A CONVERSATIONAL MAN-ROBOT INTERFACE USED IN ROBOTICS EDUCATION

Catalin Buiu, Ioan Dumitrache, Alexandru Dumitrascu

Department of Automatic Control and Systems Engineering,
"POLITEHNICA" University of Bucharest, Faculty of Automatic Control and Computer Science,
Spl. Independentei 313, 77206 Bucharest, Romania
Email: cata@ics.pub.ro

Abstract: This paper presents a conversational interface that uses the speech recognition and synthesis and animation abilities of two Microsoft software agents in order to assure a more natural and efficient interface with an autonomous robot. The interface is used in the robotics educational process. The basic commands to be given by voice (back, right, etc.) and the control of the agents are implemented in Visual Basic. More complex control programs can be implemented and tested in Matlab and after that executed with a simple voice command. This program is used as a part of an integrated educational strategy in autonomous robotics and was evaluated by a group of students. First experimental results related to the use of this conversational interface compared to traditional robot simulation environments and control of real robots are presented in this paper with related discussions.

Keywords: control education, autonomous robots, speech recognition, conversational interface, software agent.

1. INTRODUCTION

Robotics is a quickly changing and very open field. Important recent advances have raised a high public interest in this key scientific application area of the 21st century. Robotic fairs attract a large number of visitors, while the applications have a wide range, from military to education and entertainment.

Robots that work directly with humans, as assistants or teammates, are being designed and developed. Human-robot interaction can be defined as the study of humans, robots, and the ways they influence each other (Fong, *et al.*, 2003). The traditional view of the robot as a tool, operating on human command is changing rapidly, and the robot will become, more or less, a "partner" (Fong, *et al.*, 2001). In order to achieve this, high performance man-robot interfaces are needed.

Robotics education requires careful design of student-robot interfaces. One attractive possibility that is described in this paper is to command the robots by using voice. Section 2 presents various voice-controlled robots. As many educational robots have not speech recognition modules, the idea was to use the speech recognition and synthesis abilities of an animated Microsoft software agent. An additional animated agent is used as an interface help agent. The agents and the application are presented in great detail in Section 3 which is the main technical part and describes the interface with its basic features.

The paper is ended with student evaluations in Section 4 and conclusions in Section 5.

2. VOICE-CONTROLLED ROBOTS

Speech recognition involves capturing and digitizing the sound waves, converting them to phonemes, constructing words from phonemes, and analyzing the words in the given context in order to assure correct spelling for similar words. The speech recognition engines are software drivers that convert the acoustical signal to a digital signal and deliver recognized speech as text to the application.

Various robots can be controlled now by using voice commands, including teleoperated and Internet robots. Intelligent vision and voice controlled robotic workstations provide disabled people with the ability to perform various functions using a robotic arm which is controlled by the users own voice. The robot can be used to perform simple pick and place operations through a voice activated enhanced tele-operation interface. Voice-controlled robots aid surgeons to perform complicated cardiac and abdominal microscale operations, to avoid hand tremors and they also offer visual magnifications.

Due to the recent technological advances and the availability of voice recognition software, many of the existing robots have speech recognition kits, which can be trained to recognize words or phrases in any language, e.g. Son of Zylatron and OCT-1c. Voice Extreme Toolkit™ offers interactive speech recognition using a programmable module, development board and suite of development software.

The RB5X robot produced by General Robotics can speak any language in the world, through the use of international phonemes. In September 2003, AIBO (Artificial Intelligence Robot) MIND was demonstrated to allow owners to interact with the robot in numerous ways, including voice and tactile touch sensors, along with the ability to remotely access the robot and retrieve digital images on their PC via e-mail commands or an Internet browser.

Voice controlled robots work for the Museum of Science and Industry in U.S. The Sound Reversing Car is a voice control robot car, while the GEMINI robot is a life size robot with speech recognition and synthesis abilities. It accepts commands by voice and has an integrated voice command language (VOCOL).

The Voice Interactive Robot project presents an embedded system that uses the real-time JStamp native execution Java hardware and software and Legos Mindstorms and Technics hardware along with VoiceDirect 364 speech recognition module and ChipCorder 2560 playback/record chip to control robot automation. In (Marin, 2002) it is presented an educational robot that is controlled via web. It has some built-in modules that include speech recognition. It accepts voice commands like "Grasp the object one" or "Grasp the cube". However, the set of possible commands is quite limited and not easy to extend.

