

Towards multimodal Human-Robot Interaction in large scale virtual environment

Pierre Boudoin, Christophe Domingues, Samir Otmane, Nassima Ouramdane, Malik Mallem

To cite this version:

Pierre Boudoin, Christophe Domingues, Samir Otmane, Nassima Ouramdane, Malik Mallem. Towards multimodal Human-Robot Interaction in large scale virtual environment. 3rd ACM/IEEE international conference on Human robot interaction (HRI '08), Mar 2008, Amsterdam, Netherlands. pp.359-366, 10.1145/1349822.1349869. hal-00339448

HAL Id: hal-00339448 <https://hal.science/hal-00339448v1>

Submitted on 29 Nov 2008

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Towards Multimodal Human-Robot Interaction in Large Scale Virtual Environment

Pierre Boudoin 1 Nassima Ouramdane¹

Christophe Domingues¹

Samir Otmane¹ Malik Mallem¹

1 IBISC Laboratory University of Evry / CNRS FRE 2873 40, rue de Pelvoux, 91000 Evry, France

pierre.boudoin@ibisc.univ-evry.fr christophe.domingues@ibisc.univ-evry.fr

samir.otmane@ibisc.univ-evry.fr nassima.ouramdane@ibisc.univ-evry.fr

malik.mallem@ibisc.univ-evry.fr

ABSTRACT

Human Operators (HO) of telerobotics systems may be able to achieve complex operations with robots. Designing usable and effective Human-Robot Interaction (HRI) is very challenging for system developers and human factors specialists. The search for new metaphors and techniques for HRI adapted to telerobotics systems emerge on the conception of Multimodal HRI (MHRI). MHRI allows to interact naturally and easily with robots due to combination of many devices and an efficient Multimodal Management System (MMS). A system like this should bring a new user's experience in terms of natural interaction, usability, efficiency and flexibility to HRI system. So, a good management of multimodality is very. Moreover, the MMS must be transparent to user in order to be efficient and natural.

Empirical evaluation is necessary to have an idea about the goodness of our MMS. We will use an Empirical Evaluation Assistant (EEA) designed in the IBISC laboratory. EEA permits to rapidly gather significant feedbacks about the usability of interaction during the development lifecycle. However the HRI would be classically evaluated by ergonomics experts at the end of its development lifecycle.

Results from a preliminary evaluation on a robot teleoperation tasks using the ARITI software framework for assisting the user in piloting the robot, and the IBISC semi-immersive VR/AR platform EVR@, are given. They compare the use of a Flystick and Data Gloves for the 3D interaction with the robot. They show that our MMS is functional although multimodality used in our experiments is not sufficient to provide an efficient Human-Robot Interaction. The EVR@ SPIDAR force feedback will be integrated in our MMS to improve the user's efficiency.

Categories and Subject Descriptors

H.5.2 **[Information Interfaces and Presentation]**: User Interfaces – Interaction styles, I.2.9 **[Artificial Intelligence]**: Robotics, I.3.6 **[Computer Graphics]**: Methodology and Techniques, I.3.7 **[Computer Graphics]**: Three-Dimensional Graphics and Realism.

General Terms

Design, Experimentation, Human Factors.

Keywords

Human-Robot Interaction, Multimodal Interaction, Empirical Evaluation, Usability, Virtual Environment.

1. INTRODUCTION

To increase and to get a more natural Human Robot-Interaction (HRI), several modalities like Augmented Reality and 3D Vision have been successfully utilized. However to increase human performance in 3D Interaction, the design of multimodal systems is required.

One of the issues encountered in designing a HRI system is to avoid puzzling the HRI when a HO switches from one task to another. This constraint is very tough to overcome because a multimodality system implies the co-existence of different devices. The Multimodal Management System (MMS) we have developed permits to change or to combine different devices transparently for the user. This system permits to select the best device when many of them are used or to allow a combination of many devices in order to have a unique interpretation of data (position, orientation, actions event).

