COPYRIGHT © 2011 IJCIT, ISSN 2078-5828 (PRINT), ISSN 2218-5224 (ONLINE), VOLUME 01, ISSUE 02, MANUSCRIPT CODE: 110108

Implementation of Speech Recognition Based Robotic System

Md. Abdullah Al Ahasan, Md. Abdul Awal *and* Sheikh Shanawaz Mostafa

Abstract—Robotics has a momentous feature and future in our daily life. It can make our life fast and easier. But for real life man and robot interaction like household or workplace creates a new question about the controlling the robot. It is quite impossible to command a robot through a keyboard or interfaces like this. The problem can be solved by the way the human solved it, speech. Speech has a huge opportunity to solve this problem. But the speech recognised based robot controlling system is rarely implemented. Speech recognition is a broader solution which refers to technology that can recognize speech without being targeted at single speaker. This paper describes a robot which can recognize the voice command and acted as it is programmed to do. Our Speech recognition program converts spoken words to text. Then text compared to grammar helper for its activity, if the text matches with any command with the database. It executes the command.

Index Terms— parallel port, robotic arm, SDK engine, and stepper motor.

1 INTRODUCTION

A robot is an automatically guided machine, able to do tasks on its own. Another common characteristic is that by its appearance or movements, a robot often conveys a sense that it has intent or agency of its own. [1][2] The Robotics Institute of America (RIA) defines a robot as a "reprogrammable multi-functional manipulator designed to move materials, parts, tools, or specialized devices through variable programmed motions for the performance of a variety of tasks."[3]The RIA subdivides robots into four classes: devices that manipulate objects with manual control, automated devices that manipulate objects with predetermined cycles, programmable and servo-controlled robots with continuous point-to-point trajectories, and robots of this last type which also acquire information from the environment and move intelligently in response.

In the industrialized circumference, robotics is a step beyond mechanization. Robot plays an increasingly important role in the world economy and in daily experience. Engineers try to combine robots with mathematical and organizational tools to create complex systems for a rapidly expanding range of applications and human activities. Many roles for humans in industrial processes presently lie beyond the scope of normal automation robots. Human-Machine Interfaces (HMI) or Computer Human Interfaces (CHI), formerly known as man-machine interfaces, are usually employed to communicate with computers, such as entering and monitoring temperatures or pressures for further automated control or emergency response. But complex work likes pattern reorganization, speech reorganization needs human expertise.

Our robotic arm is an automatic manipulator, which is easily programmable and controlled by basic voice command. As the robot is wirelessly connected to user so this robot can be used non-human friendly situation. The links of this manipulator are connected by joints allowing translational (linear) displacement. The links of the manipulator can be considered to form a kinematic chain. The end of the kinematic chain (consists of nearly rigid links (members) which are connected with joints (kinematics pair) allowing relative motion of the neighboring links) of the manipulator is called the end effectors and it is analogous to the human grip. The end effectors can be designed to perform any desired task such as welding, gripping, mixing etc., depending on the program. It acquired its information using a speech recognizer then by comparing and converting the speech to text it manipulates its grip.

2 BACKGROUND PROBLEM

How nice would it be if the computer could listen our language! Why can't we just talk to the computer, indicating what we want, just like we do with other human beings?

Now that we've look the specific task of speech recognition. Speech recognition technologies allow computers equipped with a source of sound input, such as a microphone, to interpret human speech. This input can be classified as: uninterrupted recognition and interpreted recognition. Interpreted recognition has transcription, control, mix of both. Uninterrupted recognition is basically just recording of the sound .Since it does not involve trying to understand what's in the waveform, there's no need for further processing. Transcription is the dictation of words into a text editor. While it is normally done by keyboard, it can also be done by speech. Control is the control of various applications or the operating system (OS)

