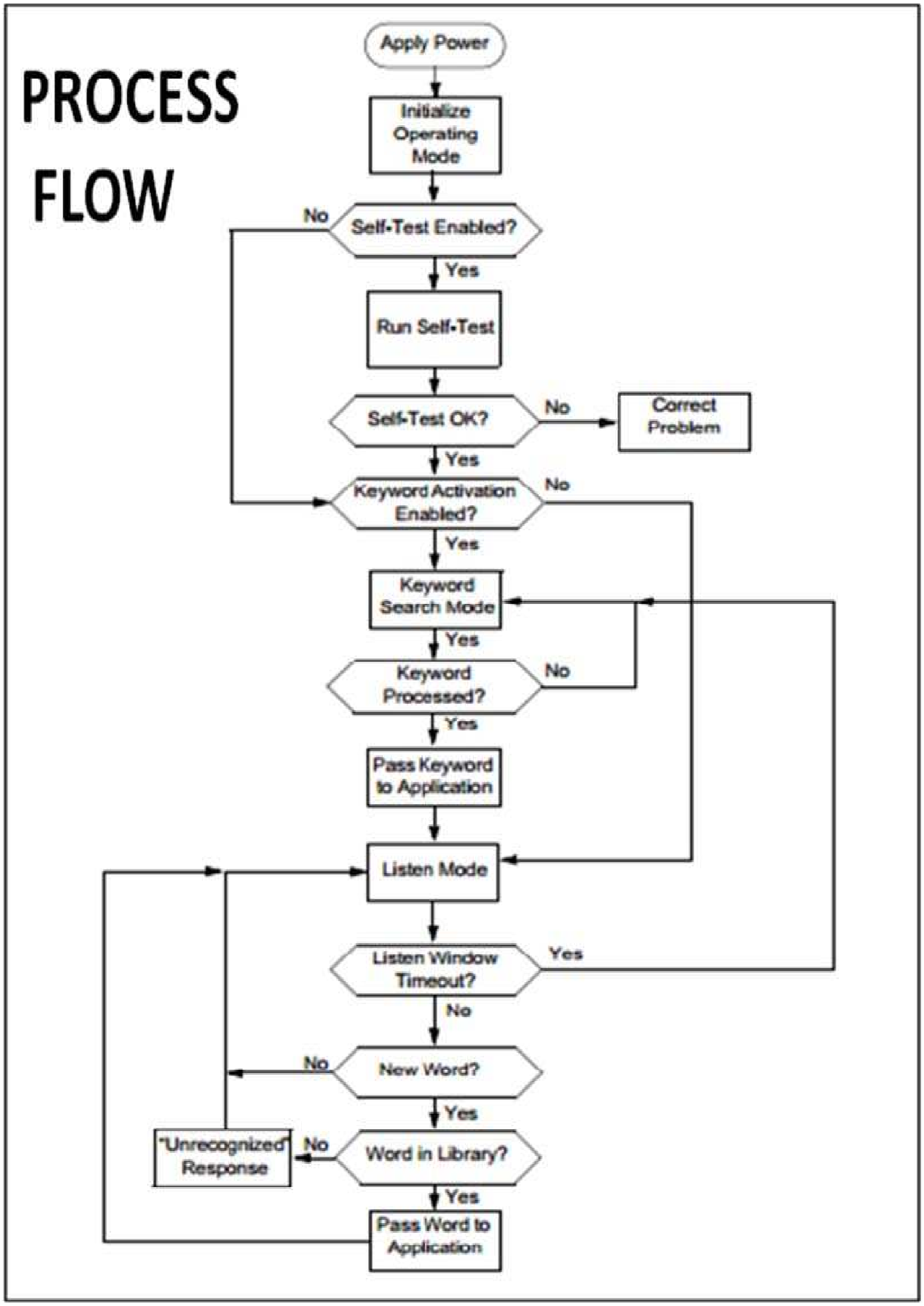


Fig3: Speech Recognition Module

Overview of speech recognition

The dsPIC30F Speech Recognition Library uses a recognition algorithm based on discrete Hidden Markov Model (HMM) of words (one HMM model for each word in an application word library). A word spoken through a microphone connected to the dsPIC30F application board is analyzed on a frame-by-frame basis using RASTA-PLP- algorithm and quantized into feature vectors of sound characteristics against a vector codebook. The quantized feature vectors are then examined to determine what word HMM model they most closely match. The dsPIC30F Speech Recognition Library can operate with a word library of up to 100 words. The word library is built around a keyword that is readily interpreted. Depending on the operating mode used, this keyword can be used to self-test the library and to trigger a recognition session. Successful recognition requires the words to be separated by a pause of at least one-half second but less than some specified period (normally programmed for five seconds). After a pause that times out, a new recognition session must be started. Optionally, the operating mode can be set to disable self-testing and/or keyword activation. When keyword activation is disabled, there is no timeout. Words must only be separated by at least 500 milliseconds



VII. CONCLUSION

Speech is the primary, and the most convenient means of communication between people. Whether due to technological curiosity to build machines that mimic humans or desire to automate work with machines, research in speech and speaker recognition, as a first step toward natural human-machine communication, has attracted much enthusiasm over the past five decades.

We have also encountered a number of practical limitations which hinder a widespread deployment of application and services. In most speech recognition tasks, human subjects produce one to two orders of magnitude less errors than machines. There is now increasing interest in finding ways to bridge such a performance gap. What we know about human speech processing is very limited.

Although these areas of investigations are important the significant advances will come from studies in acousticphonetics, speech perception, linguistics, and psychoacoustics.

