15th Triennial World Congress, Barcelona, Spain

NOVEL INTERACTIVE CONTROL INTERFACE FOR CENTAUR-LIKE SERVICE ROBOT

Jussi Suomela and Aarne Halme

Helsinki University of Technology, Automation Technology Laboratory,
PL5400, 02015 TKK, Finland, Email: jussi.suomela@hut.fi

Abstract: The paper considers the case, where a service robot and human operator work together in the same working area. A novel control interface is presented, which considerably improves the communication and interaction between robot and operator. Commands and supervision are made based on cognition by speech and gestures – the most natural communication for humans. An innovative yo-yo hand controller provides unique control possibilities for all kind of teleoperation from robot driving to manipulator control. Interaction with the task environment is improved by the aid of virtual reality models and novel pointing technology. *Copyright © 2002 IFAC*

Keywords: Human machine interface, robot control, supervisory control, cognitive systems, speech control

1. INTRODUCTION

Development of mechatronic machines has been fast on the level of automation of machine functions, but relative slow in the area of user interfaces. Little of those possibilities provided by today's computer and multimedia software technology have been utilized. Under-development of the user interfaces means that all potential benefits of mechatronisation cannot be utilized because the human operator is not able to fully use his strong cognitive abilities like understanding the environment, the work problem and planning related operations. Instead, he has more or less to concentrate on steering the machine. User interfaces are constructed still on the traditional idea of commanding and monitoring single operations rather than controlling task execution. Improvements are made mainly in ergonomics. New computer-controlled machines are coming, however, more and more robotic like (development e.g. in modern construction or agriculture machines) and the user should be able to move to a more comprehensive level of planning and monitoring tasks (supervisory control). In this level his brains and senses are utilized most effectively in the human-machine process. This doesn't mean that the direct basic level controls are not needed or used at all, but the user can be freed from those operations if and when the machine knows how to proceed with the task. In practice such needs are already coming up when teleoperation is presently being taking use e.g. in mining machines (Suomela and Halme 2001). The machines are not teleoperated through the whole work cycles, but run in automatic or semi-automatic modes as much as possible. This improves the work environment and one operator may control several machines, which in turn improves efficiency.

The current digital multimedia technology offers versatile environment to develop new type of user interfaces for robotic machines. Machines can also be equipped with more sensors and senses, which give information concerning behaviour of the machine and the task environment. This gives possibility for the user to guide the work by "talking" with the machine (in a cognition level) by using concepts and symbols, which are perceived and understood by both parties. Research on this kind of new user interfaces is topical especially in robotics, but the results are beneficial and can be applied also in more traditional work machines. Up to now the relative high prices of multimedia technology has limited its use in practice, but there has been a rapid change in this situation when PC-technology can now provide cheap and powerful means to almost all applications.

2. NOVEL INTERACTIVE HMI

In traditional supervisory control (Sheridan 1992) the machine is doing a pre-taught task and operator is supervising the work. In case of fault or unexpected situation operator will guide or even take the machine under direct teleoperation. Task teaching is usually very near to the traditional robot programming. The commands to the machine are given either through keyboard (supervisory control) or through special devices like joysticks (direct teleoperation).

Traditional control is enough for robots working in repetitive tasks, like most workmachines do. New generation service robots, like WorkPartner service robot introduced later on, are designed for varying tasks in varying environments. Interaction with the user is essential. When robot and operator are both working it is also clear that operator cannot carry heavy control equipment. The easiest way to decrease the amount of equipment is to improve the autonomy of the robot and use human like communication methods like speech, gestures (Fong et al. 2000) or even brainwaves (Amai et al. 2001).

In our novel interactive HMI the environment modelling and task definition are made interactively by the operator and the machine. This process is possible only when the operator and the machine are on the same level of cognition and presence, i.e. they understand the environment and the target objects similarly. This is why the interface is also called cognitive HMI. Augmented reality provides excellent tools to combine the human and robotic perception and cognition. It is used especially during modelling (Halme and Rintala 1997) and task definition.

In addition to the artificial control interfaces, normal muman communication means like speech and gestures are used as much as possible

Although the collaborative way of control makes it easy to teach new tasks to the robot the "pre-taught" matters are important. The larger the robot's know-how is the easier it will be controlled (like it is more to control a trained dog than an untrained one). Once a working place is modelled, new objects or tasks taught to the tele-robot they should be always available in the future. The work tasks, environmental models and visual models of objects are saved in an Internet database where they can be downloaded online when needed. This database can be divided between several machines irrespective of their location.

