

# Speech Recognition System For Controlling The Robot

<sup>1</sup>Ande Stanly Kumar, <sup>2</sup>Dr.K.Mallikarjuna Rao,  
<sup>3</sup>Dr.A.Bala Krishna.,

<sup>1</sup>Assoc.Professor,Sri Vani School of Engineering,Vijayawada.

<sup>2</sup>Professor,JNTU College of Engineering,Kakinada.

<sup>3</sup> Professor,SRKR Engineering College,Bhimavaram.

**Abstract:-** Automatic speech recognition by machine has been a goal of a research for a long time, which concurrent the inter disciplines like mechanical, electronics and coputer engineering. 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. 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. In this paper we propose voice recognition to control robot using finger print comparison by Euclidean square distance, band pass filters and java technology.

**Key words:** Concurrent Engineering, Euclidean square distance, LPC, Voice recognition, Finger print.

## 1. INTRODUCTION:

### 1.1 Voice Recognition

The term "voice recognition" is sometimes used to refer as speech recognition where the recognition system is trained to a particular speaker, hence there is an element of speaker recognition, which attempts to identify the person speaking, to better recognize what is being said. Speech recognition is a broad term which means it can recognize almost anybody's speech - such as a call-centre system designed to recognize many voices. Voice recognition is a system trained to a particular user, where it recognizes their speech based on their unique vocal sound.

### 1.2 Mechatronics:

Mechatronics basically refers to mechanical electrical systems and is centred on mechanics, electronics, computing and control which, combined, make possible the generation of simpler, more economical, reliable and versatile systems. The term "mechatronics" was first assigned by Mr. Tetsuro Mori, a senior engineer of the Japanese company Yaskawa, in 1969.

### 1.3 Embedded Systems :

A combination of hardware and software which together form a component of a Concurrent systems. An embedded system is designed to run on its own without human intervention, and may be required to respond to events in real time.

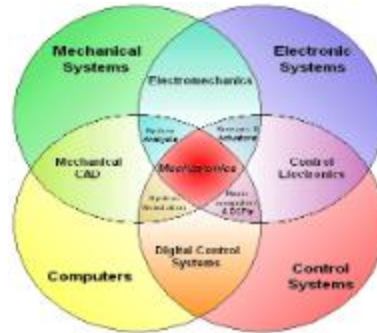


Fig : 1. Components of Concurrent Engineering.

## 2. LITERATURE:

Treumnuk & Dusadee, implemented [1] the Speech Recognition on FPGA with segmentation technique. Sriharuksa & Janwit implemented [2] a complete design and layout of an ASIC Design of Real Time Speech Recognition. They introduced a novel method for isolating the rove of higher order polynomials in Linear predictive systems. Y.M. Lam et al. implemented [3] fixed point implementations for speech recognition, they achieved recognition rate of 81.33%. SoshiIba et al. proposed [4] the framework takes a three-step approach to the robot programming i.e multi-modal recognition, intention interpretation, and prioritized task execution.

In previous works, speech recognition system was implemented [5] on ATMEL 89C51RC microcontroller to control the movement of Wheelchair. They were used the LPC model for speech recognition and achieved recognition rate of about 78.57%. Thiang implemented [6] the speech recognition for controlling movement of Mobile Robot on ATmega162 Microcontroller, with the techniques used by them were Linear Predictive Coding (LPC) combined with Euclidean Squared Distance and Hidden Markov Mode (HMM) and highest recognition rate achieved was 87%. Coming to this paper, it describes continuation work to the previous works.

## 3. SPEECH RECOGNITION SYSTEM FOR ROBOT:

Speech is a natural source of interface for human-machine communication, as well as being one of the most natural interfaces for human-human communication [8]. However, environmental robustness is still one of the main barriers to the wide use of speech recognition. Speech recognition performance degrades significantly under varying environmental conditions for many application areas.

