INTERNATIONAL JOURNAL OF COMPUTATIONAL MATHEMATICAL IDEAS
ISSN:0974-8652 / www.ijcmi.webs.com / VOL 3-NO2-PP 92-98 (2011)

Speech Recognition of Industrial Robot

D RAMBABU[*], R.NAGA RAJU[#], B.VENKATESH[@]
[*]II/II M Tech, Sasi Institute of Engineering & Technology, Tadepalligudem, Andhra Pradesh
Assistant Professor, Sasi Institute of Engineering & Technology, Tadepalligudem, Andhra Pradesh
[@] Assistant Professor, Sri Venkateswara Institute of science & Information Technology, Tadepalligudem, Andhra Pradesh.

E-mail:*rambabu.visit@gmail.com,[#] rnr808@gmail.com,[@]venkatesh.visit@gmail.com

ABSTRACT

This paper describes about the speech recognition system that was implemented on speech processor to increase the performance and is designed to recognize the word used as the command for controlling the movement of Industrial Robot. Basically there are two approaches used to recognize the speech signal. Here the used approach works on time domain and is known as Hidden Markov Model. It is built as both reference model and for recognition of words. The Robot moves in accordance with the voice commands like forward, reverse, Left, Right, pick, place and Stop. Experiments were done with several numbers of samples and it was found that maximum recognition rate was achieved with this novel method.

Keywords: Processor, PIC 16F877 Microcontroller, HMM, DC motors.

I. INTRODUCTION

Automatic speech recognition by machine has been a goal of a research for a long time. The first speech recognizer appeared in 1952 and consisted of a device for the recognition of single spoken digits. Speech recognition is the process of converting an acoustic signal, captured by a microphone or a telephone, to a set of words. The recognized words can be the final results, as for applications such as commands & control, data entry, and document preparation. They can also serve as the input to further linguistic processing in order to achieve speech understanding. There are many works carried out in this area. Initially speech recognition was implemented on Digital Signal Processors. Later it was implemented on Microcontrollers to decrease the cost.

The speech recognition system has also been implemented on some particular devices. Some of them are personal computer (PC), digital signal processor, and another kind of single chip integrated circuit. A framework to address the quantization issues which arise in fixed-point isolated word recognition was introduced in [2]. The system was developed using C++ language which is implemented on a PC. Reference [3] describes a speech recognition system using SPHINX-II, an off-the-shelf speech recognition package [4]. In reference [5] and [6], speech recognition system has been tried to be implemented on a FPGA and an ASIC.

This paper introduces the speech recognition which was implemented on a Processor. The processor where the speech recognition was implemented on is HM2007.

The speech recognition system that is implemented on the Processor is used to recognize the word in a speech signal. The words are used as the command for controlling movement of a Robot. Therefore, the system was designed to recognize limited number of the words. This can be extended by expanding the external memory of Processor.

There are only seven words used as the command for controlling movement of the Industrial Robot. They are forward, reverse, left, right, pick, place and stop, which are used to stop the Industrial Robot, to move forward the Industrial Robot, to move reverse the Industrial Robot, to turn left the Industrial Robot, and to turn right the Industrial Robot respectively. Basically there are two approaches to perform the speech recognition. The first approach is Linear Predictive Coding (LPC) and Euclidean Squared Distance (ESD). The second approach applied in this system is Hidden Markov Model (HMM), which is one of the speech recognition approaches that classified as the statistical pattern recognition. HMM is used as the

recognition method. As the feature extraction method, a simple segmentation and centroid value is applied. Here the used approach was HMM.

Section 2 and 3 of this paper describes about the design approach of speech recognition of industrial Robot more detail.

II DESIGN APPROACH

This section focuses on implementation of speech recognition of industrial Robot which includes hardware approach and software approach. The various stages in software development process viz., requirements, design, implementation, integration and an insight into the hardware design of a Robot are discussed.

A. Hardware Implementation

Three main components of Voice recognition systems are user interface, Processor and Robot. The Processor handles the communication between the user interface and Robot. It is controlled with the C programming proprietary libraries which allows control of the data. The data flow from the voice recognition components and processor is unidirectional, as the data can only be read from the voice recognition components.

