

Speech Recognition System For Robotic Control And Movement

Biraja Nalini Rout, Abhijeet Das, Sasi Rekha Sankar

Abstract: In a current scenario, voice and data recognition is one of the most sought after field in the area of artificial intelligence and robotic [1] engineering. The idea specializes on deriving a voice to voice intelligent system which operates purely on audio/voice instructions using a specialized voice recognition module, a micro controller, a set of wheels and a movable arm to operate. The working involves real time voice inputs feeded to the VR module which equivalently processes the audio signals and produces the output in audio format. It consists an IDE for both Windows and UNIX based operating system for manipulating and processing instructions both at software and hardware levels. The system also can perform a basic set of manual operations decides through the expert system. The VR module processes the data using multilayer perceptron to generate the required result. Movable arm operates to pick and place objects as per the given voice instructions. Its usability involves substituting manual work at both personal and professional levels.

Index Terms: Euclidian square Distance, LPE, HM2007, ATMEL 89c51

1. INTRODUCTION

The HRVI model is the stands for human robot voice interaction. This model is divided into two parts of mechanism one is for human-computer and other one is for computer-robot. Various regular expression techniques are used which is helpful in handling the exceptional condition while communication is going on between the sources. The first module is human-computer interaction through voice (VTT). Long distance communication possible through this interface and the respective operations are done by the system. The command-operation database is maintained with the additional operation of addition, deletion, updating through voice from direct client. The system can also give the voice response to the user. The second module is computer-robot [7] interaction through voice. There is no need of human interaction to give commands to the user. The user command will comethrough computer and checked by the voice chip (HM2007). There is also another database is maintained in the robot. The Hidden Markov rules is used to handle the exceptional commands through relative speech detection and rectification. The CMOS LSI single chip recognition is the core to handle the voice command. The commands given the chip after getting authenticated will be ready for execution module. The respective operations that are mentioned for that command will be executed. All the existing system will be taken care of in this model like phonetics, lexicology, syntax and pragmatics. Use filters to measure energy levels for various points on the frequency spectrum. Analogue to digital

conversion and sampling and quantizing will be carried through this level. Filtration process [5] is the added advantage for this system because the actual command should be separated from the background noise. Pitch, length and loudness are used to indicate "stress". According these parameter the useful steps will be taken.

2. HM 2007

HM2007 is a single chip CMOS voice recognition LSI circuit with the on-chip analog front end, voice analysis, recognition process and system control functions. A 40 isolated- word voice recognition system can be composed of external microphone, keyboard, 64kSRAM and some other components. Combined with the microprocessor, an intelligent recognition system can be built.

Features of:

- Single chip voice recognition CMOS LSI
- Speaker dependent
- External RAM support
- Maximum 40 word recognition (.96 second)
- Maximum word length 1.92 seconds (20 word)
- Microphone support
- Manual and CPU modes[4] available
- Response time less than 300 milliseconds
- 5V power supply
- 48 pin PDTP, 52 pin PLCC, 48 pad bare chip

3 registers are there

- Input buffer (write only)
- Output buffer (read only)
- Status buffer (read only)

3. Proposed Architecture

Automatic speech recognition by machine has been a goal of a research for a long time, which concurrent the inter discipline like mechanical, electronics and computer 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

- *BIRAJA NALINI ROUT, is currently pursuing undergraduate's degree program in software engineering in SRM University, INDIA, PH-9790711958. E-mail: biraja.nrsoft@gmail.com*
- *ABHIJEET DAS, is currently pursuing undergraduate's degree program in software engineering in SRM University, INDIA, PH-9789910847. E-mail: abhijeet.das4u@gmail.com*
- *Sasi Rekha Sankar, is an Assistant Professor of Department of Software Engineering in SRM University, INDIA, PH-9791994531, E-mail: sasirekha.s@ktr.srmuniv.ac.in*

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 [5] and java technology.

