

Speech Enabled Intelligent Human Machine Interaction with Mobile Robotics

P. G. V. Suresh Kumar

Associate Professor, Department of IT & SC, AAIT, Addis Ababa University, Addis Ababa, Ethiopia

ABSTRACT: This paper is based on an application for controlling robot using speech. Recent developments in speech recognition technology have made an impact in various areas and are deployed for a wide range of applications. The work described here is an example of speech recognition technology coupled with mobile robotics. It provides a glimpse of the ample possibilities of human-machine interaction in the field of robotics. A 4-channel radio controlled Wheelbot and Microsoft's speech SDK is deployed in this work. The application has been incorporated with the ability to recognize predefined commands and the robot has been designed to move according to the commands given by the user.

KEYWORDS: Mobile Robotics, Speech Recognition, Man-Machine Interface, Wheelbot.

I. INTRODUCTION

Speech recognition is the process of converting speech signals to a sequence of words to identify the utterance [1], [2]. Speech can be captured by a system using an audio input device such as microphone or a telephone. A system may allow a user to input a limited number of words or sentences. Such system is commonly referred to as command and control application [1], [2], [3]. Speech recognition has three main components namely the SR engine, application program and an application program interface. All such systems can be viewed as having a three layered architecture.

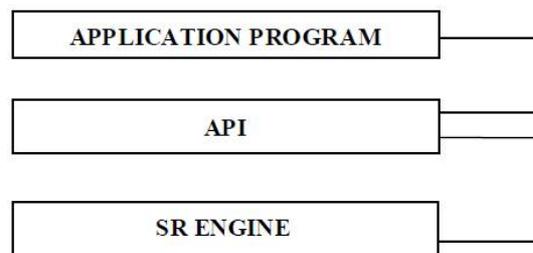


Fig1: Architecture of a speech enabled application.

A speech-enabled application and an SR engine do not directly communicate with each other. All communication is done using an API. Speech API provides a high level interface between an application and the speech engine. SR engine is responsible for recognizing commands. It analyzes the user input speech and compares against a grammar which defines rules of the possible sequence of words or phrase and successfully initiate the command. Based on these comparisons the engine assigns a confidence score. Confidence score is an attribute with numeric value that denotes how confident the engine is about the speech recognized. The resulting action of the system depends on the confidence threshold set by the application developer. This threshold determines the level below which an utterance is rejected and can be identified as false recognition [3]. In this project Microsoft SAPI is used as the API. SAPI controls a number of aspects of a speech system. It controls the audio input, whether from a microphone, file, or a custom audio source and converts audio data to a valid engine format. It loads grammar files and resolves grammar imports and does grammar editing. Compiling standard SAPI XML grammar format and parsing semantic tag is results is also done by SAPI. It finally returns results and other information back to the application. The SR engine uses SAPI for grammar interfacing.

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Vol. 5, Issue 11, November 2016

It polls the SAPI for information about grammar, performs recognition and generates event such as false recognition to provide information to the application.

II. THE WHEELBOT

A radio controlled Wheelbot with a pair of TX-2B/RX-2B CMOS LSI mounted on the controller and the Wheelbot respectively has been used in the experiment. The TX-2B/RX-2B has five control keys for controlling the motions i: e; forward, backward, left, a right and stop functions. The controller uses an infrared tower to transmit signals to the Wheelbot. These signals are received by the Wheelbot when it is in the range of controller. The range of the controller is observed to be 7 to 8 meters approximately. These signals are interpreted as instructions to activate one of the five functions [5]. Fig2 shows six important pins of TX-2B chip. Pin 1, 4, 5 and 14 are responsible for movements in right, backward, forward, and left directions respectively. Whichever of these pins is fed logic 0 the corresponding function is activated (actually a coded signal is generated which when received by the Wheelbot the corresponding functions is generated). Pin 3 is grounded and pin 9 is connected to the voltage supply.