The prototype user interface demonstrates new possibilities in human–machine interaction, like commanding tasks by analysing and advising operations rather than controlling individual motions or sequences. The interface also supports machine learning under operator supervision during working. The test platform is a service robot, called WorkPartner (Fig.2), designed for interactive working with humans in outdoor environment, like yards, parks construction sites etc

3. SPEECH AND GESTURES IN INTERACTIVE ROBOT CONTROL

Human cognition receives sensor information from all five senses. In the perception the vision is the most important sense receiving most of the data. In communication hearing is the most important sense because of sound based communication. Most of the information is transmitted in the spoken form, but also visual information like expressions and gestures are important and informative in human communication. In fact, it has been shown recently by neuro-scientists that humans have a long history in evolution for such kind of communication, because brains have specialized cells, so-called "mirror cells", which are activated when communicating with expressions or gestures. When designing a natural communication between human and machines, which is easy for human cognition, speech and gestures (vision) should be thus looked at the first place.

3.1 Speech

The oral communication is one thing that separates human from animals. It's the most efficient way to change information between humans. Spoken commands have already been used quite long in machine control. Even mobile phones can be today controlled by speech. Speech control is typically limited to few commands, which usually must be given by certain person whose voice has been "taught" to the machine. Even today the speech recognition systems have difficulties to recognize voices of different people without pre-training. The changing background noise especially in the outdoor applications makes the situation much worse. Also the recognition of free speech with all possible words is difficult.

If the recognition of words is difficult for a machine the understanding of the meaning of the sentence is almost impossible. This is why the control is usually limited to certain commands. In most cases there is no need to converse with the robot and a limited amount of commands is enough.

In our cognitive HMI prototype system spoken commands are used for two purposes: to change robot states and to give tasks to the robot. State commands are usually one - two word commands introducing the new state like "manipulator teleoperation". Task commands are typically consisting two or more words, which define the task, object and possible target position for example "bring that box". Task commands start always with a verb in imperative mode. From the verb the robot can conclude the needed amount of additional definitions. For example command "bring" means that the object is taken to the operator and needed definitions are the object to be brought and its location. The word "that" means that operator is further defining the object e.g. by pointing it, in which case both the object and its location are defined. All commands are started with an initiation word "Partner" to avoid noise commands, which are typical for speech recognition systems. When robot has understood a command, it signs for by speech or/and by a simple gesture like nodding the camera head or waving a hand.

3.2 WoPa language

WorkPartner's communication system is based on two languages and an interpreter between them. The commanding language – WoPa Language – is used between an operator and the robot and is based on combined spoken and gesture commands. The internal language is used inside the robot control system to configurate, initiate and command different services of the robot subsystems, like locomotion, navication, perception, etc. Manager, which is the highest-level always-running task in the control system, includes an interpreter, which breaks down the command sequences coming trough the operator HMI. Assisting by the planner it further code the actions into the internal language, as illustrating in Fig.1.

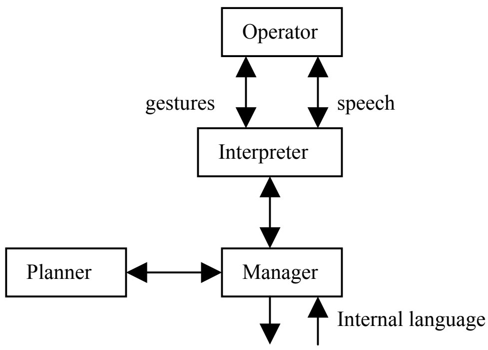

Fig. 1. WorkPartner language processing

In the following we consider in detail only the WoPa - language.
Interpreter receives a spoken command, breaks it up to primitives and checks that the syntax of the command sentence is sensible. If the syntax is accepted, command is transferred to the manager, which checks wheather the commands include all the information needed to perform a task. If information is missing, the robot asks for the missing data. At the same time gestures given by the operator are detected either by the calour camera or control yo-yo hardware (see ch. 4.2). For example the command "Partner, bring that ball" will produce the following primitives:

Bring

Primitive information	Verb: Drive to object, grip object and lift it, drive to operator
Needed addition definitions	Object and its position
Example	bring that ball

That

Primitive informaation	Pointing: Operator is pointing an object or a location, read pointing information
needed addition definitions	Name of the object
Example	bring that ball

Ball

Primitive information	Object
Possible adjectives	Color (red)
Properties	Can be gripped from any direction, gripping in the middle of the ball (great circle)

Both the interpreter and the manager would accept this command because the syntax is right and it includes all needed information for a task execution. In case of command "Partner, bring ball" the interpreter accepts the command because the syntax is right and sends it to the manager. The manager checks it's world model database to find out if the object "ball" is unique and does its specifications include data concerning the location. If not, the manager asks the user more information. In case of command: "Partner, bring", the interpretor notices the missing object and asks it directly from the operator without sending the command to the manager.
If syntax or information content were defective, the interpreter would ask the missing information from the operator by speech.