To gather feedback about our Multimodal Management System, a preliminary evaluation has been carried out on a teleoperation task involving the ARITI software framework [3] for assistance to the HO and the IBISC semi-immersive VR/AR platform EVR@ for the 3D interaction with the robot. ARITI software uses the Follow-Me technique [4] to select and manipulate objects in the virtual or remote world. Devices used for this evaluation were optical tracked data gloves (2 modalities) and Flystick (1 modality). We have used our Empirical Evaluation Assistant developed in the laboratory in order to carry out the validation of our experiment. This system permits to perform light evaluations during the development lifecycle, leading in most cases to positive final evaluations by ergonomics experts.

This paper is structured as follows. Section 1 will briefly review related work about multimodality. Then section 2 presents the ARITI system and the semi-immersive Virtual Reality/Augmented Reality EVR@ platform used for our experiments. In section 3 we introduce the Multimodal Management System we have designed. Next the EEA system is described in section 5. Finally, section 6 gives insights about the feedback given by EEA.

2. RELATED WORK

We have studied different multimodal frameworks, especially devoted to devices management. Two drew our attention:

The Sylvia Irawati's team [1] proposed a complete framework. The most interesting part is their using of objects ontology to make the interaction more natural and user-friendly. Moreover the object ontology they've proposed supports constraints definitions. This point is very interesting because a robot has got constraints on its articulations and could have some others constraints on the environment in which it evolutes. So, it could be interesting to be inspired by some parts of their framework.

The Ed Kaiser's team [2] worked on mutual disambiguation. Their work testified an interesting approach to manage multimodality with the use of what they called Multimodal Integrator. The aim of this integrator is to find the best multimodal interpretation with the preliminary rated inputs. The principle is to unify inputs data in:

- Amalgamating redundant or complementary data via a logical test set;
- Taking care about the spatiotemporal aspect of data;
- Taking care about data's hierarchy.

This work has inspirited us for designing our Multimodal Management System (MMS).

3. SYSTEM DESCRIPTION 3.1 Hardware and software of the EVR@ platform

IBISC laboratory owns a semi-immersive VR/AR multimodal platform called $EVR@¹$. It permits stereoscopic display, wireless hand/head gesture tracking and force feedback. At IBISC laboratory we have ART optical tracking system, SPIDAR [9] and 5DT Data Gloves. Each device is associated to a specific server which is accessed via the C++ *VRPN* library by clients. The interactivity between the user and the VE is done by using *Virtools 4.0* as a front-end. *Virtools* is efficient software for prototyping and testing 3D IT or HRI because it offers a fast and graphical way to compute them and link them with hardware devices and VEs by connecting specific building blocks to each other.

3.2 The ARITI software framework

 \overline{a}

ARITI (Augmented Reality Interface for Teleoperation via Internet, see [3]) is a software framework assisting the HO in complex and repetitive teleoperation tasks.The teleoperation of the remote robot is accomplished in two steps where Mixed Reality and Virtual Guide paradigms are used (see fig. 2).

• Mixed Reality: First, the HO interacts with a local virtual model of the remote robot, called "phantom robot" (PR), in the master site. Second, the remote robot moves to the PR position after validation of the HO. This permits the HO to predict the position of the remote robot without being puzzled by the transmission delay thru the internet.

• Virtual Guides are used when PR is approaching an object to be selected and manipulated by the robot. They offer a good compromise between freedom of movement given to the HO and reduction of the power of concentration needed in the critical parts of the teleoperation task.

Besides theses assistances, the HO is also proposed Computer Supported Collaborative Work (CSCW). The principle of CSCW is to propose tools for communication, share or exchange of data, in the aim of a common achievement for the users. The ARITI framework may be used either on a master site consisting of a simple PC with a mouse and a keyboard thru a web interface (see fig. 3) or the IBISC semi-immersive VR/AR platform EVR@.

Figure 1. The 6DOF FANUC Robot used in our experiments.

Figure 2. An operator using the 6DOF robot with Spidar, Data Gloves and optical tracking on the EVR@ platform within the ARITI framework.

¹ For more details, see the EVR@ WEB site at http://evra.ibisc.univ-evry.fr/

Figure 3. The web interface of the ARITI framework used to control the 6DOF robot with keyboard or mouse.