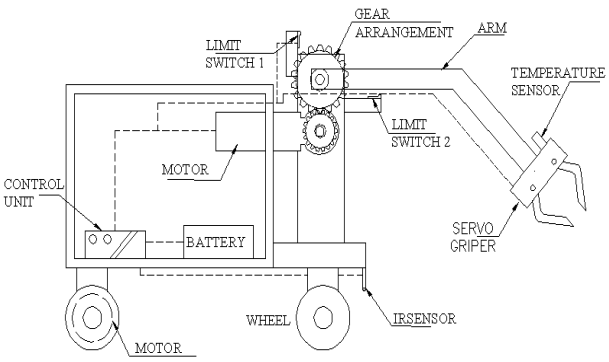


Fig. 1 Architecture Level Design

As a result, the processor must perform the functions like data transmission from the user interface, decoding of data signals and manipulation of the decoded signals.

Major parts (blocks) used to implement the hardware are:

- Microphone
- Filter & amplifier
- Analog to Digital convertor
- Processor
- Memory
- Comparator
- Control signal generator
- PIC Microcontroller
- Object (Robot)

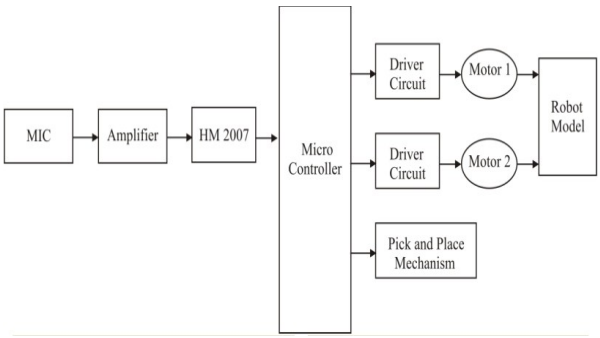


Fig 2 : Block Diagram of Industrial Robot

Initially the sample word passes through filters and amplifiers. The Basic algorithm code checks the ADC input at a rate of 4 KHz. If the value of the ADC is greater than the threshold value it is interpreted as the beginning of a half a second long word. The words to be matched are stored as template in a dictionary so that sampled word template can be compared against actual, later.

Once a template is generated from a sample word it is compared against the dictionary template using probability calculation and finds the template in the dictionary that is the closest match. Based on the word that matched the best the program sends a control signal to the Robot to perform basic operations like forward, reverse, left, right, pick, place and stop.

At start up as part of the initialization the program reads the ADC input using timer counter 0 and accumulates its value 256 times. By interpreting the read in ADC value as a number between 1 to 1/256, in fixed point, and accumulating 256 times. The average value of ADC was calculated without doing a multiply or divide. Three average values are taken each with a 16.4msec delay between the samples. After receiving three average values, the threshold value is to be four times the value of the median number. The threshold value is useful to detect when a word has been spoken or not.

B. Hardware Details

Actual components used to implement the hardware are

- Microphone
- Speech processor
- PIC 16F877 Microcontroller
- HY 6264
- IC7430
- L298N

❖ DC Motors

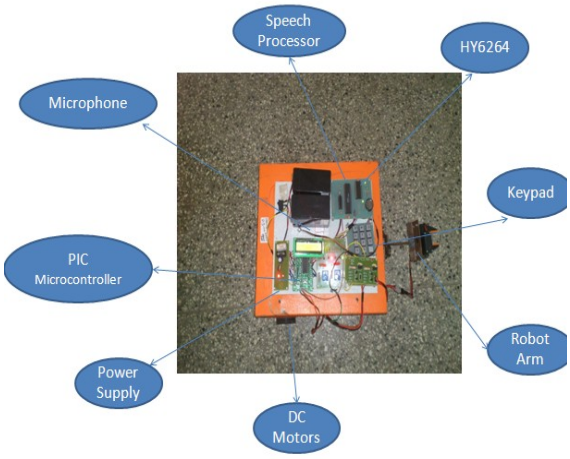


Fig. 3 Components used in Hardware

A microphone is an acoustic-to-electric transducer or sensor that converts sound into an electrical signal. Condenser microphones span the range from telephone transmitters through inexpensive microphones to high-fidelity recording microphones. They generally produce a high-quality audio signal and are now the popular choice in laboratory and studio recording applications.