- ***Md. Abdullah Al Ahasan*** *is with the Discipline of Computer Science and Engineering, Khulna University (www.ku.ac.bd), Khulna-9208, Bangladesh. E-mail: ahasanzia@yahoo.com.*
- ***Md. Abdul Awal*** *is with the Discipline of Electronics and Communication Engineering, Khulna University (www.ku.ac.bd), Khulna-9208, Bangladesh. E-mail: awalece04@yahoo.com.*
- ***Sheikh shanawa Mostafa*** *is with the Dept. of Biomedical Engineering, Khulna University of Engineering and Technology (www.kuet.ac.bd), Khulna-9203, Bangladesh. E-mail: shawan44@gmail.com.*

by speech. A mixture of both is also possible, for example when asking a text-editor to do something with the text. The computer usually receives a single wave-form full of speech and other noises like background noise, distorted input and meaningless noises by speaker. Examples of the last are repeating, pauses, and continuation noises (uhhm..., errr...). Finding out what is what is more difficult than it seems. Humans can hear the direction of the sound, and can focus accordingly. Also, they can interpret body language and van use context. With only one ear (the microphone) and no visual information, the computer is severely impaired. - Variation in speaker (having a cold, emotional state), and between speakers/accents/languages - word boundary detection. With a continuous stream of voice, segmenting sentences into words can be hard due to difficulty in word boundary detection. [4]

Another problem is to interfacing these speech recognizers to a heard were. Programs which send data to port were pretty easy in old DOS days and in Win95/98. Being a very secure operating system compared to other previous windows operating system, Windows NT assigns some privileges and restrictions to different types of programs running on it. It classifies all the programs in to two categories, User mode and Kernel mode ie; running in ring3 and ring0 modes. User mode programs are running in ring3 mode and Kernel mode programs are running in ring0 mode. The normally written programs fall in the user mode category. The user mode programs are restricted to use certain instructions like direct accessing etc. Whenever the operating system finds that a user mode program is trying to execute such instructions, the operating system stops execution of those programs and will display an error message. But in the same time Kernel mode programs are in no way restricted in executing such instructions. [5]

3 SPEECH RECOGNISE ROBOT TIONS

The generalized concept of this robot speech recognise robot is that its pick up sound from surrounding area by using a microphone. Microphone is wirelessly connected to computer (Fig.1).

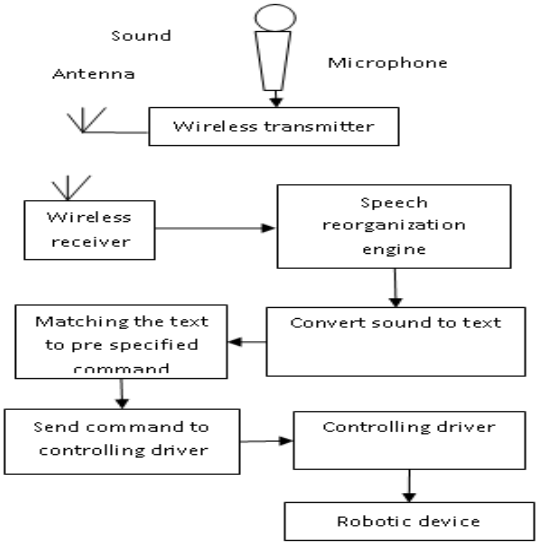

Fig. 1 Block diagram of speech recognise based robotic system

The audio transmits through the wireless system to the computer interface. The receiver driver of computer system captures the audio. The capture audio then converted to text string by sound to text converter subsystem of the program. This text string compared to the database for exaction command. If a command is found against the text the implementation command is send to robot for do the activity.

This system can be divided two sub parts. Software and hardware

3.1 software

C# language is used for programming the software of this project. And The Microsoft Speech SDK 5.1 is use for speech reorganization. It adds Automation support to the features of the previous version of the Speech SDK. The design of the new API included the concept of strictly separating the application and engine so all calls were routed through the runtime sapi.dll. This change was intended to make the API more 'engine-independent', preventing applications from inadvertently depending on features of a specific engine. The new API was initially a pure COM API and could be used easily only from C#. A recognizer object can be used that runs in a separate process (sapisvr.exe) in windows. All applications using the shared recognizer communicate with this single instance. This allows sharing of resources removes contention for the microphone and allows for a global UI for control of all speech applications. The Flow diagram is shown in Fig2.