In this paper, speech/voice recognition system is implemented to recognize the word used as the command for controlling the movement of robot. The proposed novel method will increase the recognition rate. Especially monitor the need of Embedded Systems in Industrial applications to control the movement of either simple or Bulky devices. There are two approaches used to recognize the speech signal. The first approach is Linear Predictive Coding combined with Euclidean Squared Distance (ESD). In this approach LPC is used as the feature extraction method and Euclidean Squared Distance is used as the recognition method. The second approach is Hidden Markov Model, which is used to build reference model of the words and also used as the recognition method. Feature extraction method used in the second approach is a simple segmentation and centroid value. Both approaches work on time domain. Experiments have to do in several variations of observation symbol number and number of samples. The robot can move in accordance with the voice command. Maximum recognition rate will be expected here by introducing a novel method.

### 3.2. Design of robot system:

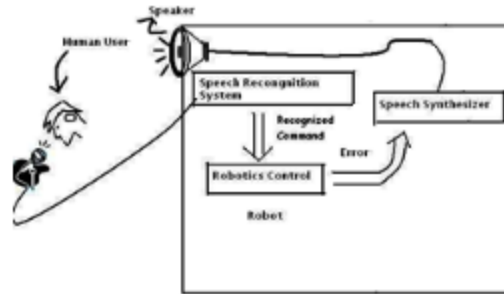


Fig. 2. Layout of robotic system

The layout of robotic system designed as shown in figure 2. The design had been done in the field of robotics and there exists a line follower robots, sensor robots and used speech to control a robot. It would make a robot which obeys human speech commands and performs errands.

#### 4. MATHEMATICAL EXPRESSION FOR SPEECH ANALYSIS:

In order to analyze speech, needed to look at the frequency content of the detected word. To do this is used several 4<sup>th</sup> order Chebyshev band pass filters. To create 4<sup>th</sup> order filters, and cascaded two second order filters using the following "Direct Form II Transposed" implementation of a difference equations.

$$y_1(n) = b_{11}x_1 + b_{12}x_1(n-1) + b_{13}x_1(n-2) - a_{11}y_1(n-1) - a_{12}y_1(n-2)$$

$$y_2(n) =$$

$$b_{21}x_2 + b_{22}x_2(n-1) + b_{23}x_2(n-2) - a_{21}y_2(n-1) - a_{22}y_2(n-2)$$

$$y_{out}(n) = g y_2(n)$$

Where the coefficient a's and b's were obtained through Matlab using the following commands.

$$[B,A] = \text{cheby2}(2,40,[\text{Freq1}, \text{Freq2}]);$$

(Where 2 defines a 4<sup>th</sup> order filter, 40 defines the stop band ripple in decibels, and Freq1 and Freq2 are the normalized cutoff frequencies).

$$[\text{sos2}, g2] = \text{tf2sos}(B2, A2, 'up', 'inf');$$

#### 4.1 Fingerprint Calculation:

Finger print calculation is needed a way to encode the relevant information of the spoken word. The relevant information for each word was encoded in a "fingerprint". Fingerprints are compared using the Euclidean distance formula between sampled word fingerprint and the stored fingerprints to find correct word.

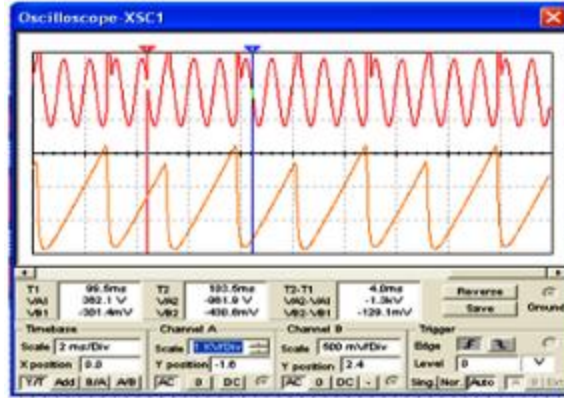


Fig 3 : Signal Response of Speech Amplifier

Euclidean distance formula is:

$$d = \sqrt{\sum_{i=1}^n (p_i - q_i)^2}$$

$P = (P_1, P_2, \dots, P_n)$  and  $Q = (Q_1, Q_2, \dots, Q_n)$

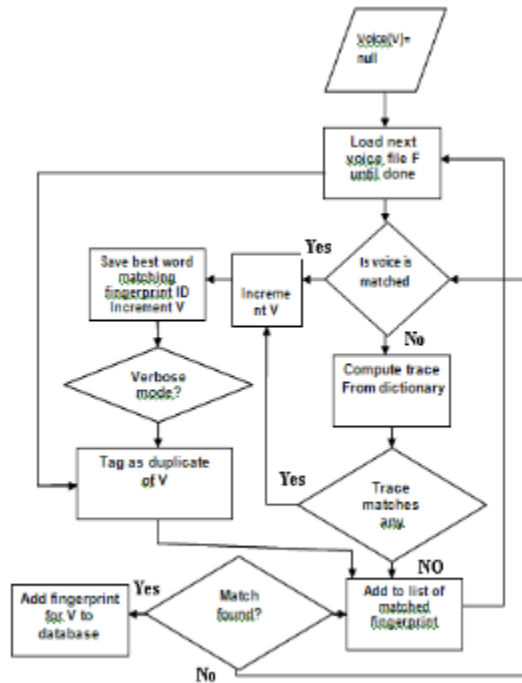


Fig 4 : Flow Chart for voice recognition

Where P is a dictionary fingerprint and Q is the sampled word fingerprint and  $p_i$  and  $q_i$  are the data points that make up the fingerprint. To see if two words are the same we compute the Euclidean distance between them and the words with the minimum distance are considered to be the same. The formula above requires squaring the difference between the two points, but since by using fixed point arithmetic, found that squaring the difference produced too

large of a number causing our variables to overflow. Thus implemented a "pseudo Euclidean distance calculation" by moving the sum out of the square root reducing the equation to

$$D = \sum_{i=1}^n |p_i - q_i|$$

## 5. HARDWARE/SOFTWARE OF ROBOT:

The signal coming out of the microphone needed to be amplified. We had two different versions of operational amplifier, LM741 and LM 741. The LM741 has a slew rate of 0.015 V/ $\mu$  s, on the other hand LM741 has 0.3V/ $\mu$  s. The LM741 has a better slew rate and it gave us better response to input signals so we used it when we designed our amplification circuit as shown in figure 3.

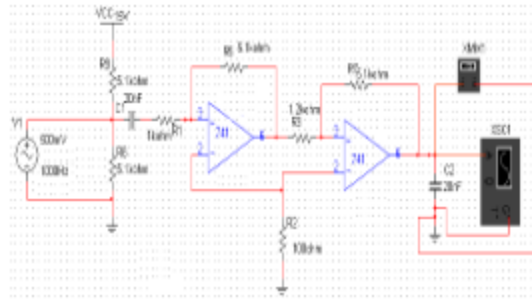


Fig 5 : Signal Response of Speech Amplifier

The signal processing of speech requires lot of computations, which implies required fast processor, but to operate at 16 M Hz. In order to minimize the number of cycles we used filtering the audio signal we had to write most of the code in assembly. We wrote all of 10 digital filters in assembly which made them very efficient and significantly improved our performance over a C/java code implementation.

Fig:5 .Circuit Diagram of Speech Amplifier

Fig:3. Signal Response of Speech Amplifier

Fig:4. Flow chart for voice recognition:

### Flow Chart of Speech Comparison:

The Basic algorithm of code was to check 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 sample word passes through 8 band pass filters and is converted into a fingerprint. The words to be matched are stored as fingerprints in a dictionary so that sampled word fingerprints can be compared against them later. Once a fingerprint is generated from a sample word it is compared against the dictionary fingerprints and using the modified Euclidean distance calculation finds the fingerprint in the dictionary that is the closest match. Based on the word that matched the best the program sends a PWM signal to the robot to perform basic operations like left, right, go, stop, or reverse.