3.3 Gestures

Although the speech is the most efficient and used method to communicate between humans, expressions and gestures are very important part of the conversation when the debaters are face to face. According to the latest studies the human brains have specialized cells called "mirror cells" which react to the expressions and gestures. This explains the very good recognition of gestures in all communication situations.
There are also lots of situations where gestures are the only way to communicate. Deaf people can communicate with gestures like hearing people with speech. In noisy environments and situations where speech can't be used, gestures are used to communicate with the co-workers. Typical branches where plain gestures are used to inform or control other people are military, construction and vehicle control. Each area has their own gestures, but they have much in common. Most of the gestures used in human communication are dynamic. There are two reasons: human reacts to movements even they happen apart from the focus area of vision and the amount of possible gestures is almost infinite when they include movements.
Traditionally the gestures (Fong et al., 2000) or gazes (Heinzmann and Zelinsky, 2001) have been tracked

with image processing. However, image-processing systems are not optimised for fast motion recognition and they have difficulties to work in outdoor conditions with changing illumination. More over, the distance between the operator and robot can be long, which makes visual recognition difficult. Because of these reasons, we are also studying an alternative way to signal gestures to the robot, namely use mechanics of the HMI, see Ch 4.2.
Gestures are useful especially in situations where background noise is too high for spoken commands. However, from purely cognitive point of view it is propably the easiest way for the operator to use mostly speech and to have response from the robot by speech and gestures.

4. HARDWARE

4.1 WorkPartner – Centauroid service robot

The WorkPartner robot (Halme *et al., 2000*), (Halme *et al., 2001)* is shown in Fig 2. It is a mobile two-hand service robot, which moves by a hybrid locomotion principle. The robot has four legs equipped with wheels and an active body joint. The weight is about 200 kg and the payload about 40 kg. The actuation system is fully electrical and the power system a hybrid one with batteries and a 3 kW combustion engine. The locomotion system allows motion with legs only, with legs and wheels powered at the same time or with wheels only. With wheels, the machine can obtain 7km/h speed on a hard ground. The purpose of the hybrid locomotion system is to provide a rough terrain negotiating capability and a wide speed range for the machine at the same time.
The platform is equipped with a two-hand human like manipulator system, as illustrated in the Fig. 2. The manipulator is a lightweight construction (design weight about 30 kg) and can handle loads up to 10 kg.

Fig. 2. WorkPartner robot

WorkPartner is a large-scale R&D project done in three phases in 1998-2005. The project is public and can be followed on the Web site:
www.automation.hut.fi/IMSRI/workpartner

4.2 Human Machine Interface (HMI)

Studying and developing of Cognitive Human Machine Interfaces is one of the subprojects related to WorkPartner project. The HMI hardware under development is divided into three parts: Operator hardware, Machine hardware and Home base hardware, see Fig. 3. These parts are linked together with WLAN communication.
Operator. As explained earlier, communicate with WorkPartner is done by speech and gestures. There is an option to control the robot without any additional hardware on the operator. Spoken commands can be given to the onboard microphone and static gestures are noticed by the image processing. However, it is obvious that this is possible only in very good conditions, usually valid only indoors and in short distances. For more realistic environment conditions special operator hardware is needed.
Operator hardware is included in so called operator's coat, shown in Fig 4. The coat itself is a bright colour textile including the following hardware: PC, PDA, Speechmike^{TM} and yo-yo controllers.
The bright colour of the waistcoat is used to help the machine to recognize and track the operator.