Future systems need to have an efficient way of representing, storing, and retrieving knowledge required for natural conversation. This paper attempts to provide a comprehensive survey of research on speech recognition and to provide some year wise progress to this date. Although significant progress has been made in the last two decades, there is still work to be done, and we believe that a robust speech recognition system should be effective under full variation in: environmental conditions, speaker variability etc.

Speech Recognition is a challenging and interesting problem in and of itself. We have attempted in this paper to provide a comprehensive cursory, look and review of how much speech recognition technology progressed in the last 60 years.

Speech recognition is one of the most integrating areas of machine intelligence, since, humans do a daily activity of speech recognition. Speech recognition has attracted scientists as an important discipline and has created a technological impact on society and is expected to flourish further in this area of human machine interaction.

REFERENCES

1. Sadaoki Furui, *50 years of Progress in speech and Speaker Recognition Research, ECTI Transactions on Computer and Information Technology*,Vol.1. No.2 November 2005.

2. W. M. Campbell_, D. E. Sturim W. Shen D. A. Reynolds_, J. Navr´atily *"The MIT- LL/IBM Speaker recognition System using High performance reduced Complexity recognition"* MIT Lincoln Laboratory IBM 2006.

3. Zahi N.Karam,William M.Campbell *"A new Kernel for SVM MIIR based Speaker recognition* "MIT Lincoln Laboratory, Lexington, MA, USA.

4. Asghar .Taheri, Mohammad Reza Trihi et.al, *Fuzzy Hidden Markov Models for speech recognition on based FEM Algorithm*, Transaction on engineering Computing and Technology V4 February 2005,IISN,1305-5313

5. GIN-DER WU AND YING LEI *"A Register Array based Low power FFT Processor for speech recognition"* Department of Electrical engineering national Chi Nan university Puli ,545 Taiwan

6. Nicolás Morales1, John H. L. Hansen2 and Doorstep T.Toledano1 *"MFCC Compensation for improved recognition filtered and band limited speech"* Center for SpokenLanguage Research, University of Colorado at Boulder, Boulder (CO), USA

7. M.A.Anusuya , S.K.Katti *"Speech Recognition by Machine: A Review"* International journal of computer science and Information Security 2009.

8. Goutam Saha, Ulla S. Yadhunandan *" Modifield MelFrequency Cepstral coefficient"* Department of Electronics and Electrical communication Engineering India, Institute of International Journal of Computer Applications (0975 – 8887)Volume 10– No.3, November 2010

9. Technology ,Kharagpur Kharagpur-721302 West Bengal,India.