The application reported in this paper was developed at DLAB, a distributed laboratory for robotics education and research that has concentrated on various autonomous robotics related applications (Buiu, 2002). These include intelligent control strategies (fuzzy logic, neural networks, and hybrid geno-fuzzy systems) and experiments in local and remote collective robotics (Buiu, 2002). DLAB uses 5 Khepera robots (Fig. 1) and one Koala robot, both manufactured in Switzerland by K-Team, a world leader in mobile robotics.

Khepera is equipped with two motors and a belt of eight proximity sensors. Fig. 2 shows the numbering, position and orientation of sensors and actuators embedded on Khepera. Khepera can be controlled in three different ways: via a serial link protocol, by downloading the control program to the SRAM of the Khepera, or by creating a customized EEPROM. Khepera accepts two kinds of instructions: commands and tools. Commands consist in giving speed and position, reading data from the sensors; tools can achieve more complex tasks. Commands are given to Khepera through a serial port, that is programmed without any flow control with 1 start bit, 2 stop bits, 8 data bits. Transmission speed can be 9600, 19200 and 38400 bps. An example of a speed control command is D,5,5 which sets a speed of 5 on the left and right motors of Khepera. To stop the Khepera, one may give the D,0,0 command and to turn right D,5,-5 is an example.

Fig. 1. Khepera robot

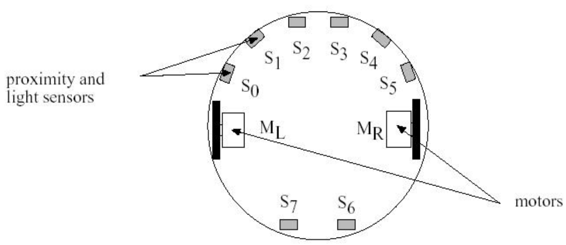

Fig. 2. Sensors' position

3. THE CONVERSATIONAL MAN-ROBOT INTERFACE

3.1 Mobile robotics education

One can identify some interesting and challenging aspects in mobile robotics educations. Robots are expensive pieces of hardware. Simulators are not widely available. An important and challenging field of robotics is cognitive robotics. Cognitive robotics is concerned with the design of robots that function in a changing, incompletely known and unpredictable environment by using high-level cognitive abilities. The emphasis here is on controlling real (as opposed to merely simulated) robots and this is why the student needs an extensive period of training on real robots. Especially, man-robot interface is a delicate issue for the new-comers in this field.

Mobile robotics education needs to combine the use of simulators and the work on real-world robots. How to efficiently do that, how to interact with complex autonomous mobile robots and what to do when there are few robots to experiment with, these are open questions that were addressed at DLAB (Buiu, 2002).

This paper proposes the use of voice for controlling mobile robots, using the speech recognition abilities of an animated software agent. Another animated agent is used for pedagogical purposes acting as an interface help agent. The main advantage of using conversational software agents is that they offer the students a more "human" interface with the robot, the application becoming more friendly and realistic.

Boring control programs can transform into attractive control experiences when the student is able to use his own voice and when he is assisted by an animated character as a tutor.

This kind of scenario promotes the acquisition of know-how and is a part of DLAB's educational strategy, which was designed to combine traditional classroom education with remote experimentation on multiple robots while promoting an active participation of the students in various tasks that can be also competitive. DLAB's educational strategy in robotics (Buiu, 2002) is based on simulation and local control of robots, followed by remote control of one or more robots (remote collective robotics). The use of interactive software agents is a solution for the lack of human tutors, both in local and remote experiments (Albu, *et al.*, 2001).

3.2 Animated conversational agents

Lifelike characters are one of the most exciting technologies for human– computer interface applications. They are synthetic agents apparently living on the screens of computers. Characters can be lifelike in a "human-like" or an "animal-like" way. These lifelike autonomous characters co-inhabit learning environments with students to create rich, face-to-face learning interactions. Recent research with lifelike pedagogical agents has indicated positive effects on learners' attitude towards learning and performance (Baylor, 2002a, 2002b; Baylor and Ryu, 2003a; Moreno, et al., 2001).

Other studies have shown that using artifical characters (or *personas*) can improve the users' satisfaction with the system providing a more personal and social interaction. As André's and Rist's study (2000) suggested, people can learn more about a subject matter if they are willing to spend more time with a system. In the context of robotics education, the use of animated persons attract students to spend additional time on working with the robot.

The lifelike agents have been used in various educational applications. For example, they can be used to animate Web lectures and have been shown to be useful tools for delivering audio-video material on the net. They are very easy to program and are ideal when the user needs a guide. An animated agent can easily replace a text-based online help system.

For the purpose of the research reported in this paper, Microsoft Agent was chosen from a list of available interface agent development tools because it offers a wide range of useful features. There is a pre-existing library of animations as well as text-to-speech and vocal language interaction capabilities.