Speech Processor is a single CMOS LSI circuit that analyzes the analog signal obtained from the microphone. Speech signals are captured with a microphone which are filtered and converted into a digital signal by an analog filter. The speech signal must be filtered to remove any frequencies outside the range of normal speech. Filtering certain frequencies also reduces the bandwidth of the speech signal resulting in less required computation power. Depending on the operation of the Speech Processor (training or recognition mode), data is either written to or read from the SRAM. The SRAM is divided into data banks where each data bank has its own unique 8 bit binary value. If the Speech Processor is in recognition mode, the Speech Processor will attempt to match the speech signal with the entries in the SRAM and return the corresponding data bank's 8 bit binary value on its data bus. If no match is detected, a reserved 8 bit value is transmitted.

When the Speech Processor matches the received speech signal to a specific stored phrase of word, the corresponding 8 bit value will be constantly held on the eight pins of the data bus. The 8 bit value will only change when the Speech Processor recognizes a different speech signal. Therefore, the voice recognition components will always transmit a signal of the last speech signal recognized.

Microcontroller is the heart of the project which is used store and controls the operations of robot. In this project AT89C51 microcontroller was used. It is a 40 pin, 8 bit controller manufactured by Atmel group. The advantage of this microcontroller is it has flash type reprogrammable memory with this user can erase the program within few minutes.

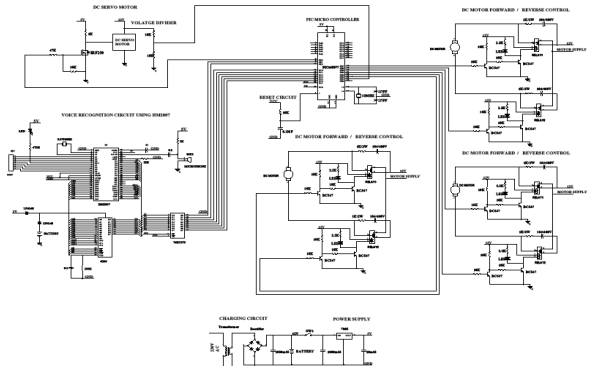


Fig 2.1 Circuit diagram of Robot

The HY6264 has a retention mode that guarantees data to remain valid at a minimum power supply voltage of 2.0V. Using CMOS technology, supply voltages from 2.0V to 5.5V have little effect on supply current in data retention mode. Reducing the supply voltage to minimize current drain is unnecessary with the HY6264 family.

The L298 is an integrated monolithic circuit in a 15-lead Multi watt and Power SO20 packages. It is a high voltage, high current dual full-bridge driver designed to accept standard TTL logic levels and drive inductive loads such as relays, solenoids, DC and stepping motors. Two enable inputs are provided to enable or disable the device independently of the input signals. The emitters of the lower transistors of each bridge are connected together and the corresponding external terminal can be used for the connection of an external sensing resistor.

Two rear wheels which are actuated by two DC motor with gearbox. One DC motor drives one rear wheel. Diameter of front wheel is 10 cm and diameter of rear wheel is 22 cm. Specification of the DC motor is 12 V, 2 A, and 200 rpm. Maximum linear speed of the wheelchair is about 0.461m/s or about 1.66 km/hr. The first part is DC motor control circuit. The circuit consists of controller, diver, and DC motor speed sensor circuit. They are forward, reverse, left, right, pick, place and stop, which are used to stop the Industrial Robot, to move forward the Industrial Robot, to move reverse the Industrial Robot, to turn left the Industrial Robot, and to turn right the Industrial Robot respectively.

III HIDDEN MARKOV MODEL

HMM is doubly embedded stochastic process that is not observable (it is hidden), but can only be observed through another set of stochastic processes that produce the sequence of observations.