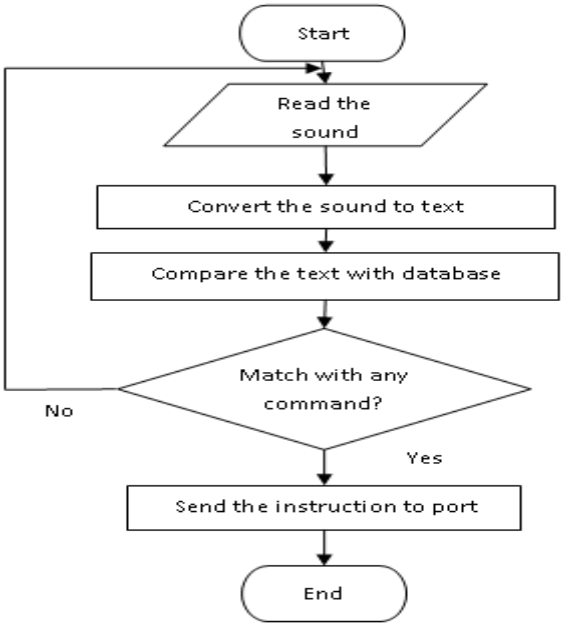

Fig. 2 Flow diagram of the robot controlling program.

Using some of persuasive word like start, stop, move, close, up and down SDK engines convert this word in audio string and compare this audio with recent audio activity pick up by microphone. It compares the human voice with its own grammar helper. If matched with its grammar helper then it passes a text string to C# and then the text string is compared with the predefined text that given by software developer. The graphical user interface (GUI) based robot controlling program developed using C# is shown in Fig. 3.

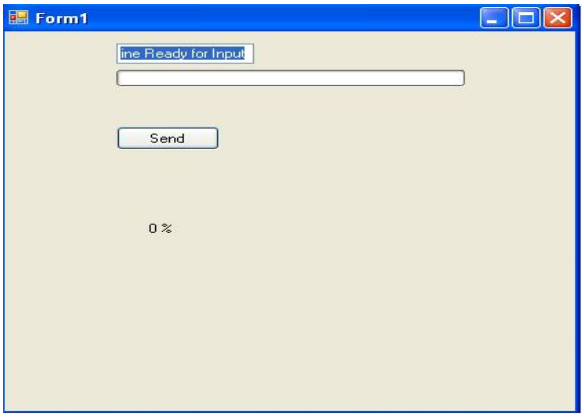

Fig. 3 GUI of the robot controlling program

3.2 Hardware

There are three main hardware computer, microphone, and robotic arm itself. Computer is for running the software measuring the audio input through microphone.