According to the Microsoft Agent Web Page, "Microsoft® Agent is a software technology that enables an enriched form of user interaction that can make using and learning to use a computer, easier and more natural". The technology is not available on the Macintosh operating system. With the Microsoft Agent set of software services (currently version 2.0), developers can easily enhance the user interface of their applications and Web pages with interactive personalities in the form of animated characters (Fig. 3): Merlin, Genie, Robby, Peedy.

Merlin is older and therefore is wiser, more serious and rational. He behaves like a real wizard using a magic wand and talking in a magician's way.

Fig. 3. Microsoft animated agents

Genie is in a certain way similar to Merlin. Peedy is a parrot that needs to be taught. Since he is young, he is curious and funny, but also inexpert. It has more interactive abilities and options than the other three agents. Robby the robot has a serious steel voice and is much more indicated to technical content and applications.

These characters can move freely within the computer display, speak aloud (and by displaying text on screen), and even listen for spoken voice commands. This ability to recognize speech is especially exploited in this application. When used effectively with a conversational interface approach, a Microsoft agent will broaden and humanize the interaction between robots and humans. In summary, it brings animation, interactivity, and versatility, all of them with high pedagogical impact.

3.3 Experimental set-up

The application reported in this paper is a part of DLAB strategy in teaching autonomous robotics. It is used for local control of one robot using voice. It illustrates how to integrate the conversational abilities of a Microsoft software agent in controlling a Khepera robot and allows to test simple control commands (back, forward etc.) and more complex control algorithms developed in Matlab. An animated agent is helping the student in using the interface. For this application, the conversational agent is the Robby Microsoft agent, while the tutor agent is Merlin.

System requirements for the reported application are:

- Microsoft Windows® 95 or later.
- Internet Explorer version 4 or later.
- A Pentium 100-megahertz (MHz) PC (or faster).
- At least 16 megabytes (MB) of RAM.
- At least 1 MB free disk space for the Microsoft Agent core components (version 2.0 available free on the Internet).
- An additional 1.6 MB free disk space for the Lernout & Hauspie® TruVoice Text-To-Speech Engine for speech output.
- A Windows-compatible sound card.
- A compatible set of speakers or headphones.
- A high quality microphone.
- Microsoft Visual Basic Runtime.
- Microsoft Access 97 or 2000.
- Matlab 5.3 or later.
- KMatlab functions and KiKS. kMatlab is a set of Matlab routines that permit the user to interact with Khepera over a serial connection. It includes Windows DLLs to perform the system-level serial-port communication, and a library of useful Matlab .m files to read sensors, set speed, etc. Kiks is a Khepera simulation environment (Nilsson, 2001) (evaluation copy freely available on the Internet).

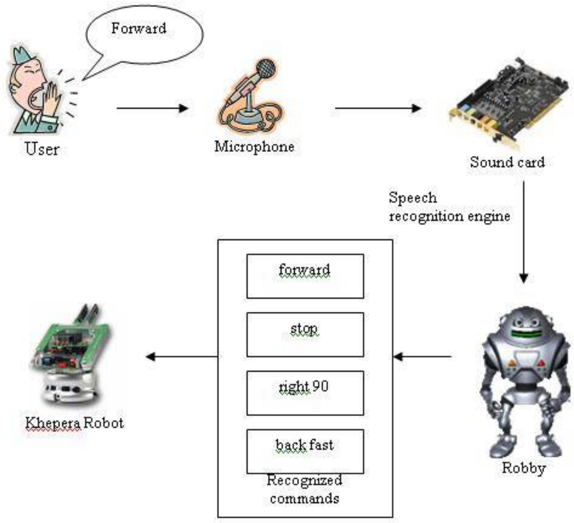

Fig. 4. Experimental setup of the man-robot interface

The structure of the experimental setup is presented in Fig. 4. The input device is a microphone. Speech recognition accuracy is much higher if a high-quality headset microphone is used. Otherwise, the ambient might bury the commands given. The speech input is recognized by the speech recognition engine and executed by the robot. After installing the speech recognition engine, it may be useful for the user to go through a series of exercises in order to enhance speech recognition accuracy.

The application allows the user to give basic ("forward", "right", etc.) or complex commands (a basic command plus an option, e.g. "forward fast", "right 90") to the robot by using voice.

An interesting contribution of this interface is that the user can give voice commands that correspond to Matlab® control programs, so any control algorithm that was developed in Matlab and tested previously in a simulation environment (Nilsson, 2001) can be executed as a result of a voice command.