Modern speech recognition systems use various combinations of a number of standard techniques in order to improve results over the basic approach described above. A typical large-vocabulary system would need context dependency for the phonemes. So phonemes with different left and right context have different realizations as HMM states. Decoding of the speech would probably use the Viterbi algorithm to find the best path, and here there is a choice between dynamically creating a combination Hidden Markov Model which includes both the acoustic and language model information.

A Elements of HMM

An HMM is characterized by the following

- 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 set of states of the model.
  $S= \{S_1, S_2, S_3, \ldots, S_N\}$
  State at time t is q_t states are interconnected.
- M, the number of distinct observation symbols per state, i.e., the discrete alphabet size.

$V=\{V_1,V_2,V_3\ldots..V_M\}$ The observation symbols correspond to the physical output of the system being modeled.

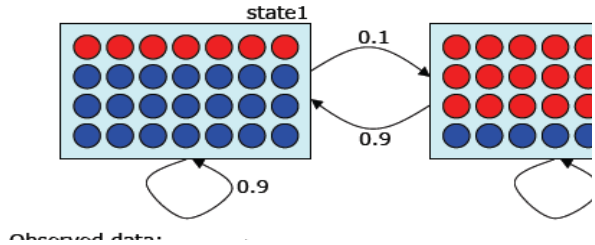
- The state transition probability distribution
  $A= \{a_{ij}\}$ where$a_{ij} = P[q_{t+1} = S_j \mid q_t = S_i]$,
  $1 \le i,j \le N$
- The observation symbol probability distribution in state j, $B=\{b_j(k)\}$
  Where $b_j (k) = P [v_k \text{ at } t \mid q_t = S_j]$, $1 \le j \le N$, $1 \le k \le M$
- Initial state distribution $\pi = \{\pi_i\}$
  $\pi_i = p[q_1 = S_i]$, $1 \le i \le N$

B Procedure to generate an observation sequence:

Given the appropriate values of N, M, A, B, and π, the HMM can be used to generate an observation sequence $O= O_1O_2\ldots O_T$ as follows

- Choose an initial state $q_1=S_i$ according to the initial state distribution π.
- Set t =1
- Choose $O_t= v_k$ according to the symbol probability distribution in state S_i, i.e., $b_i(k)$.
- Transit to a new state $q_{t+1}=S_j$ according to the state transition probability distribution for state S_i, i.e., a_{ij}.
- Set t=t+1; return to step 3 if t < T; otherwise terminate the procedure.

Consider the similar example with probabilities as given below



Observation probabilities in

State 1: Red → 0.25 [7/28 balls] and Blue→ 0.75 [21/28 balls]

State 2: Red → 0.75 [21/28 balls] and Blue→ 0.25 [7/28 balls]

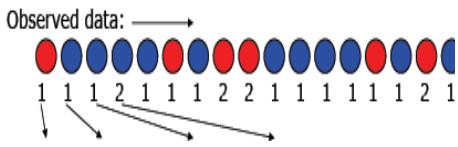


Fig. 4 An example of HMM

C. Soloutions to Basic Problems of HMM

In order to train an HMM, we must optimize ***a*** and ***b*** with respect to the HMM's likelihood of generating all of the output sequences in the training set, because this will maximize the HMM's chances of also correctly recognizing new data. Unfortunately this is a difficult problem; it has no closed form solution.

This general method is called *Estimation-Maximization* (EM). A popular instance of this general method is the *Forward-Backward Algorithm* (also known as the *Baum-Welch Algorithm*), which is describe now.

Previously defined $a_j(t)$ as the probability of generating the partial sequence and ending up in state *j* at time *t*. Now we define its mirror image, $b_j(t)$, as the probability of generating the remainder of the sequence, starting from state *j* at time *t*. $a_j(t)$ is called the *forward term*, while $b_j(t)$ is called the *backward term*. Like $a_j(t)$, $b_j(t)$ can be computed recursively, but this time in a backward direction.

$$\beta_j(t) = \sum_k a_{jk} b_k(y_{t+1}) \beta_k(t+1)$$

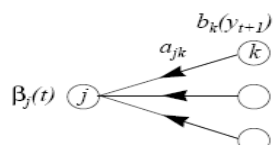


Fig. 5 The backward pass recursion

This recursion is initialized at time *T* by setting $b_k(T)$ to 1.0 for the final state, and 0.0 for all other states.