3.3 Presentation of the Follow-Me technique

The *Follow-Me* technique (see [4] for the specification of the model) is the 3D interaction technique used within the ARITI framework. When one grabs an object with his hand, his arm has a fast and roughly precise movement far from the target, a precise and slower movement when approaching the target and a grabbing strategy associated to the geometry of the object. Each of the three steps may be associated to a particular interaction strategy and may have what we will call a specific interaction granularity. The Follow-Me technique takes inspiration from this simple grabbing example. It owns two main characteristics:

- the VE is divided into three zones (see fig. 4) in which the interaction has its own granularity, whether one wants to navigate freely and realistically in the VE (free manipulation zone), approach more securely to a target but without loosing any degree of freedom (scaled manipulation zone) and finally approach to the target and manipulate it easily with high security (precise manipulation zone);
- In the precise manipulation zone, virtual guides are used to handle both precision and security of manipulation, which induces a loss of freedom for the user. The aim of a virtual guide is firstly to anticipate the most probable action of the user into the VE and then to make him perform his action as simply as possible with high precision. In order to achieve this goal, the virtual guide puts limitations to the user's possible actions in the virtual world so that he has to *follow* the virtual guide with a specific subspace of VE (with a specific orientation).

4. MULTIMODALITY MANAGEMENT SYSTEM

In order to manage multimodality, we define a main component: *Smart Core* composed of two specific components: *Smart Tracking* and *Smart Action* (see fig. 5). All these component were developed in Virtools 4.

Figure 4. The 3 areas of the Follow-Me model used in the ARITI framework.

Figure 5. MMS *Smart Core* **component.**

The first component: *Smart Tracking* manages the tracking of an effector from one or many 2D or 3D pointing devices, such as optical tracking, SPIDAR, mouse … and returns a position and an orientation. The second: *Smart Action* manages actions triggering independently from the device(s) used.

In fact, the *Smart Core* component allows operator not to be puzzled about the device or combination or device he can use, because the device management is transparency to his eyes. Indeed, multimodality system can be seen in two different ways:

- One consists on having a system, which works with many devices but which are not used at the same time (case 1).
- Another consists on having a system which supports simultaneous combination of devices (case 2).

The *Smart Core* component is able to work in both cases. In order to know in which situation we are, *Smart Core* component just have to look if there is only one device activity (i.e. the variations of position, orientation or states) (case 1) or many devices activities in the same time (case 2). This fact is very important because operator should be able to choose his favorite device(s) and so to be the most efficient and at ease with the application. In other words, *Smart Core* could be considerate like a virtual device resulting from unification of the data incoming from the different devices (see fig 6).

Figure 6. The Smart Core component is assimilated to a virtual device.

4.1 Multimodal Tracking Management

The *Smart Tracking* component produces the Multimodal Tracking Management. In fact, we choose to use a temporal management because it allows us to solve the three main problems we identified, namely:

- What pointing device should I listen if there are many tracking devices activities?
- How to make coincide referential of each pointing devices to an absolute referential?
- How to compensate a loss of tracking from a pointing device?

First of all, we rate the accuracy of each pointing device we had, to get a priority list. This list indicates us what device we must listen in priority. Obviously optical tracking was the first of this list due to his accuracy and then comes the SPIDAR [5] (less precise 3D pointer) and the mouse (precise but 2D pointing device).

 In order to answer to the first question, we search for activity incoming from each pointing device in the order of the list. As soon as activity exists on the position and orientation from a pointing device, it means that the operator is using this device. So we listen the pointing device which gets the "best" activity (i.e. which has the highest priority). Figure 7 shows the algorithm of the *Smart Tracking* component and figure 8 represents an example of the *Smart Tracking* activity resulting from the unfolded algorithm.

In order to answer to the second question, we firstly decided that optical tracking referential should be the absolute referential due to the extremely high accuracy of this device. So when other pointing device are used we calculate their new referential to make coincide to the absolute referential.

In order to answer to the last question, *Smart* Tracking integrates a device switcher, which allows to switch from a pointing device to another when there is a loss of data in the tracking process. This switching is made possible by the mapping of every pointing device referential.