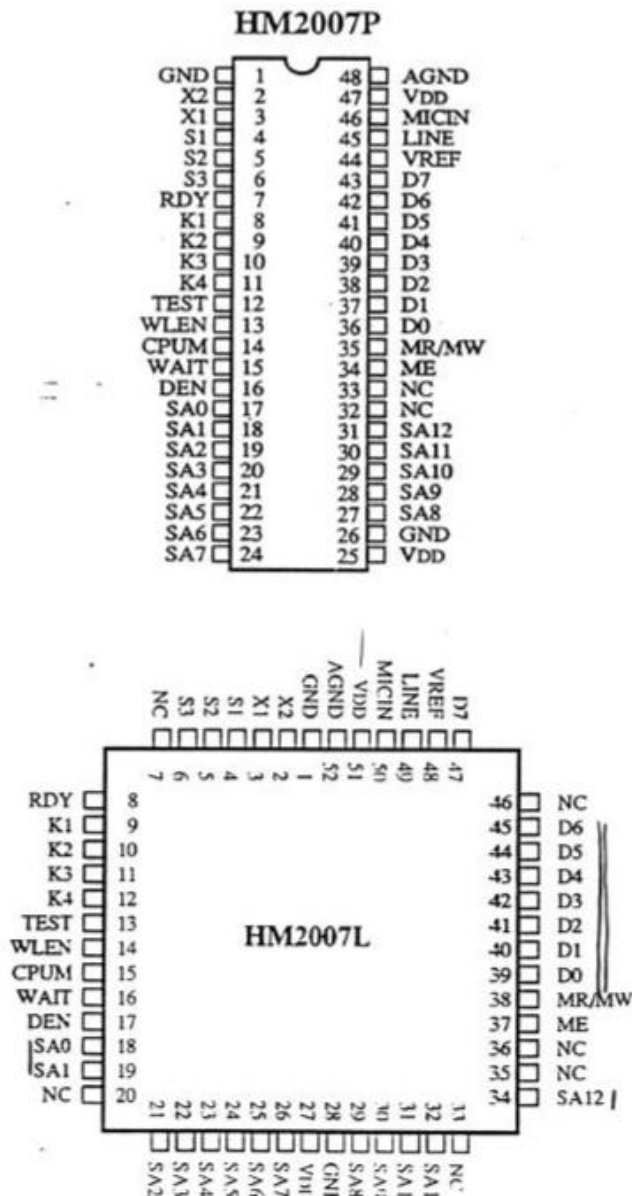


Fig 1. Hm2007 Chipset

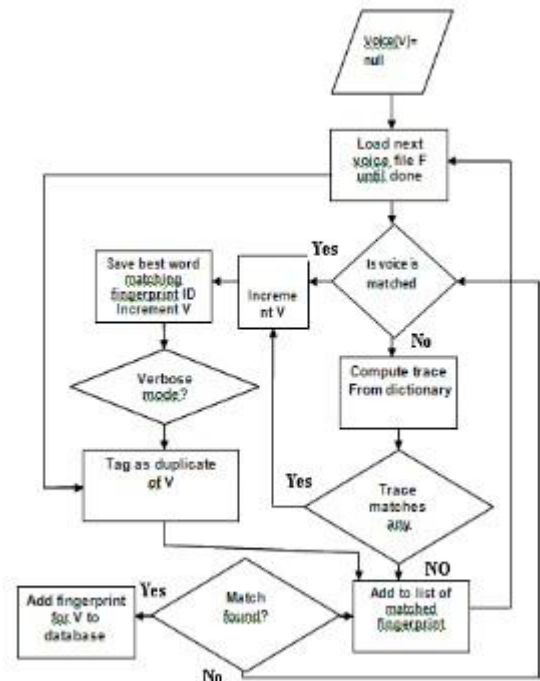


Fig 2: Flow Chart for Voice Recognition

Considering 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

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

The system is basically divided into two parts:

1. Human – Computer interaction
2. Computer – Robot interaction

3.1 Human – Computer Interaction

This module is basically text to voice application along with advancement of DQLand DML handled by the C# embedded program which is also connected with the ADO.NET and SQL lite database.

Three flags are there which are: S1, S2, and S3 [2]

- When S1 register is low then status register data come
- When S1 register is high then output register data come
- S2 is high means data is stored and retrieved from the bus structure
- S3 is high then data comes from input buffer

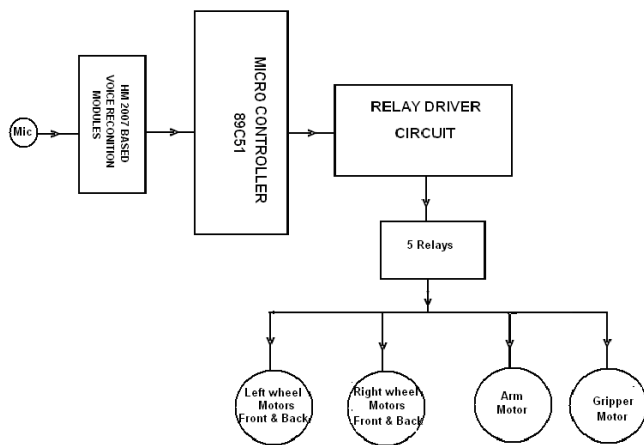


Fig 3. Proposed Architecture

Read write operations are controlled by this registers [6] setting value 0 and 1

- 0,0 - Training
- 0,1 - Read
- 1,0 - Write
- 1,1 - Output buffer

3.2 Computer-Robot Interaction

Functional Description:

There are two mode of functional [2] unit

- Manual mode
- CPU mode

Manual mode:

In this operation mode, a keypad a SRAM and other components may be HM2007 to build a simple recognition system. The keypad and digital display are used to communicate with and program the HM2007 chip. The keypad is made up of 12 normally open momentary contact switches.