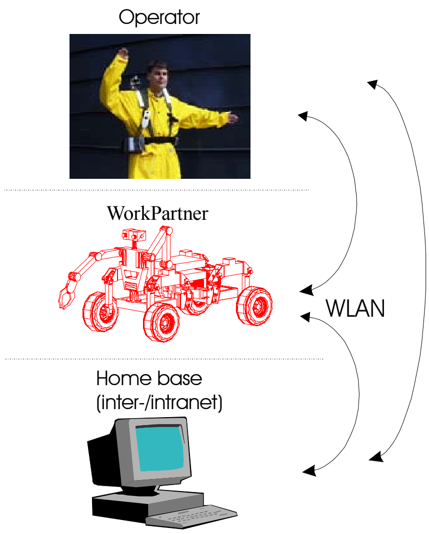

Fig. 3. HMI system structure

HMI computers: The original idea was to utilize a PDA as a main HMI-processor. However, the speech recognition, modelling and need of analog and digital inputs were too demanding for today's PDAs, and a normal laptop was chosen. The GUI of the PDA was anyway considered superior to the laptop GUI in the planned, mobile tasks, and the PDA was utilized as laptop terminal. The communication between laptop and PDA was implemented with Bluetooth. The PDA/PC computer is used for pointing, direct teleoperation of the machine and interactive modelling of the environment. It also processes the data from other devices (Speechmike^{TM}, control yo-yos) and communicates with the machine and home base by using WLAN.

Control yo-yos (Hand trackers) are an innovative taylormade hardware shown in Fig. 5. The device is used for detecting gestures, object pointing, and direct teleoperation of robot motion. Yo-yos are wire potentiometers equipped with 2DoF direction sensors. The ring at the end of the potentiometer wire is simply held on the operator's thumb when he/she moves arms. Wrist positions are calculated from the length and the two direction angles of the wire and transmitted to the robot. In the future the yo-yo handles will also have buttons to control the grippers of both hands.

Fig. 4. Operator hardware

The original idea of these yo-yos was to provide direct manipulator control for the operator. After some tests it was noticed that they are perfect hand controllers for all kind of teleoperation. First of all the 6 (2 x 3) DoF provides possibilities to control complex structures like manipulators. Large movement range allows accurate control in outdoor conditions even when operator is using heavy clothes and gloves. If this large amount of DoFs are not needed it's easy to make a simpler yo-yo. Interesting example is a so-called "Intelligent suit", where 1DoF pull tags are used for communication control (http://www.reimasmart.com/).

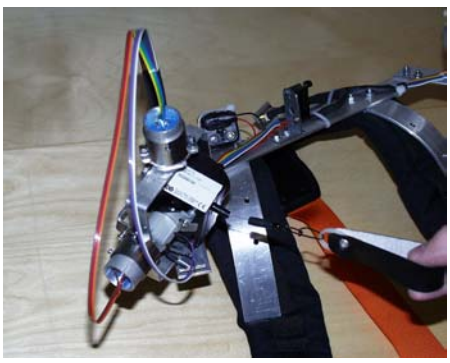

Fig. 5. Structure of the control yo-yo

Speechmike™ (Philips) is used as a control and communication device. It contains microphone, loudspeaker and trackball. Microphone and loudspeaker are used for the verbal communication between operator and machine. Trackball can be utilized for pointing of objects instead of yo-yos. This device can be replaced with a headset microphone when yo-yos are available.

Pointing. Pointing is a typical and effective way for a human to show something to somebody else. However, when pointing something with a hand it is extremely difficult to follow the "hand line" so that the position of the line end can be determined accurately enough. Because of this we have chosen another way for pointing. Instead of showing by hand the operator uses the control yo-yos or trackball to control the camera/laser pointer head of the robot. By turning the visible laser spot to the target object operator can be sure that both him and the machine have common understanding about the target and its position.

In the case of long distance between operator and robot, when the laser spot cannot be seen, the operator uses the PDA computer for pointing. The image from the robot's camera is shown in real time on the PDA and operator can turn the camera by pointing the target from the image (touch screen) and the servo head turns the centre point of the image (optical axis) to the pointed target.

Robot: The onboard hardware consists of camera/laser-pointer head, panorama camera, inertial navigation unit, 2D and 3D laser scanners. 3D-laser scanner is only for the environment modelling and it's not carried on during the work phase. 2D scanner is located in the body of the torso. It is connected to the navigation unit and used both for relative navigation and obstacle/target detection. The head unit is the interface between operator and machine. It includes camera(s) and a laser- pointer.

Homebase: Homebase is any additional hardware connected to the intra/internet. Homebase equipment provides supporting processing power for demanding modelling and image processing tasks and a database for ready environment models, tasks, objects, etc.

5. HMI FUNCTIONS

5.1 Modelling

Execution of the work tasks demands proper model of the existing environment. Primary modelling is made by the aid of a 3D laser scanner using a SLAM (Simultaneous Localization and Mapping) method (Forsman, 2001). In practice an environment, like parking area or home yeard, is scanned from 2 or 3 different locations. The 3D measurement data and the colour images are transferred to the modelling processor, which is placed in the home base. Modelling processor forms first the unified model from the partial 3D scanned data by registering the data sets, and then processes a basic geometrically valid 3D virtual model, which is asimplified occupancy grid model of the environment. Certain clear objects like trees and building walls, are separated and recognized automatically, but most of the other objects are ammended manually by the operator. The virtual basic model and set of related colour image are stored together. Having the basic environment model available the operator views and edit it with his computer by recognizing and adding named objects and other data. The objects and data form symbols, which can be used in the WoPa-language. Editing is possible on-line in the CHMI so that the operator can, for example, delete non-static objects from the model or add objects, which are related to the current task.

