Embedded Vehicle Control System based on Voice Processing using DSPIC

Pradeep L. Yadav
EXTC Department
Dwarkadas J. Sanghvi College of Engineering
Vile-Parle (West), Mumbai

Sanjay B.Deshmukh
EXTC Department
Dwarkadas J. Sanghvi College of Engineering
Vile-Parle (West), Mumbai

ABSTRACT
The paper reports one Microcontroller car which processes DSPIC30F forming Speech Recognition System. The Speech Recognition system not only has high recognizable veracity, small volume, economy-power consumption, lower cost, high operation speed and real-time speech recognition. Furthermore, the function of speech cue offers a favorable interface for human-computer interaction in the system. These characteristics embody fully predominance of the embedded speech recognition**.** 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.

1. 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. The research of this paper extends the above NC (noise classification) based framework by additionally taking the SNR-range of the individual noise types into account. The architecture of the NC technique is illustrated in Fig.1 where the Hidden Markov Model Database (HMD) contains number of HMM model sets each trained on data corresponding to a known noise type and a chosen SNR value. The NC and the SNR estimator shown in the figure estimate the noise type and the SNR value of the noise contaminated input signal, respectively. An HMM model set is then selected from the HMD and used for recognition. With this more detailed partitioning of the training database the PDF's of each HMM set have less variance, and a better model discriminability is expected.

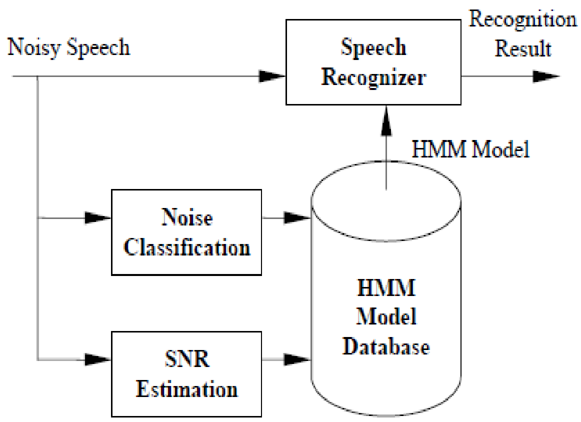

Fig.1: NC framework architecture

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

2. SPEECH CONTROL USING HMM

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

a. 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.

b. M, the number of distinct observation symbols per state, i.e., physical output of the system being modeled.
c. 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.
d. 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
e. 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).
For training the HMM for multiple speakers, the HMM parameters corresponding to each speaker is averaged. Compared to Rabiners's [2] approach, this has a number of advantages such as lower data requirement, higher detection accuracy and lesser computation complexity. Feature extraction, training and other pre processing stages of HMM are implemented in software (C/C++) in the offline mode and the recognition process is implemented in the SOC as online process. For the recognition, viterbi decoder is implemented in hardware.

3. PROPSED EXPERIMENTAL SCENARIO

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**.**

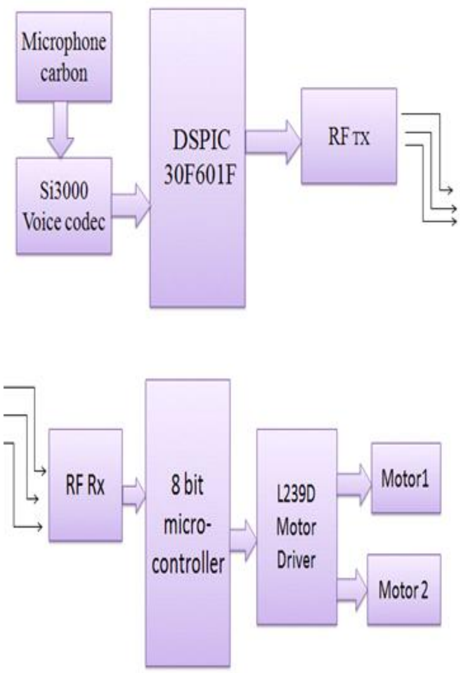

Fig.2: Block diagram of Speech Recognition Module

3.1 Motor driver circuit

The L293D Fig. 3 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

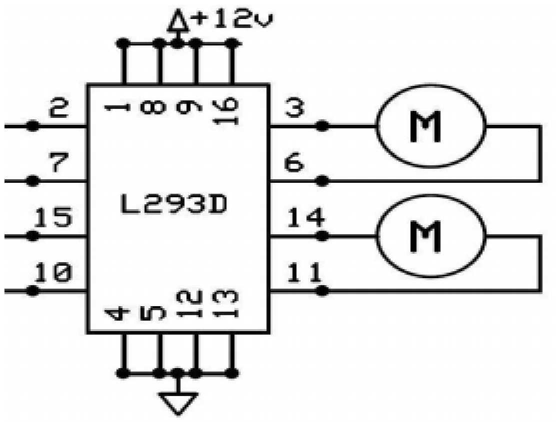

Fig.3: motor driver

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].