Figure 7. *Smart Tracking* **algorithm.**

Figure 8. MMS *Smart Tracking* **activity.**

4.2 Multimodal Action Management

The *Smart Action* component has been created in order to manage control commands or action given by the operator such as click, button pressed or gesture. *Smart Action* returns an action independently from the hardware. For example, in our application, if data gloves are connected, the "close the hand" gesture will be equivalent to click with a mouse. The associations between gestures and buttons are defined depending on the needs of the application. These correspondences have been hard coded into the *Smart Action* component, but we plan that they will be dynamically loaded via an XML file.

5. EMPIRICAL EVALUATION SYSTEM

Ergonomic evaluation is a mandatory step to detect usability problems for creating intuitive and transparent human computer or robot interaction. Whereas their exist guidelines based on predictive models to build effective 2D HCI, it is not the case for 3D HRI. The main reasons are: no strong models, new interfaces and devices, fewer experts. Due to a lack of norms and ergonomics experience, analytical approaches (predictive models) cannot be used to evaluate 3D HRI. Empirical approaches should be used to gather feedback about our MHRI [6].

5.1 Specification

Empirical Evaluation System (EEA) is intended to be used during the development lifecycle of a 3D Interaction Technique, HRI or MHRI by non experts of ergonomics. Typically, it is dedicated to developers. It has no aim to bypass a complete evaluation process made by ergonomics experts. The main objectives of the EEA system are:

1) To assist experimenters *before the experiment:*

-Help for selection of pertinent variables to be traced during the experiment and submitted to statistical analysis after the experiment (correlation detection, hypothesis testing by using ANOVA [6]);

-Help for selection of known or personalized protocols to be applied in the experiment;

-Help for selection of known or personalized qualitative questionnaires [7] given to the users.

This assistance is carried out by using a *database, which* centralizes the knowledge acquired during past experiments.

2) To assist experimenters *during the experiment:*

-Trace of pre-selected variables during the whole experiment in a log file;

-Real time display of pre-selected variables.

The aim is to permit an *easy debugging* and to detect erroneous behaviors of the users.

- 3) To produce a feedback about the studied *after the experiment:* -Results of statistical analysis made over traced quantitative variables and qualitative variables (questionnaires);
	- -Possibility to replay the experiment off-line;
	- -Integration of the whole experiment results in the *database*.
- 4) To permit collaborative work over the *database* data:
	- -Share the results of experiments with experts outside the Lab; -Annotate the experiments.

5.2 Software architecture of the EEA

In order to achieve the specification of the EEA system, we have built two distinct tools and utilized existing free software. The global architecture and software implementation is given in figure 9.

The first tool is dedicated to *Experimental Protocol Conception*, which we call *EPC* tool. It includes paragraphs *1.* and *4.* of our objectives. The *EPC* tool permits the access to the *database*. We fulfill paragraph 4. By choosing a WEB based architecture centered on an *Apache 2* server. The database is implemented with a *MySQL* server, which is accessed via *SQL* queries from the *EPC* tool written in *PHP* and *AJAX*.

The second tool is dedicated to *Measurements and Debugging*, which we call *MD* tool. It includes paragraph *2.* of our objectives. It has been implemented by making specific *Virtools building blocks* that we call *Probes*. The probes may be connected to building blocks which output has to be measured, traced and displayed in real time. Figure 10 shows five probes connected to the tested 3D IT given in section 6. They permit to measure the duration of a user's experiment and how many mistakes he has made. A *Core* component permits to initialize the measurement schema of all pre-selected variables by using a configuration file created by the *EPC* tool before the experiment.

Figure 9. Architecture of the *Core* **module of the** *MD* **tool. The boxes at the right correspond to the preselected variables the experimenter wanted to trace.**

Figure 10. View of five probes connected to ARITI main blocks (Object Selection and Selection) in the *Virtools 4.*

It also permits to synchronize the data gathered by the different probes, by using dedicated modules (see fig. 11). The *Synchronization* and *Wait* modules permit to synchronize probes and core. Probes send synchronization messages to these modules. When the synchronization is done the core launches a module (e.g. "Speed" or "Acceleration") for computing speed or acceleration of specified object on the virtual environment.