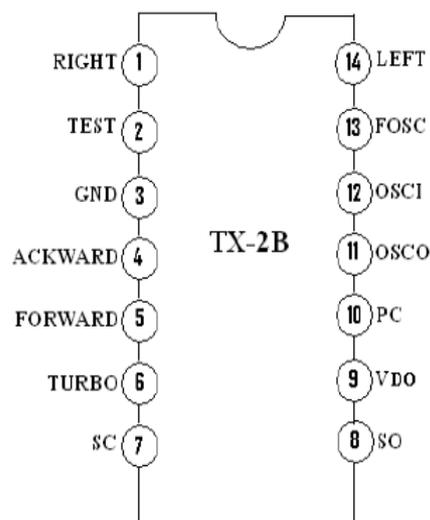


Fig 2: Pin configuration of TX-2B chip [5]

III. MOTIVATION

PC parallel port is a 25-pin female connector usually in the back of the computer. It has three different sections as shown in Table1. These are data lines, control lines and status lines. There are 8 data lines, and they are used to let information out of the port. The control lines are another 4 outputs. The status lines are parallel port's only inputs. There are 5 of them. Each section has its own address and acts independently from the rest. These addresses are in sequential order. That is, if the data port is at address 0x0378, the corresponding status port is at 0x0379 and the control port is at 0x037a [6],[7]. In order to use the port, the address of the port, the command to access the port and the number to be save in it must be identified. These ports work with numbers. It can be expressed in Hex, Binary or Decimal. For instance to set the 8 data lines to 11111111, either 255 in decimal or 0xff in Hex is to be used. Similarly to set them to 10000001 either 129 in decimal or 0x81 in Hex is to be used. [6], [7].

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Vol. 5, Issue 11, November 2016

Port	Address(Hex)
Data line	378h
Status line	379h
Control line	37Ah

Table1: Port address of 3 different ports

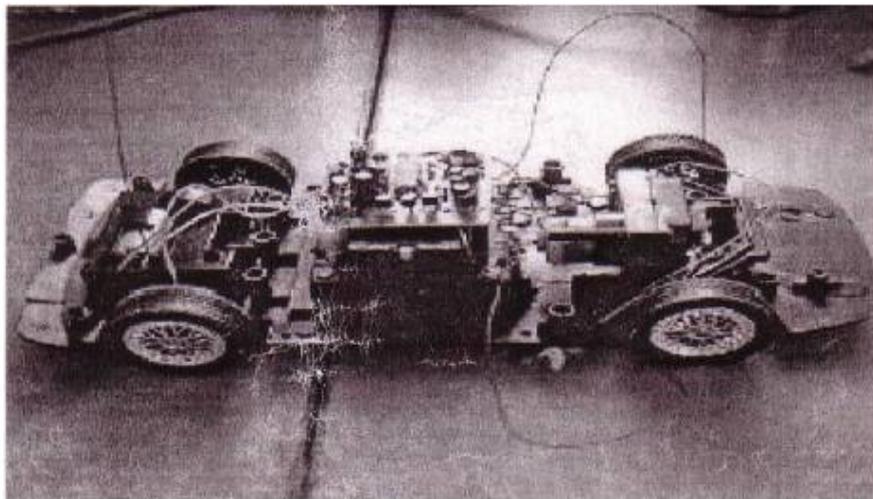


Fig 3: Wheelbot

A. DATA PORT

The data pins are TTL level output pins. These pins ideally put 0V when they are in low level logic and +5V when they are in high level logic. However the voltages can vary from ideal conditions when the circuit is loaded [7], [8]. As shown in the Table 2 below we observed the voltage output to be 4.5V.

Pin	Functions	Voltage (Vcc)
2	D0	4.5
3	D1	4.5
4	D2	4.5
5	D3	4.5
6	D4	4.5
7	D5	4.5
8	D6	4.5
9	D7	4.5

Table 2: Observed output at data pins

B. INTERFACING CIRCUIT TO PARALLEL PORT

The transmitter [TX-2B] circuit gets minimum 3.0V voltage from some external source [5].only four different data pins 2, 3, 4&5 are used in order to connect the four ends of the circuit to the PC printer port which are responsible for generating signals for four different instructions to the robot namely "Go Back", "Advance", "Turn Left" & "Turn Right" as shown in Fig 4.

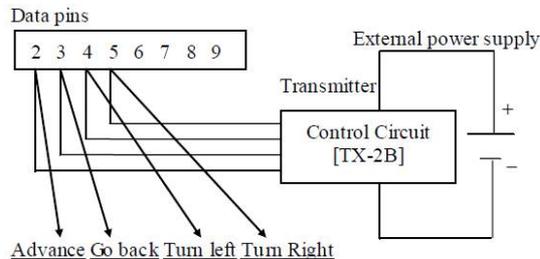


Fig 4: PC printer port and IR control interface circuit

IV. CONTROLLING OVER VOICE

The controller of the Wheelbot is attached (using printer cable) to the printer port of the computer which runs an application capable of recognizing six commands (run, back, left, right and straight). A CFG grammar is written, compiled and stored in a DLL file which is loaded by the application at run time to recognize these six commands. The grammar also restricts the system to recognize only these six commands. The application program is designed to make an output on the particular pin on the DATA port go high and low depending upon the command recognized[8]. When the output on a particular data pin is low (logic 0) the corresponding pin on the controlling activates a particular function. Thus the Wheelbot can be controlled using speech.