Power on (wait pin)

When pin wait is "L" HM2007 will go for memory checkup and check whether the SRAM is perfect or not. When it is "H" then memory checkup step will be skipped and forwarded to recognition mode.

Recognition mode

When WAIT is H then this is ready to take the voice input, once the voice is inputted RDY is set to H. It is recommended that the recognition should be done before the training will be done otherwise the result will be unpredictable.

TRN AND CLR

This is a two mode operation for training and clearing the pattern. The operation will be carried out with reference of word number which is unique identification for each and every command.

- 24 for training
- 01 for clear command
- 99 for clear all command

CPU mode

This mode is the inter combination of all the operation controlled by the CPU directly. Specified CPU code are available for the command execution in the ROBOT.

RECOG	0001
RESULT	0010
TRAINING	0100
UPLOAD	0101
DOWNLOAD	0110
RESET	0111

Power On and Recognition mode are same as the manual mode.

Result

The recognition result is will be ready in buffer. When the data in the buffer has been read state will be set to 10. The result is sent in two part each part required two read action.

1. First part is for word number
2. Second part is for matching number

UPLOAD pattern

UPLOAD command is available for uploading the voice command along with its execution process. Two more word is required for this in which first one is low 4 bit second one is high 4 bit.[6] In this word length and data of the packet are combined. The word length required two word one 4 bit from low bit one 4 bit for high bit, similarly data pattern will required two word one from low 4 bit and other from 4 bit high.

DOWNLOAD pattern

This process is same as the previous one only the flow will be in the reverse order. This is also combined of two part each having two read action consisting of word length and pattern of the frame.

RESET Operation

RESET command is given by the processor to reset the command as reference to word number which is the identification for each command.

3.3 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 4th order Chebyshev [3] band pass filters. To create 4th order filters[5], and cascaded two second order filters using the following "Direct Form II Transposed" implementation of a difference equations.

$$\begin{aligned}
 y_1(n) &= b_{11}x_1 + b_{12}x_1(n-1) + b_{13}x_1(n-2) - a_{11}y_1(n-1) - a_{12}y_1(n-2) \\
 y_2(n) &= b_{21}x_2 + b_{22}x_2(n-1) + b_{23}x_2(n-2) - a_{21}y_2(n-1) - a_{22}y_2(n-2) \\
 y_{out}(n) &= y_2(n)
 \end{aligned}$$

The Euclidean distance formula for recognition is below:

$$d = \sqrt{\sum_{i=1}^n (p_i - q_i)^2} \quad \mathbf{P} = (p_1, p_2, \dots, p_n) \text{ and } \mathbf{Q} = (q_1, q_2, \dots, q_n)$$

3.4 Robot Operation

When the voice chip HM2007 authenticated the command which is given by the user or the processor as voice input it will update the status register for execution of the same. The execution is carried by the microprocessor and the functionality is mentioned in the database. According to the functionality it will set bit [6] code in the microprocessor in the microprocessor to execute the process. Socket setting is required to operate the back to back operation. Various operations are involved like

1. Go forwardgoes in forward direction cleaning all the way
2. Go backward...goes reverse direction cleaning all the way
3. turn left turns left
4. turn right turns right
5. Arm up pick & Place arm is moved up
6. Arm down Arm is moved down
7. stop movement stopped
8. open Gripper is opened
9. close Gripper is closed
10. automate picks the object from one position, places at particular place, then comes back to default position

4. COMPARISON WITH EXISTING SYSTEMS

1. HM2007 Chipset

HM2007 Chipset helps the developed robotic system to easily configure and recognize the voice patterns using Hidden Markov Models. Multiple chip configuration support.

2. Voice Filtration

Using Chebsy's [3] Voice Filtration Fourth Order Equation as explained above it is possible to filter [5] voice commands up to certain extent.

3. Faster Respond

The system will take maximum of 1.92 sec for word recognition

4. Dual Mode

The operation can be done in dual mode both in Manual and CPU mode[4] means user can directly give the voice command or it can be pass through the processor. Responsive time will be less than 300ms.