The fulfillment of the objectives of paragraph 3 is done by using the log files produces by *MD* tool during the experiment and configured in the *EPC* tool before the experiment. Results (correlation detection, hypothesis validation) are produced by *R Software* scripts, which had read the log files.

Figure 11. Software architecture and implementation of the EEA System. Statistical analysis is done by the *R* **software from the log files generated by the MD tool.**

5.3 Experimental protocol design with EPC tool

Three steps are required for creating your own experimental protocol. The first step consists in choosing properly the indicators and variables (the main ones are described in [6], [7] and [8]). The experimenter is helped during this process. Each parameter (indicators and variables) may be associated with many publications, examples and different help messages. Variables and indicators are gathered in different categories. The next step is the creation of evaluation scenarios. The experimenter chooses the good disposition of the different parameters. The final step is the questionnaires' step. The experimenter has to choose the automatic or semi-automatic mode. The automatic mode permits to create questionnaires directly by the parameters chosen. Semi-Automatic creates questionnaire directly but the experimenter may modify, add or delete questions.

6. PRELIMINARY EVALUATION

EEA system has been used for the first time to gather feedbacks about a 3D IT developed at IBISC laboratory, called Follow-Me [4], which is still under study. Results show that EEA has permitted to refine some characteristics of this 3D IT. In this article, we apply the EEA system to perform an evaluation on our MHRI for the ARITI system on a robot teleoperation tasks. We have experienced our EEA system to check the easiness of use of the tracked Data Gloves comparing to the Flystick for selection and manipulation with the Follow-Me 3D interaction technique implemented in the ARITI framework. Position and orientation of the robot's grip are obtained by optical tracking. With Data Gloves control the user must close his fist to close the grip but with the Flystick users must use a button.

6.1 Experimental protocol

In our experiment (see fig. 12), HO controls the virtual model of our FANUC LRMate 200i robot with two kinds of Interaction Device: Flystick or Tracked Data Gloves (Hand tracking and Data Gloves). The SPIDAR was not used. The task to accomplish in this experiment is to move, as fast as the HO can, one of the three objects (square, circle or triangle shaped) from their start location

Figure 12. *A HO realizing the experiment by using a Flystick***.**

to their specific arrival location, using Tracked Data Gloves then the Flystick.

Before performing a trial of the experiment, objects to be selected and devices to be used are randomly chosen. (For devices: Flystick and Data Gloves or Data Gloves and Flystick and for the objects : Blue, Red, Yellow; Red, Yellow, Blue; Yellow, Blue, Red).

Indicators variables are the selection and manipulation time, the number of errors number made on selection and manipulation. Comfort, satisfaction and device usability are collected by a questionnaire at the end of session. We also gather information about the skill in VR practice, age, gender and handedness.

6.2 Results and Analysis

8 users (4 men, 4 women) have utilized the ARITI system. Each user completed 12 trials, 6 trials per interaction device (users performed 4 parts of 3 trials). Ten minutes per interaction device was allowed to use the robot before performing the evaluation. The kind a device used has a great impact on the object selection time (the dependence of the selection time variable over the kind of device is high after an ANOVA testing leading to $p < 0.001$, that means device has a large influence on selection time). Flystick control provides less time to complete the task than Data Gloves control (see fig. 13). Another ANOVA testing shows that Object color/position does not have effect on the selection time.

Figure 13. Average selection time.

However, there exists a significative but smaller difference between the two devices for the manipulation time (ANOVA provides less significant results than selection time with $p < 0.01$). As for selection, Flystick appears to give better results, as shown in fig. 14.

Figure 14. Average manipulation time.

As we can see in fig. 15, skilled users have more facility to control the robot with the tracked gloves than the naïve users. The experiment shows that skilled users may perform better than naïve users when using Data Gloves (ANOVA gives significant results on relation between selection time with and the skill in VR the user pretends to have, with $p < 0.001$). In the case of the Flystick control, results are not as much significant as the Gloves control. It seems that the Flystick is more adapted for novice users. One explanation may be that the control with the Data Gloves is so natural that users want to control robot's arm as if it was their own arm, but the mechanical constraints of the arm of the robot limit the user's movements in the virtual world, which may be puzzling. So there is incoherence between what the user wish and what the robot can do.