Forward Algorithm

Given input of length *N* observations

- Consider all possible paths (state sequences) of length *N* through HMM, and ending in its final state.
- Compute probability of each path (multiply together individual edge and local node probabilities)
- Sum all path probabilities

Initialization $\alpha_1(i) = \pi_i b_i(o_1)$

Recursion $\alpha_t(j) = \left[\sum_{i=1}^{N} \alpha_{t-1}(i) a_{ij}\right] b_j(o_t)$

Let us use $\alpha_t(j)$ to mean: the probability of observing the partial stream of observations $x_1, x_2, x_3 \ldots x_t$, and ending up at state *j*. It is the sum of the probabilities for all paths leading up to state *j*, while observing the partial sequence. Suppose define $\alpha_t(j)$ in terms of α_{t-1} (all predecessors of *j*), as in DP or DTW, and get an efficient solution.

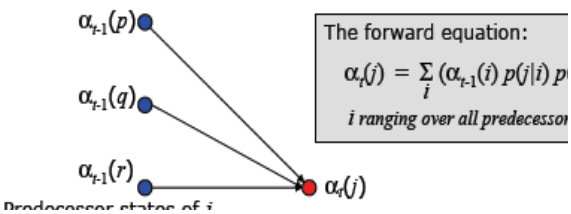


Fig. 6 An Example of calculations of HMM Forward Algorithm

Algorithmically a trellis is used to avoid exponential explosion of considering all possible paths. Hence, *P*(*X* | *word*) can be computed using a trellis, somewhat similar to DP and DTW

- The first difference: since this is a probabilistic model, the component edge and node probabilities are *multiplied*, rather being summed
- The second difference: Partial path likelihoods arriving at a node are *summed*, rather than the max (or min) being chosen.

D HMM Training System

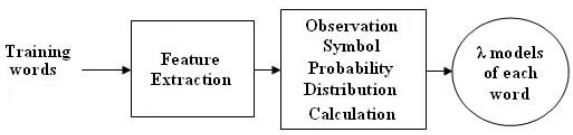


Fig.7 HMM training system

Parameters of HMM λ model consist of number of state (N), number of observation symbol (M), state transition probability distribution (A), observation symbol probability distribution (B), and initial state probability distribution (π).

E HMM Recognizer system

Firstly, the speech signal of an unknown word is converted to the observation sequence of that word. This is done by performing the feature extraction and vector quantization process. Then, the observation sequence probability of unknown word for a given model λ (P (**O** | λ)) is calculated for each HMM word model. The probability of observation sequence calculation is done by using the forward procedure method. By applying the parameter A and π of the designed HMM (16) and (17) to the forward procedure equations, the probability of the observation sequence, **O** = (O_1,O_2 … O_N), for the given model can be calculated by using the following equation:

$$P(\mathbf{O} \mid \lambda) = b_1(\mathbf{o}_1) \cdot b_2(\mathbf{o}_2) \cdot b_3(\mathbf{o}_3) \cdots b_N(\mathbf{o}_N)$$

Where b_N (O_N) is probability of N-th observation symbol (O_N) in state N.

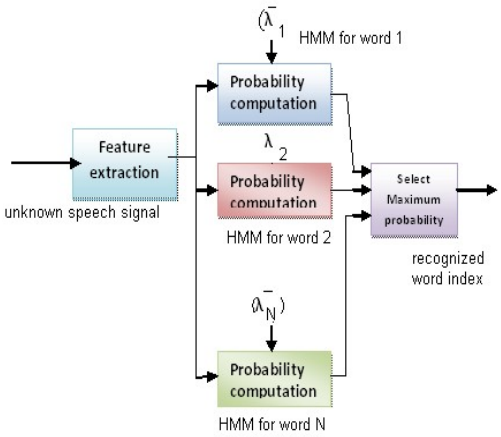


Fig.8 Recognition system of HMM

IV PROJECT RESULTS

The following figure shows the Robot controlled Module in top view and it consists of various components.