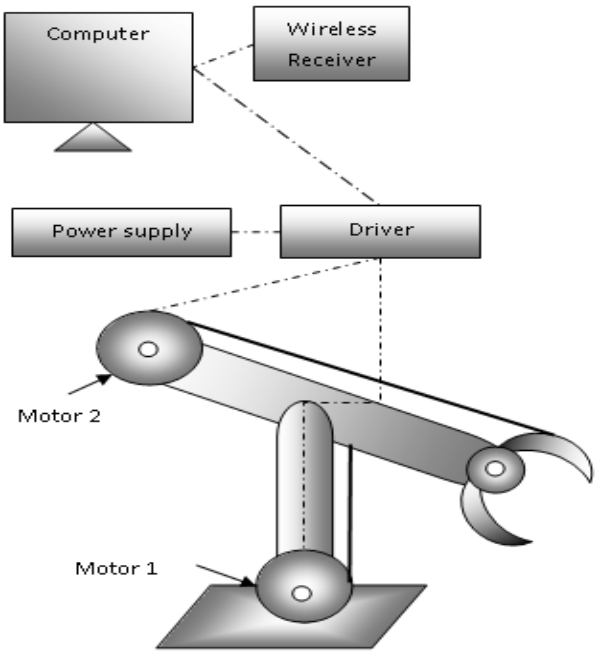

Fig. 4 Schematic diagram of the robotic arm

Robotic arm section is very simple design. As we manipulate it by two steeper motor and these steeper motor is controlled by the driving circuit and the circuit is controlled by the computer command. Familiar types of different pull or gear is used hear. It is used here as the liver. Motor one controls the movement of ups and downs. Motor two controls the movement of opening and closing. Here it can be stopped at any movement willingness of user. The robotic arm can move two of its joints. These joints are created by EM 546 (collected from old printer) 4 pole stepper motor. The griper joint can move 0-100 degree. And motor 1 can move 0-120 degree. Stepper motors are used as uni polar and their excitation character is shown in Table 1.Moreover the Schematic diagram and the implementation of our speech reorganisation based robotic arm are shown in fig4 and Fig 5 respectively.

TABLE 1
EXCITATION TABLE FOR STEPPER MOTORS

Pole 1	Pole 2	Pole 3	Pole 4
+	-	-	-
-	+	-	-
-	-	+	-
-	-	-	+

Fig. 5 Implementation of our speech reorganisation based robotic arm

4 CONCLUSION

In this paper, we try to implement the robotic arm manipulation system and manipulate it by voice command. It can understand any human voice; doesn't need to depend on specific user voice. To do that sometime it can't recognize command because of different pronunciation and different tone of different people. It has an extra facility it can be controlled by wireless. So that user doesn't need to physically appear in the working area, it can be used in. It has a great advance provided that wireless control system always in the range. In near future we have a plat to build it commercially for mankind. It is helpful for all of the robotic research and project based work.

REFERENCES

[1] ROBOT" Merriam-Webster Online Dictionary. 2010[Online]. Available: http://www.merriam-webster.com/dictionary/robot

[2] "Robot" American Heritage Dictionary. 2010 [Online]. Available: http://education.yahoo.com/reference/dictionary/entry/robot

[3] Bright Hub 2010. [Online]. Available: http://www.brighthub.com/engineering/mechanical/articles/71937.aspx

[4] Joris de Ruiter, Natural Language Interaction - the understanding computer, 2010 [Online], Available: www.few.vu.nl/~jdruiter/published_work/Natural_language_interaction.pdf

[5] Architecture of Windows NT. Wikipedia, 2010 [Online], Available: http://en.wikipedia.orsg/wiki/Architecture_of_Windows_NT

Md Abdullah Al Ahasan is a student of Computer Science and Engineering (CSE) discipline in Khulna University. His research interest is on the Software Development, and Programming on Artificial Intelligence and Expert Systems. Recently, he is engaged in research activities with this paper. Moreover, he is a member of Bangladesh Electronic society.

Md. Abdul Awal was born in Dinjpur in 1st august 1986. He got his B.Sc. Engineering degree in Electronics and Communication in the year of 2008 from Khulna University, Khulna-9208, and Bangladesh.He is now the faculty member of Electronics and Communi cation Engineering Discipline, Khulna University, Khulna-9208, Bangladesh. Now he is studying as a part time student in M. Sc. Engineering in Khulna University of Engineering and Technology. His current research interested is wireless communication, information technology, network design and biomedical signal processing. His number of published paper are 5 among them international recognized journal and proceedings of international and local conference. He is also a member of Bangladesh Electronic society.

Sheikh Shanawaz Mostafa was born in Khulna on 9th July 1988. He got his B.Sc. Engineering degree in Electronics and Communication in the year of 2010 from Khulna University, Khulna-9208, Bangladesh. He is currently a M. Sc. Engineering candidate in the Department of Biomedical Engineering, Khulna University of Engineering and Technology, Khulna, Bangladesh. His current research interests are in wireless communication, information technology, network design and biomedical signal processing. His number of published paper are 5 among them international recognized journal and proceedings of local conference.