The Visual Basic loading procedure for Robby is presented:

```
frmStart.RoboAgent.Characters.Load "Robby",
"Robby.acs"
Set RoboAg =
frmStart.RoboAgent.Characters("Robby")
' show the agent on the screen
RoboAg.Show 10
```

The corresponding loading code for the help agent is:

```
Private Sub mnuLoadHelpAg_Click()
If HelpAg Is Nothing Then
frmStart.HelpAgent.Characters.Load "Merlin",
"Merlin.acs"
Set HelpAg =
frmStart.HelpAgent.Characters("Merlin")
HelpAg.Show
```

```
'MsgBox "Help agent already loaded!",
vbInformation, "Help ..."
End If
End Sub
```

The agent is easy to install and configure. One need to freely download (see the Microsoft agent website) three packages: Microsoft agent system, the Microsoft command and control speech recognition engine, and the Text-To-Speech engine.

The agents ship as part of the standard operating system beginning with Windows 98 Second Edition. Some custom Windows installations may have not Microsoft Agent installed.

The application is developed in Visual Basic (VB) so the basic movement commands and the corresponding options are implemented in VB. The database containing the commands and their possible options was implemented in Microsoft Access 2000, whose facilities were enough for this application.

The application has two forms:
1) The main one, by which one can give voice commands to the robot (Fig. 5);
2) The second one, in which the user may configure the commands according to his needs and preferences (Fig. 6).

Both are described elsewhere (Buiu and Dumitrascu, 2004), where full technical details are given.

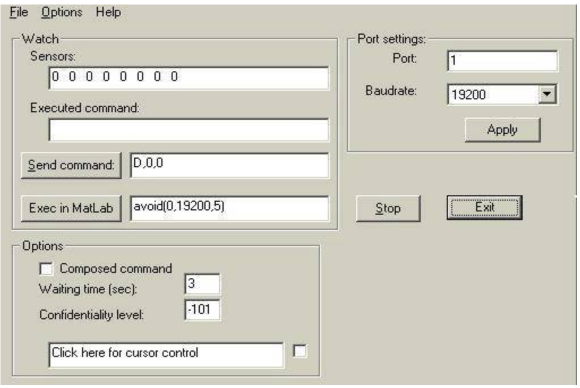

Fig. 5. Main command form

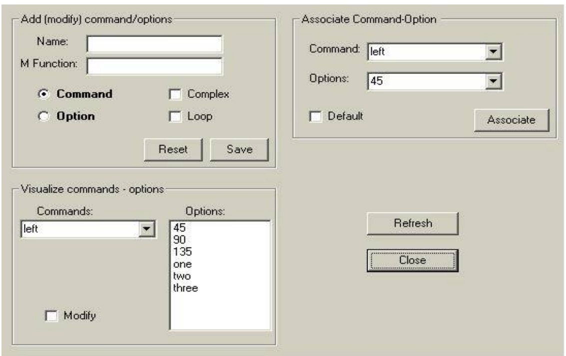

Fig. 6. Command configuration form

4. STUDENT EVALUATIONS

First experiments at DLAB show that the conversational interface works well, the quality of the speech recognition process being influenced by the environment and by the way the user is pronouncing the words (for example, the commands are not to be pronounced very quickly one after another).

After the first experiments, it’s obvious that the whole experimental setup is very attractive to the students and has an important educational impact. The animated agents, with integrated speech recognition and synthesis are adding new dimensions to the control experiments providing an important benefit: motivation. A human-like presence can motivate the students to interact more frequently with the robots.

Various experiments have studied the impact of animated pedagogical agents, but none exists in the field of robotics education. The *persona* effect usually designates a strong positive effect on students’ perception of their learning experience due to the presence of an animated pedagogical agent. Some experiments were conducted in this direction (Baylor, 2002b), but little is known regarding the value of presenting more than one pedagogical agent to the learner simultaneously.

To empirically investigate the students’ perception of animated agents in robotics control experiments, 20 undergraduate students (90% male and 10% female) enrolled in a cognitive robotics course participated at the evaluation of the interface. Cognitive robotics is concerned with endowing robotic or software agents with higher level cognitive functions that involve reasoning, for example, about goals, perception, actions, the mental states of other agents, collaborative task execution, etc.

The participants had to use various ways to control a Khepera robot. They had to solve the same set of tasks. First they used a Matlab program to control a simulated robot (Task 1). The implemented commands are simple (back, forward etc.) and were followed by a simple obstacle avoidance program. After that, they used the same Matlab programs to control a real robot (Task 2). In the end, they used the reported conversational interface to give a real robot simple commands (as above) and to make it avoid obstacles (Task 3).

