

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

ABSTRACT

This paper elucidates the research and implementation of voice automated mobile robot. The robot is controlled through connected speech input. The language input allows a user to interact with the robot which is familiar to most of the people. The advantages of speech activated robots are hands-free and fast data input operations. In future, it is expected that speech recognition systems will be used as man-machine interface for robots in rehabilitation, entertainment etc. In view of this, aforementioned system is a source of learning process for a mobile robot which takes speech input as commands and performs some navigation task through a distinct man-machine interaction with the application of the learning. The speech recognition system is trained in such a way that it recognizes defined commands and the designed robot navigates based on the instruction through the Speech Commands. The medium of interaction between humans and computers is on the processing of speech (words uttered by the person). The complete system consists of three sub-systems, the speech recognition system, a central controller and the robot. We have studied the various factors such as noise which interferes speech recognition and distance factor. The results prove that proposed robot is capable of understanding the meaning of speech commands.

Keywords

Speech Recognition system, .net framework, SAPI, voice automated mobile robot, application programming interface, device driver interface.

1. INTRODUCTION

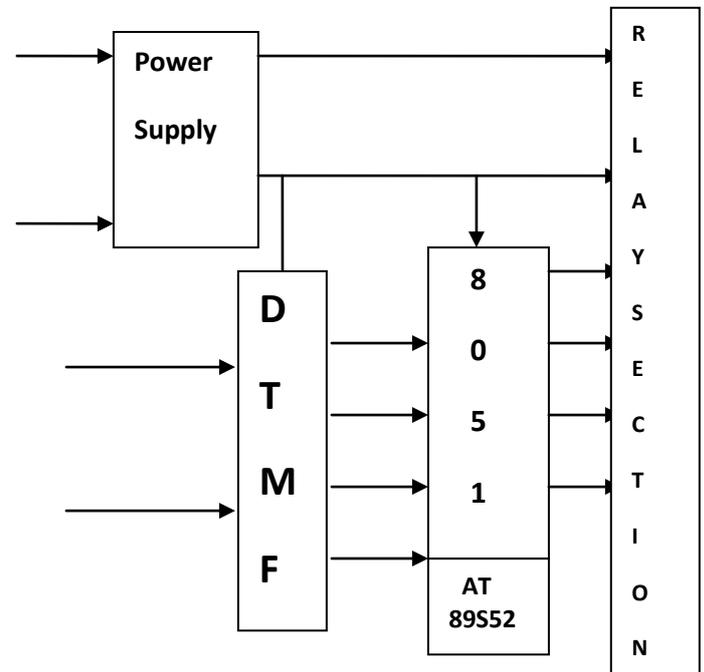
Voice based robotic control is an interesting project, mainly used for industrial and surveillance applications. It gives exact concept of controlling a robot by a voice instruction.

Robot is capable of understanding and synthesizing human speech for communication. A voice recognition unit built around a high speed processor that ensures various operations of the system to be performed by voice command. A few of commands recommended for operation are listed as: START, STOP, FORWARD, REVERSE, RIGHT, LEFT, SLOW, FAST, OK, UP, DOWN, CLOCK, ANTICLOCK, CLOSE, OPEN.

2. IMPLEMENTATION FRAMEWORK

The speech recognition system is programmed in the manner that the system has to be trained in spoken language (or vocal utterances) by the user so that circuit is to recognize the voice of a specific user. This board allows us to experiment with many facets of speech recognition technology. It has 8 bit data which interfaced with any microcontroller.

The 89S51 microcontroller contains four ports of each eight pins. In this project one port is dedicated for speech recognition. Relays are interfaced through ULN driver circuit to control the electrical appliances. A simple yet powerful program is written in assembly language and burned into the microcontroller to record and accept voice instructions and to control the devices. This project uses regulated 5V, 1A power supply.



3. SPEECH RECOGNITION SYSTEM

SAPI

SAPI is an API developed by Microsoft to allow the use of speech recognition and speech synthesis within window applications. This window extension gives workstations the

ability to create human like audio output from printed text. This ability adds a new dimension to human/pc interaction .Speech recognition services can be used to extend the use of PCS to those who find typing too difficult or time consuming.

SAPI is Part of the window open service architecture (WOSA [3] model) .speech recognition (SR) and text to speech (TTS) services are actually provided by separate modules called engines. user can select the speech engine they prefer to use as long as it conforms to the SAPI interface.

The SAPI 5.0 runtime was consist of two interfaces the application programming interface (API) and the device driver interface (DDI). The SAPI 5.0 AP I dramatically reduces the code overhead required for an application to use speech recognition and synthesis ,making speech more usable for applications by developers.

The SAPI 5.0 DDI and API remove the implementation details such as multithreading and audio device management making speech synthesis and recognition engines and applications convenient.

.Speech recognition is a technology that able a computer to capture the words spoken by a human with a help of microphone. These words are later on recognized by speech recognizer, and in the end, system outputs the recognized words. The process of speech recognition consists of different steps that will be discussed in the following sections one by one.