3.2 Speech recognition module

The fig. below is the block diagram of speech recognition module

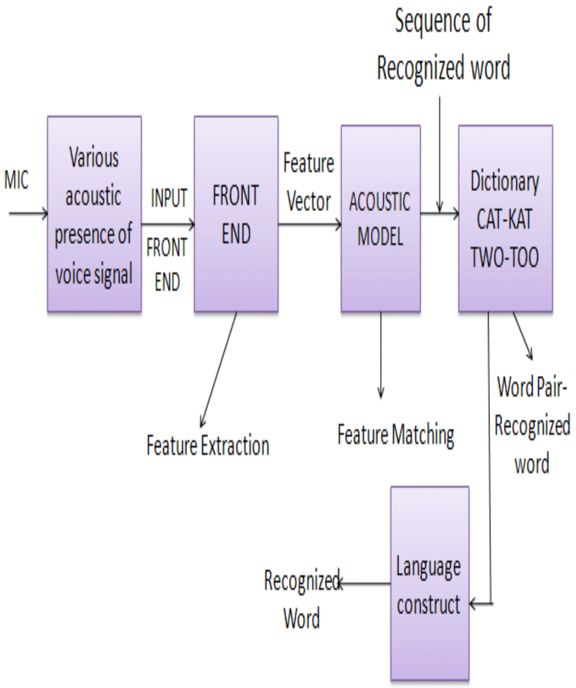

Fig.4: speech Recognition model

Speech recognition consists of two main modules, feature extraction and feature matching. The purpose of feature extraction module is to convert speech waveform to some type of representation for further analysis and processing, this extracted information is known as feature vector. The process of converting voice signal to feature vector is done by signal-processing front end module. As shown in above block diagram input to front-end is noise free voice sample and output of it is feature vector. In feature matching, the extracted feature vector from unknown voice sample is scored against acoustic model, the model with max score wins, and its output is considered as recognized word.

4. SPEECH ENHANCEMENT

Speech enhancement techniques tend to suppress the noise which corrupts the speech signal. Besides the methods using several microphones, many different types of speech enhancement systems using a single microphone have been proposed and tested [4]. All these systems are based on techniques which intend to recover the clean speech signal by enhancing the signal-to noise ratio. The performances depend upon the type of noise which corrupts speech and the information they require about noise. It should be noted at this point that the increase of SNR will improve the quality of the speech signal but not always its intelligibility. Therefore, as far as speech recognition is concerned, a trade-off has to be found between SNR improvement and recognition accuracy. Four main types of methods are used for speech enhancement:

4.1 Noise subtraction: This is very common method which assumes that noise and speech are uncorrelated and additive. In the spectral subtraction approach, the power spectrum of cleaned speech is obtained by subtracting the noise power spectrum from the spectrum of noisy speech . The method assumes that the noise varies slowly so that the noise estimation obtained during a pause can be used for suppression. Obtaining a good estimate of the noise spectrum is obviously the most difficult part of the method. Non-linear spectral subtraction has been proposed in order to overcome some limitations of basic subtraction. It basically consists in overestimating the noise spectrum, either in a uniform way, or else based on the perceptual evidence that the ear is more sensitive to the peaks of a power spectrum than to the valleys and that noise in the frequency regions of the valleys contributes the most to perceptual distortions.

4.2 Filtering: Traditional adaptive filtering techniques like Wiener or Kalman filtering have been used for speech enhancement, but more for speech transmission than for recognition purposes. The Wiener filter yields an optimal solution to the adaptive filtering problem in the sense of least mean square error. It necessitates the estimation of some parameters of the noise. Unless the noise is stationary and perfectly known that must usually be done iteratively. A recursive optimal estimation can be obtained with a Kalman filter**.** Kalman filtering can also be used with a colored noise assumption instead of white noise.
Experimental results show an improvement in intelligibility even though no automatic recognition was carried out.

4.3 Use of Markov models*:* hidden Markov models (HMM) decomposition is a method which makes it possible to separate speech from additive noise. The recognition of a noisy utterance can therefore be carried out by extending the classical Viterbi decoding algorithm to a search in the state space defined by the two models. This method is rather computationally demanding, but it has been demonstrated to perform satisfactorily even in bad SNR conditions.

4.4 Speech mapping: speech enhancement can be viewed as the process of transforming noisy speech into clean speech by some kind of mapping. For instance, spectral mapping has been implemented by a set of rules obtained by vector quantization techniques. It is also possible to implement arbitrarily complex space transformations thanks to connectionist neural networks. Even simple models such as multi-layer perceptions have been trained on learning samples to realize a mapping of noisy signals to noise-free speech which has been tested with success in an auditory preference test with human listeners.

5. EXPERIMENTAL RESULTS

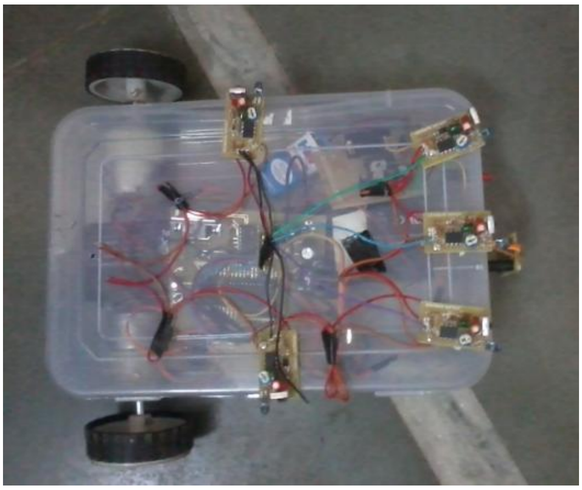

Fig4: Robotic car develop for voice recognition module along with Obstacle sensors

Figure above shows a robotic car which is controlled by a dsPIC30F speech recognition module. There are obstacle sensors placed on robotic car for obstacle avoidance. TSHOP 1738 will be used for detection of reflected infrared signals. Based on the information from six different sensors, we will control the robo car. Also we will use indication light for left right indication to the fellow human controlled cars.

6. 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 acoustic phonetics, 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 s 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.

7. REFERENCES

[1] Sadaoki Furui, 50 years of Progress in speech and

[2] Speaker Recognition Research, ECTI Transactions on Computer and Information Technology,Vol.1. No.2 November 2005.

[3] 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.

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

[5] 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

[6] 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

[7] 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

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

[9] 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