The objective was to find students’ assessments of the animated agents’ presence on a number of dimensions. They had to ask several questions as follows:
Q1. How much time did you spend with the program? (minutes)
Q2. How do you rate the help system (availability, modality to present information)? (1..10)

Q3. How motivated you feel to make the experiments? (1..10).
Q4. Which one of the three educational setups would you show to a fellow beginner robotics student?
Q5. Do you feel motivated to contribute to further versions of the program you used? (1..10).
The first preliminary results are shown in Table 1 (average values).

Table 1. Student evaluation of the conversational interface

	Task 1	Task 2	Task 3
Q1	21	39	45
Q2	7.15	7.75	8.45
Q3	8.35	8.55	9.15
Q4	3	5	12
Q5	8.25	8.45	9.25

The first findings reveal that animated agents, with speech recognition and synthesis abilities, have an important positive impact on students. The students spend more time on average with the conversational interface, while the simulation experiments tend to require less experimental time. As first results show, students may choose this kind of interactive robotic control frequently and for longer periods of time.

The help offered by the animated agent was rated higher than the usual written support. Responses to the third question show that human-like presences in the control experiments can bring important benefits on the motivational level. The students are more readily to show the conversational interface to other fellow students than other kind of experiments. They feel also much more motivated to contribute themselves to further versions of the conversational interface.

5. CONCLUSIONS

A conversational man-robot interface has been developed in order to be used for teaching basic concepts of autonomous robotics. The program uses an animated Microsoft software agent whose speech recognition and synthesis capabilities are utilized. This allows the user to give simple and complex predefined voice commands to a Khepera robot. New commands can be added at any time using Matlab functions. Another animated Microsoft agent is used for guiding the student.

Further implementations plan to add learning capabilities to the agent. In what extent the agent can learn a certain user and recognize it in different environmental situations, it is still an open question.

We plan to extend the code in order to allow voice command of a remote robot and of a team of two robots that is already used in remote collective experiments. The interface will be tested and further refined using the feedback of more students in robotics.

ACKNOWLEDGEMENTS

This work was supported in part by the Swiss National Science Foundation and Romanian Government under CNCSIS grants.

REFERENCES

Albu, M., F. Mihai, C. Buiu and D. Gillet (2001). Enhancing the Internet-based control in teaching laboratories using software agents. In *Preprints of IBCE 2001*, Workshop on Internet based control education, pp. 47-52, 12-14 December, Madrid

André, E. and T. Rist (2000). Presenting through Performing: On the Use of Multiple Lifelike Characters in Knowledge-Based Presentation Systems. In *Proceedings of the International Conference on Intelligent User Interfaces*, IUI ' 00, New Orleans

Baylor, A. L. (2002a). Agent-based learning environments for investigating teaching and learning. *Journal of Educational Computing Research*, **26(3),** 249-270.

Baylor, A. L. (2002b). Expanding preservice teachers' metacognitive awareness of instructional planning through pedagogical agents. *Educational Technology Research & Development*, **50(2),** 5-22.

Baylor, A. L., and J. Ryu (2003a). Does the presence of image and animation enhance pedagogical agent persona? *Journal of Educational Computing Research*, **28(4),** 373-395.

Buiu, C. (2002). A distributed laboratory for automatic control and robotics education. In: *Proc. IEEE-TTTC AQTR 2002 (THETA 13) Int. Conf.*, Cluj-Napoca, Romania, 2002, Vol. II, 1-6.

Buiu, C. and A. Dumitrascu (2004). A conversational interface for voice control of autonomous robots. *Periodica Politehnica, Transactions on Automatic Control and Computer Science*, **Vol.49 (63),** No. 4, pag. 87-90.

Fong, T., C. Thorpe and C. Baur. (2001). Collaboration, Dialogue, and Human-Robot Interaction. In: *Proc. 10th Int. Symp. on Robotics Research*, Lorne, Victoria, Australia.

Fong, T., C. Thorpe and C. Baur (2003). Robot, Asker of Questions, *Robotics and Autonomous Systems*, **42(3-4).**

Marin, R. (2002). The UJI online robot: a distributed architecture for pattern recognition, autonomous grasping and augmented reality. Ph.D. thesis, Jaume University, Intelligent Robotics Laboratory.

Moreno, R., R.E. Mayer, H.A. Spires and J. Lester (2001). The case for social agency in computer-based teaching: do students learn more deeply when they interact with animated pedagogical agents? *Cognition and Instruction*, **19(2),** 177-213.

Nilsson, T. (2001). Kiks is a Khepera Simulator, M.Sc. Thesis, Umea University.