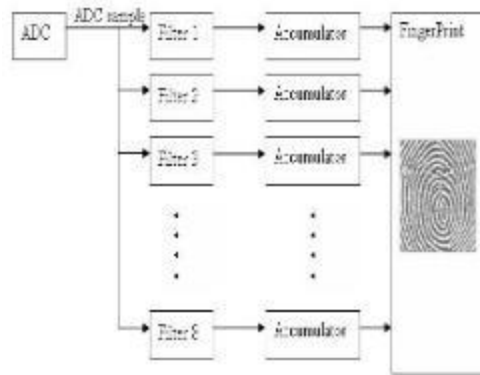


Fig.6. Finger Print Implementation

### 5.1: Initial-Threshold Calculation:

At start up as part of the initialization the program reads the ADC input using timercounter0 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.

### 5.2: PWM (Pulse Width Modulation) duty cycle calculation:

The motors in the robot were measured to have a 50 Hz PWM frequency and movement was controlled by varying the duty cycle from 5% to 10%. To generate the PWM signals we used timer/counter1 in phase correct mode. The top value of timer/counter 1 was set to be 20000 was set to have a frequency of  $50\text{Hz} = 16\text{MHz}/(8*2*20000)$ . To calculate the duty cycle the following equation was used  $\text{OCR1x} = (20000 - 40000*\text{duty cycle})$ . Where OCR1x is the value in the output compare register 1 A or B.

## 6. FINGERPRINT GENERATION:

The finger print can be done as shown in figure 6. The program considers a word detected if a sample value from the ADC is greater than the threshold value. Every sample of ADC is typecast to an int and stored in a dummy variable A in. The A in value passes through 8 4<sup>th</sup> order Chebyshev band pass filters with a 40 dB stop band for 2000 samples (half a second) once a word has been detected. When a filter is used its output is squared and that value is accumulated with the previous squares of the filter output. After 125 samples the accumulated value is stored as a data point in the fingerprint of that word. The accumulator is then cleared and the process is begun again. After 2000 samples 16 points have been generated from each filter, thus every sampled word is divided up into 16 parts. Our code is based around using 10 filters and since each one outputs 16 data points every fingerprint is made up of 160 data points.

### 6.1: Filter Implementation:

A 4<sup>th</sup> order Chebyshev filter with 40 dB stop band since it had very sharp transitions after the cutoff frequency and designed 10 filters a low pass with a cutoff of 200 Hz, a high pass with a cutoff of 1.8 KHz, and eight band passes that each had a 200 Hz bandwidth and were evenly distributed from 200Hz to 1.8 KHz. Thus we had band pass filters that went from 200-400 Hz, 400-600, 600 – 800 and so on all the way to the filter that covered 1.6 KHz – 1.8 KHz. Designed filters in this way because that most of the important frequency content in words was within the first 2 KHz since this usually contains the first and second speech formants, (resonant frequencies). This also allowed to sample at 4 KHz and gave almost enough time to implement 10 filters. Needed ten filters each with approximately a 200 Hz bandwidth so that would have enough frequency resolution to properly identify words. Originally had 5 filters that spanned from 0 – 4 KHz and were sampling at 8 KHz, but this scheme did not produce consistent word recognition.

## 6.2: Fingerprint Comparison:

Voice fingerprinting for identifying voice, either in streams or in files, and that these fingerprints be checked against a large database of previously-computed fingerprints. In this paper, explore two new voice recognition applications:

Duplicate detection and voice thumbnail generation. In duplicate detection, aim to identify duplicate voice files based only on the voice data, even if one is a noisy version of the other or if they have different durations. In voice thumbnail generation, the task is to find a representative section of the voice – a “thumbnail.”