Figure 15. Influence of the skill of the user on the selection time, depending on the selected device.

7. CONCLUSION AND PROSPECTS

The main objective of our system is to get a natural human robotinteraction in which the human operator could choose the device(s) with which he should be the more efficient, without worrying about the hardware layer.

In this paper, in order to answer to this objective, we proposed a *Multimodal Management System* (MMS). MMS allows 3D Human-Robots Interaction systems to manage multimodality from many devices, especially pointing devices, in order to have an

efficient and more natural Human-Robot Interaction. This component uses the Virtools framework and may combine easily many devices and unify their data, to obtain a virtual device with the capacity to compensate loss of data from one or more devices.

 We have tested our MMS on a preliminary experiment on a 6 dof robot teleoperation task that uses the ARITI framework to assist the user and our VR/AR semi-immersive platform EVR@ for the 3D interaction*.* Two devices were used to interact with the robot: Flystick and wireless Data Gloves.

In order to evaluate our MMS we use the EEA (Empirical Evaluation Assistant), we developed in our laboratory. The preliminary evaluation permits to have a rather good feedback on the MMS, during the use of the tracked data gloves and the Flystick. It appears that the selection time and manipulation time are lessen by using the Flystick comparing to the wireless data gloves. Moreover, naïve users seem to have difficulties for manipulating the 6 dof robot with data gloves, whereas they do not have when using the Flystick. It may be due to the fact that the 6 dof robot arm cannot reproduce the all real arm gestures, which may puzzle the user.

The future use of the SPIDAR in addition to the Tracked Data Gloves might solve the problem of incoherence between the constraints from robot and user's arm, with its force feedback capabilities. Moreover stereo versus mono viewing or the use of the AR should be tested on a future work. MMS has many advantages in terms of transparency for users. Such a Multimodal Management System is therefore more likely to promote human robot interaction (HRI).

8. ACKNOWLEDGMENTS

We wish to thank "Le Conseil Général de l´ Essonne", "Le CNRS" and the IBISC laboratory for funding this work.

9. REFERENCES

- [1] Irawati S., Calderòn D. and Ko H., 2006. Spatial Ontology for Semantic Integration in 3D Multimodal Interaction Framework, in VRCIA, ACM, Hong Kong, pp. 129-135.
- [2] Kaiser E., Olwal A., McGee D., Benko H., Corradini A., Li X., Cohen P. and Feiner S.,2003. Mutual Disambiguation of 3D Multimodal Interaction in Augmented and Virtual Reality, in ICMI-PUI '03, ACM.
- [3] Otmane, S., Mallem M., Kheddar A. and Chavand F., 2000. ARITI: an augmented reality interface for teleopeation on the internet. In: *Advanced Simulation Technologies Conference (ASTC2000)*. Wyndham City Center Hotel, Washington, D.C., USA. pp. 254–261.
- [4] Ouramdane N., Davesne F., Otmane S. and Mallem M., 2006. FOLLOW-ME: a new 3D interaction technique based on virtual guides and granularity of interaction.*"* ACM International Conference on Virtual Reality Continuum and Its Applications (VRCIA 2006), pp. 137--144, Hong Kong, China.
- [5] Sato M., 2002. Development of String-based Force Display: SPIDAR, in 8th International Conference on Virtual Systems and Multimedia, Precision and Intelligence Laboratory,

Tokyo Institute of Technology, 4259 Nagatsuta, Midori-ku, Yokohama, 226-8503, Japan.

- [6] Bowman D., Kruijff E., Laviola J. and Poupyrev I, 2001. 3D User Interfaces theory and practice, Edition Addison Wesley, USA.
- [7] Karaseitanidis I., Amditis A., Patelb H., Sharplesb S., Bekiarisa E., Bullingerc A. and Tromp J., 2005. "Evaluation of virtual reality products and applications from individual,

organizational and societal perspectives: The ''VIEW'' case study", Int. J. Human-Computer Studies 64, pp. 251–266, Elsevier Science.

[8] Gabbard J. L., Hix D. and Swan II J. E., 1999 "User-Centered Design and Evaluation of Virtual Environments", IEEE Computer Graphics and Applications, USA.