An ideal situation in the process of speech recognition is that, a speech recognition engine recognizes all words uttered by a human but, practically the performance of a speech recognition engine depends on number of factors. Vocabularies, multiple users and noisy environment are the major factors that are counted in as the depending factors for a speech recognition engine.

4. THE INTERFACE (CENTRAL CONTROLLER)

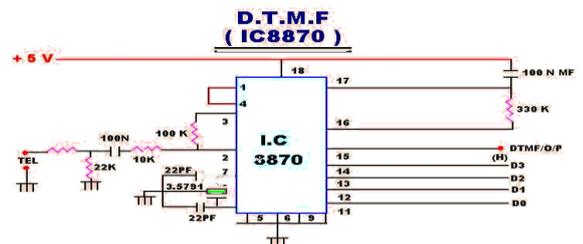
Dual tone multiple frequency(DTMF) interface system, which receives input from and outputs content to the user, can be divided into two subsystems: the DTMF(DUAL TONE MULTIPLE FREQUENCY) and THE SPEECH SYSTEM. Whenever computer recognized the particular DTMF tone which is sent through speaker and will be received by DTMF .The interface converts tone into BCD because microcontroller cannot understand analog signal as it can understand only digital signal that is done by interface.

There are so many ways to interface the device like using parallel port, serial port, USB port, dtmf. As serial port and parallel port are not available in laptop .It is very tedious job to interface with USB. So we have used DTMF here.

There is no need of DTMF encoder because we have wave file of all the tones like 0 -9 ,*,#.

0	-	0110 (A)
1	-	0001
2	-	0010
3	-	0011
4	-	0100
5	-	0101
6	-	0110
7	-	0111
8	-	1000
9	-	1001
*	-	1010
#	-	1100

But we don't know how to work on USB port and rest of port i.e. parallel port and serial port are not available on laptop that's why we use DTMF Decoder. DEDICATED IC FOR DTMF IS 8870



BLOCK DIAGRAM OF DTMF

B. Working of DTMF

In DTMF there are 16 distinct tones. Each tone is the sum of two frequencies: one from a low and one from a high frequency group. There are four different frequencies in each group.

Phone only uses 12 of the possible 16 tones. If you look at your phone, there are only 4 rows (R1, R2, R3 and R4) and 3 columns (C1, C2 and C3). The rows and columns select frequencies from the low and high frequency group respectively. The exact value of the frequencies are listed in Table I

TABLE 1: DTMF Row/Column Frequencies

LOW-FREQUENCIES

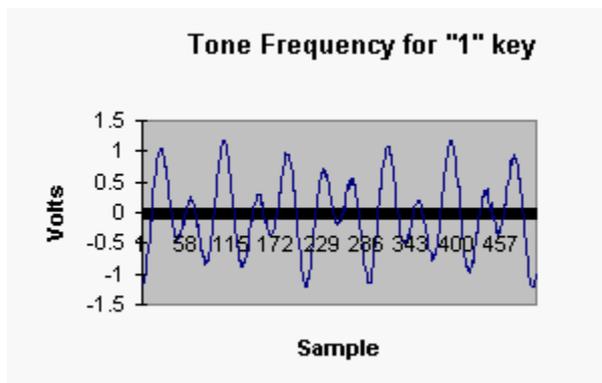
ROW #	FREQUENCY (HZ)
R1: ROW 0	697
R2: ROW 1	770
R3: ROW 2	852
R4: ROW 3	941

HIGH-FREQUENCIES

COL #	FREQUENCY (HZ)
C1: COL 0	1209
C2: COL 1	1336
C3: COL 2	1477
C4: COL 3	1633
C4 not used in phones	

Thus to decipher what tone frequency is associated with a particular key, look at your phone again. Each key is specified by its row and column locations. For example the "2" key is row 0 (R1) and column 1 (C2). Thus using the above table, "2" has a frequency of $770 + 1336 = 2106$ Hz. The "9" is row 2 (R3) and column 2 (C3) and has a frequency of $852 + 1477 = 2329$ Hz.

The following graph is a captured screen from an oscilloscope. It is a plot of the tone frequency for the "1" key:

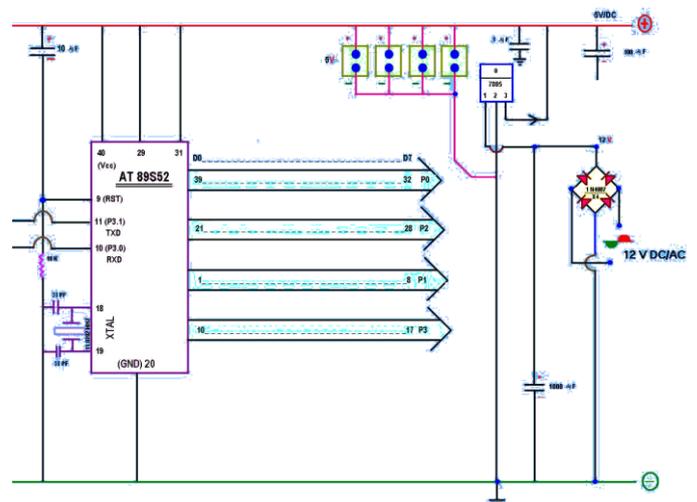


We can see that the DTMF generated signal is very distinct and clear. The horizontal axis is in samples. The frequency of the tone is about 1900 Hz - close to the 1906 Hz predicted by Table 3 ($697+1209$).