The following snap shot shows the voice controlled Robot in three dimensional views and below the bottom it has two motors.

The mechanism, used for industrial Robot arm was displayed in the below figure.

The following snapshot shows the Industrial Robot arm and its relevant connections

Whenever voice command forward is match with the reference word, then the word "Forward" will be displayed on the Screen. The following snapshot shows the command.

Whenever voice command Reverse is match with the reference word, then the word "Reverse" will be displayed on the Screen.

Whenever voice command Left is match with the reference word, then the word "Left" will be displayed on the Screen. The following snapshot shows the command.

Whenever voice command Right is match with the reference word, then the word "Right" will be displayed on the Screen. The following snapshot shows the command.

Whenever voice command Stop is match with the reference word, then the word "Stop" will be displayed on the Screen. The following snapshot shows the command.

V CONCLUSION

Speech recognition has successfully been implemented on Processor in order to control the movement of Robot. The approach of speech recognition that has been implemented is HMM. From the experimental results, it can be concluded that the response time is less than one second.

At the beginning of this work, there is a goal to recognize seven, at the end of project, seven words were recognized and Industrial Robot was respond to the voice command with in milliseconds.

REFERENCES

[1]. Treeumnuk, Dusadee. (2001). *Implementation of Speech Recognition on FPGA. Masters research study, Asian Institute of Technology, 2001).Bangkok: Asian Institute of Technology.*

[2]. Sriharuksa, Janwit. (2002). *An ASIC Design of Real Time Speech Recognition. (Masters research study, Asian Institute of Technology, 2002). Bangkok: Asian Institute of Technology*.

[3]. Y.M. Lam, M.W. Mak, and P.H.W. Leong, "Fixed point implementations of Speech Recognition Systems". *Proceedings of the International Signal processing conference*. Dallas. 2003.

[4]. Soshi Iba, Christiaan J. J. Paredis, and Pradeep K. Khosla. "Interactive Multimodal Robot Programming". *The International Journal of Robotics Research* (24), pp 83 104, 2005.

[5]. Thiang, "Implementation of Speech Recognition on MCS51Microcontroller for Controlling Wheelchair" *proceedings of International conference on Intelligent and advanced systems. Kuala Lumpur, Malaysia, 2007*

[6]. Thiang, "Limited speech recognition for controlling movement of Mobile Robot Implemented on ATmega162 Microcontroller" *proceedings on International conference on Computer and Automation Engineering.2009.*

[7]. Lawrence Rabiner, and Biing Hwang Juang, *Fundamentals of Speech Recognition. Prentice Hall, New Jersey, 1993*.

[8]. Speech recognition by machine. *By William Anthony Ainsworth, Institution of Electrical Engineers.*

[9]. www.speechrecognition.com -/ united states.

[10]. B.Yegnanarayana, "Artificial Neural Networks", Prentice-Hall of India, New Delhi, 1999.

[11]. Rabiner, L.R. and Schafer, R.W. (2005), *Digital Processing of Speech Signals*, Pearson Education

[12]. L. R. Rabiner, *A tutorial on hidden Markov models and selected applications in speech recognition,* Proc. IEEE, vol. 77, no. 2, pp. 257–286, 1989.

[13]. Daniel Jurafsky and James H. Martin (2002), *Speech and Language Processing,* Pearson Education Asia.

[14].Davies, K.H., Biddulph, R. and Balashek, S. (1952) *Automatic Speech Recognition of Spoken Digits*, J. Acoust. Soc. Am. **24**(6) pp.637 – 642.

[15].Managing editors: Giovanni Battista Varile, Antonio Zampolli. (1997). Cole, Annie; Mariani, Joseph; Uszkoreit, Hans et al. Eds. *Survey of the state of the art in human language technology*. Cambridge Studies In Natural Language Processing. XII–XIII. Cambridge University Press. ISBN 0-521-59277-1.

[16].Junqua, J.-C.; Haton, J.-P. (1995). *Robustness in Automatic Speech Recognition: Fundamentals and Applications*. Kluwer Academic Publishers. ISBN 978-0792396468.