Which converts a segment of voice to 64 floating-point numbers (a fingerprint). Its fingerprints are very robust to distortions of the original voice and the AFP lookup method uses a new technique that is about a factor of 50 faster than the fastest competing method. For each created fingerprint, a normalization factor is also created, so that the mean Euclidean distance from that fingerprint to a large collection of fingerprints computed from other voice is one. From voice, and “fingerprint” means a reference fingerprint against which traces are compared to determine the voice identity. The normalization factor is always associated with the fingerprint, so Euclidian distances  $D(\cdot, \cdot)$  between traces and fingerprints are normalized.

## 6.3: PWM signal to move ROBOT:

Once a word is recognized, its time to perform an action based on the recognized word. To perform an action of PWM signal generated using timercounter1. Control of the PWM signal generation is done by the robot control() function. For robot, generate two different PWM signals, one for moving the robot front/back and another one to steer left or right and also need to send a default PWM signal to pause a robot. Timercounter1 has two different compare registers, and can output two unique PWM signals. Use Phase correct mode to generate PWM signals because it is glitch free, which is better for the motor.

To find out a frequency and a duty cycles at which Robot turns forward/backward and left/right, and attached an oscilloscope probe to a robot’s receiver, sent different signals to the receiver using the robot remote control and measured the frequency and duty cycle for different motions.

## 7.0: CONCLUSION:

The Concurrent systems design covers a very wide range of microprocessor designs Our task is to design a control module for a robot. The robot is a simple two wheel robot that uses two stepper motors for driving. The robot can be programmed to drive autonomously a certain path. A list of driving commands are first downloaded from a PC to the robot, after which the robot will drive automatically through the program and to provides a framework to specify a system. At the beginning set a goal to recognize five words, to be recognized. However five words needed to be orthogonal to each other because filters were not giving a high enough resolution and inaccuracy in fingerprint calculations due to using fix point arithmetic made the lookup function to be error prone. As a result, it had to pick various different words that sound apart. If it had to do this again instead of trying to use the Euclidean distance formula to match words try to do perform a correlation of the two fingerprints. A correlation is less sensitive to amplitude differences and is a better way of identifying patterns between two objects. If it had faster process chip, it could modified algorithm to add more filters, perform Fourier transform, or floating point arithmetic in order to improve results.

## REFERENCES:

- [1]. **Thiang**, “Limited speech recognition for controlling movement of Mobile Robot Implemented on ATmega162 Microcontroller”, Proceedings on International conference on Computer and Automation Engineering.2009.
- [2]. **Thiang**, “Implementation of Speech Recognition on MCS51Microcontroller for Controlling Wheelchair”, Proceedings of International conference on Intelligent and advanced systems. Kuala Lumpur, Malaysia, 2007
- [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]. **Treumnuk, Dusadee**, “Implementation of Speech Recognition on FPGA”, Masters research study, Asian Institute of Technology, Bangkok 2001

- [5]. **Soshi Iba, Christiaan J. J. Paredis, and Pradeep K. Khosla.** “Interactive Multimodal Robot Programming”. The International Journal of Robotics Research, 2005, (24), 83 –104.
- [6]. **Sriharuksa, Janwit.** “An ASIC Design of Real Time Speech Recognition”, (Masters research study, Asian Institute of Technology, Bangkok, 2002.
- [7]. **Lawrence Rabiner, and Biing Hwang Juang,** “Fundamentals of Speech Recognition. Prentice Hall”, New Jersey, 1993. Speech recognition by machine. By William Anthony Ainsworth, Institution of Electrical Engineers.
- [8]. **Andre Harison and Chirag Shah,** “Voice recognition by robot”.
- [9]. [www.speechrecognition.com](http://www.speechrecognition.com) -/ united states.
- [10]. **Frank Vahid and Tony Givargis,** “Embedded System Design: A Unified Hardware/Software Approach”.