5. 80S51 MICROCONTROLLER

We are using 8 bit microcontroller but our data is 4 bit so the lower nibble of port ie P 1.0-P 1.3 are connected to D0-D3 and in higher nibble of port i.e. P1.4 –P1.7 all bit are set high. The AT89S52 is a low-power, high-performance CMOS 8-bit microcontroller with 8K bytes of in-system programmable Flash memory. The device is manufactured using Atmel’s high-density nonvolatile memory technology and is compatible with the industry-standard 80C51 instruction set and pin out. The on-chip Flash allows the program memory to be reprogrammed in-system or by a conventional nonvolatile memory programmer. By combining a versatile 8-bit CPU with in-system programmable Flash on a monolithic chip, the Atmel AT89S52 is a powerful microcontroller which provides a highly-flexible and cost-effective solution to many embedded control applications.

The AT89S52 provides the following standard features: 8K bytes of Flash, 256 bytes of RAM, 32 I/O lines, Watchdog timer, two data pointers, three 16-bit timer/counters, a six-vector two-level interrupt architecture, a full duplex serial port, on-chip oscillator, and clock circuitry. In addition, the AT89S52 is designed with static logic for operation down to zero frequency and supports two software selectable power saving modes. The Idle Mode stops the CPU while allowing the RAM, timer/counters, serial port, and interrupt system to continue functioning. The Power-down mode saves the RAM contents but freezes the oscillator, disabling all other chip.



BLOCK DIAGRAM

6. USES

1. It helps physically disabled persons by carrying some objects from one place to another place using the arm structure in the robot.
2. It guides the blind persons to reach a particular Destination by using the voice feature.
3. It is used to guide visitors in an organization by providing information about the facilities available.
4. Because of the presence of the Real-time Clock (DS1307), time-based control of the robot is possible. For example, it is used in hospitals to inform patients to take the tablets at the right time.
5. It is used in hazardous places.
6. The photo electric sensor in the robot will sense the obstacles and it will make decisions according to the obstacles it encounters.

7. FUTURE SCOPE

Some of interfacing applications which can be made are controlling home appliances, robotics movements, Speech Assisted technologies, Speech to text translation, and many more.

8. RESULT AND ANALYSIS

Table 1. Result And Analysis

COMMAND SET	CURRENTUSER	OTHERUSERS
COMMANDSET1	%85-%95	%80-%90
COMMANDSET2	%80-%90	%85-%95
COMMANDSET3	%90-%100	%95-%100

9. LIMITATIONS

Speech has difficulties to be recognized by an application. Because speech is different for every speaker, May be fast,

slow, or varying in speed. May have high pitch, low pitch, or be whispered. Have widely-varying types of environmental noise. Can occur over any number of channels .Changes depending on sequence of phonemes, May not have distinct boundaries between units (phonemes), Boundaries may be more or less distinct depending on speaker style and types of phonemes,. Changes depending on the semantics of the utterance, has an unlimited number of words, has phonemes that can be modified, inserted, or deleted.

10. SUMMARY AND CONCLUSION

A highly reliable and versatile system to accomplish a purpose design specific task such as distribution of medicine and food to the bed ridden patients specially in infected and inaccessible areas of the hospitals and medical centre have been reported. The on-board intelligence helps providing situational awareness a basic requirement of the system to be operated by voice / tele confined for ascertaining a majority of other tasks in open loop environment. The operation by voice command could best be used for handicapped.

11. REFERENCES

- [1] Dinesh P.Mittal and Goh Wee Leng School of Electrical And Electronics engineering, Nanyangang Technology Institute,Singapore.
- [2] Raj Kamal Embedded Systems: Architecture, Programming and Design, Tata McGraw-Hill Publishing
- [3] Rachna jain, Narina thakur 2nd National Conference INDIACOM-2008 Computing for Nation and Development. Bharati Vidyapeeth Institute Of Computer Applications and Management, New Delhi.
- [4] A Project By: K. Vivek, K. Chandru, and R. Manivannan Dept of Electronics and Communication Eng., PSG College of Technology, Coimbatore, India
- [5] Muhammad Ali Mazidi & Janice Gillispie Mazidi, the 8051 Microcontroller and Embedded Systems, Pearson Education.
- [6] Y.T. Yu, M.F. Lau, "A comparison of MC/DC, MUMCUT and several other coverage criteria for logical decisions", Journal of Systems and Software, 2005, in press.
- [7] P. M. Grant, Speech recognition techniques
- [8] Stephen Cook, Speech Recognition.