V. PROPOSED WORK

The application contains a single form. The form on being loaded displays an empty text box shown in fig 5 and in the background the form load event initializes some information containing speech recognition events on the form. Speech API returns certain events to the application such as a successful recognition. The application then addresses or ignores the events. Furthermore the application can selectively reject certain types of events if there is no need to work with them. The action corresponding to a recognition event is handled by appropriate SAP events handler defined by the application designer. Upon successful recognition of the command, the speech recognition event SRERecognition is returned and the name of the recognized command gets displayed in the text box, as shown in Fig 5.

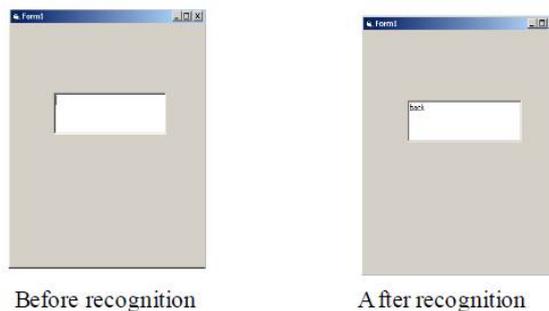


Fig 5: User interface displaying the result of a successful recognition.

VI. CONCLUSION

Applications incorporated with speech recognition technology are fragile in nature and their performance depends upon various factors such as external noise, type of input device used, amount of training given to the system and many more. For example training improves a systems performance if it is a speaker dependent system. The size of vocabulary (the number of words allowed to communicate) and the grammar also affects a systems performance. As yet no speech enabled application has enabled 100% accuracy The following graph shows the performance of the proposed system in terms of word error retain varying environment. The system was tested twice each time in a different environment. At first the system was tested in a closed room with almost no external noise. The next time it was tested in an open area

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Vol. 5, Issue 11, November 2016

with some noise. We observed that when the noise is less and the confidence threshold is set to low then system's recognition is as high as 100%. In the same environment with systems confidence threshold set to high, performance degrades to 70%. when the system was tested in a noisy environment with low confidence threshold than system misinterpreted some external noise as a command and thus behaved unexpectedly. And with high confidence the system performance comes down to 50%. The observation shown in the graph fig 5 is indicative and may vary every time the system is tested. As the system was tested by four different users the result can be better if it is used by a dedicated user with good English accent. Result can also be improved by giving more training to the system. The confidence threshold can be set using the speech properties in the control panel of the computer.

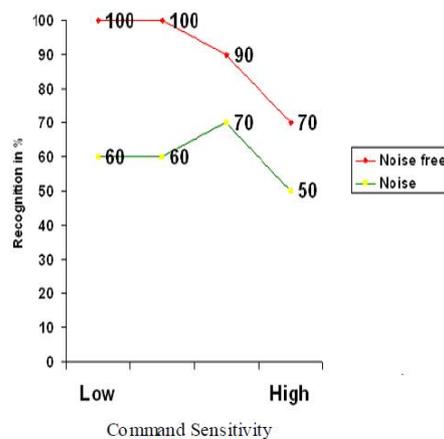


Fig 6: Graph Showing Percent of Recognition vs Command Sensitivity

VII. FUTURE SCOPE

The system can be extended in a way that the user of the machine is able to control and monitor the device on a network. The system has been incorporated with intelligence to recognize human voice instructions and respond to it. Efforts to make the system capable of tracking obstacles in its path and making decisions on its own are under progress. Efforts to improve the systems performance in the terms of reduced word error rate under varying environments is under progress [4].

REFERENCES

- [1] <http://en.wikipedia.org>
- [2] <http://cslu.cse.ogi.edu>
- [3] www.pronexus.com
- [4] Li Deng, Xuedong Huang "Challenges In Adopting Speech Recognition", In Communications of the ACM January 2004/Vol.47.No.1
- [5] <http://pdf1.alldatasheet.com>
- [6] www.arroncake.net
- [7] www.e panorama.net
- [8] Peter H. Anderson "Use of a PC Printer Port for Control and Data Acquisition" in the Electronic journal for Engineering Technology pp no 258.