5. CONCLUSION AND FUTURE WORK

This robotic [1] system does provides an efficient interface to the user training in the system and also with service access for a pick and place movable arm to bring up thing to and

from areas distributed over a particular area. The

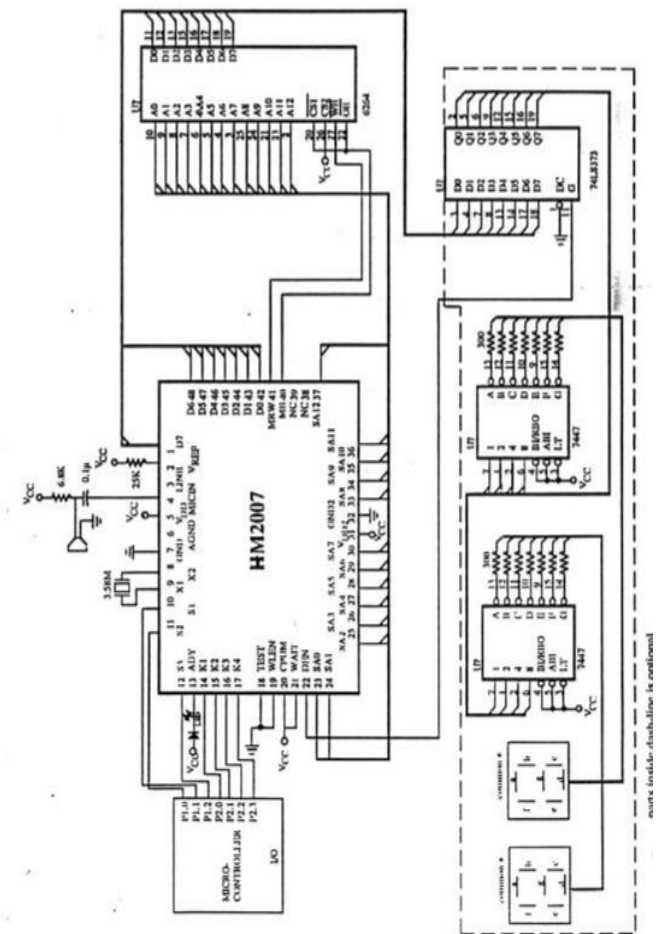


Fig. HM2007 with robotic system

system can be used along with a wireless camera to provide a live stream of the area which can be controlled from a remote place within the service area and thus would provide efficient by removing any kind of physical gestures whether wired or wireless to the robotic system which makes it an efficient system for usage. These system can also be integrated with some other efficient algorithm's to move in harsh environment providing service in various other professional and private environments to help aide in daily day to day work. These system also provides a more able interface with easy training and recognition features. Security of the system can be improved by using more complex fourth order differential equations so as to precisely detect and among various audio or voice instructions. In near future the other robotic applications can be integrated with the proposed system for the respective voice recognition of the user accessing the services.

References

- [1] Implementation of Speech Recognition Based Robotic System, Md. Abdullah Al Ahasan, Md. Abdul Awal and Sheikh Shanawaz Mostafa, 2011 IJCIT, ISSN 2078-5828 (PRINT), ISSN 2218-5224 (ONLINE), VOLUME 01, ISSUE 02, MANUSCRIPT CODE: 110108

- [2] Speech Recognition for Robotic Control Shafkat Kibria December 18, 2005 Master's Thesis in Computing Science, 20 credits Supervisor at CS-UmU Thomas Hellström Examiner: Per Lindström
- [3] AN EFFICIENT SPEECH RECOGNITION SYSTEM Suma Swamy and K.V Ramakrishnan Department of Electronics and Communication Engineering, Research Scholar, Anna University, Chennai, Computer Science & Engineering: An International Journal (CSEIJ), Vol. 3, No. 4, August 2013
- [4] Kilobot: A Low Cost Scalable Robot System for Collective Behaviors Michael Rubenstein, Christian Ahler, and Radhika Nagpal Michael Rubenstein, Christian Ahler, and Radhika Nagpal 2012 IEEE International Conference on Robotics and Automation RiverCentre, Saint Paul, Minnesota, USA May 14-18, 2012
- [5] Speech Recognition System For Controlling The Robot Ande Stanly Kumar, Dr.K.Mallikarjuna Rao, Dr.A.Bala Krishna., Assoc.Professor, Sri Vani School of Engineering, Vijayawada. Professor, JNTU College of Engineering, Kakinada. Professor, SRKR Engineering College, Bhimavaram.
- [6] Voice Automated Mobile Robot Rachna Jain , PG Research Scholar Dept. of Computer Engineering Delhi Technological University Delhi, India Dr. S.K Saxena Senior Faculty Dept. of Computer Engineering Delhi Technological University Delhi, India International Journal of Computer Applications (0975 – 8887)
- [7] Speech Recognition by Wireless Robot Nishant Dadwal 1, Praveen Pratap Singh 2, Pankaj Pandey 3, Ravikant 4, Mrs. Preeti Dhiman 5 1,2,3,4,5 Galgotias College Of Engineering & Technology, Greater Noida, India, International Journal of Electronics and Computer Science Engineering, ISSN- 2277-195