5.2 Control without operator interface

When operator interface unit is not in use operator can still use the robot. All the available speech commands can be given straight to the machine if the operator is near enough (like with Sony's Aibo robot-dog). In addition to the speech commands gestures can also be used for the robot control. Simple gestures are recognized by the image processing system of the robot. They include come here/follow me, stop and continue. Operator gestures are recognized by the servo-camera and image processing system.

5.3 Control with the operator interface

Cognitive control. This is the “normal” operating mode. Operator is wearing the coat, which is easy to track by the image processing system. Control is mostly based on spoken commands, which are given to the Speechmike™ device. In case the robot cannot recognize a target object or place, or there are several similar objects, the right target can be pointed by using the trackball or yo-yos and laser pointer in the camera head (see Ch. 4.2). All gestures can also been used during this phase. The robot responds with speech and gestures.
Manipulator teleoperation. In cases where difficult objects have to be gripped or a gripping task is taught, the operator teleoperates the manipulators. Teleoperation interface is formed by yo-yos, which locate on the shoulders of the operator coat. Operator can now teleoperate the manipulators in a very natural way by making the movements with his/her hands. WorkPartner's two-hand manipulator has the size of upper body of a normal size man. The waist of the manipulator body can be controlled by the trackball of Speechmike.
Direct driving control The operator can control also motion of the platform by using the PC. This so-called direct control mode includes both direct teleoperation and path control. In the direct teleoperation the operator can give straight “joystick commands” to the machine by moving the pen on the direct teleoperation window. In the path control mode the operator has a 2D map on the PC. With the pen he/she can draw a path to the map and send it to the machine for execution.

6. CONCLUSIONS

Cognitive HMI is a novel interface for mobile service and work robots. Instead of traditional robot control, control is based on cognitive communication methods between the operator and the robot. The environment modelling and task planning/teaching bring the cognitions of the robot and the human together on the same level in the same space. The communication is designed to satisfy both human and machine cognition. The use of limited amount of spoken commands and static gestures is easy for the robot’s cognition, like the speech and dynamic gestures are for the human cognition.
Development of the CHMI-concept considered is still in its early phase and experiences are limited. The authors see, however, that the outlined methodology is the way to go and results can be extended fairly easily to the design of new HMIs for other machines, like traditional working machines, too.

7. REFERENCES

Amai W., and Fahrenholtz J., Leger C. (2001). Hands-Free Operation of a Small Mobile Robot, *Autonomous Robots*, **Vol 11**, No 1, July 2001

Fong T., Conti F., Grange S. and Baur C. (2000). Novel Interfaces for Remote Driving: Gesture, Haptic and PDA, *SPIE Telemanipulator and Telepresence Technologies VII*, Boston, MA, November 2000

Forsman, P. (2001) Three-Dimensional Localization and Mapping of Static Environments by Means of Mobile Perception, Helsinki University of technology, Doctor Thesis, 2001

Halme, A. and N. Rintala (1997). An interactive telepresence system augmented with VR-models of the working environment. Topical workshop on virtual reality and advanced man-machine interfaces, (CD-ROM), Tampere, Finland

Halme, A., K. Koskinen, V-P. Aarnio, S. Salmi, I.Leppänen and S. Ylönen (2000). WorkPartner - Future Interactive servicerobot, The 9th Finninsh Artificial Intelligence Conference, Helsinki, Finnish Artificial Intelligence Society, pp. 35-42.

Halme, A., I. Leppänen, S. Ylönen and I. Kettunen (2001). WorkPartner - Centaur Like Service Robot for Outdoor Aplications, Proc. FSR2001 Conference, Society of Finnish Automation, Helsinki, pp. 217-223

Heinzmann J. and Zelinsky A. (2001). Visual human-robot interaction, Proc. FSR2001 Conference, Society of Finnish Automation, Helsinki, pp. 225-230

Sheridan T. B. (1992). *Telerobotics*, Automation, and Human Supervisory Control, The MIT Press

Suomela J. and Halme *A. (2001).* Tele-Existence Techniques of Heavy Work Vehicles*, Autonomous Robots,* ***Vol 11****, No 1, July 2001*