IMPROVED COMMUNICATION IN VIRTUAL WORLDS

Eric Saravanane Varadaradjou, Philippe Dax, Alain Grumbach
GET-Télécom Paris – CNRS LTCI-UMR 5141
Ecole nationale supérieure des télécommunications
46 rue Barrault 75013 Paris, France
Eric.Varadaradjou@enst.fr, Philippe.Dax@enst.fr, Alain.Grumbach@enst.fr

ABSTRACT: *The principal contribution of this article is to present the design of an advanced communication between users mediatized by a virtual world. Introducing Virtual Reality into communication acts improves communication by making the meaning more obvious. An exchange can be enriched for a better comprehension. After a short presentation of the problematic and a few existing supports, we expose the design of a conversational agent and a paradigm communication assistance.*

KEYWORDS: *Communication, Virtual Reality, Conversational Agent, Modalities, Artificial Intelligence.*

I. Introduction

Virtual reality (VR) is one of the components of the information highways. It allows a user to visualize and to interact in a Virtual World (VW). The systems of virtual reality are very dependent on the technical level of their various components. Research on cooperative work using the computer – Computer Supported Cooperative Work (CSCW) – tries to determine how data processing can help groups of people to work together on a project, to work out a problem, to make a decision, etc. [Grudin,1994]. Cooperative work calls upon techniques of virtual reality and requires a certain standard of communication.

With this title, we introduce the concept of improved communication. By "improved", we understand a communication which is the result of inferences. We differentiate it from an augmented communication that is defined in our research as label which typically facilitate identification of an object.

Our objective is to improve the communication between users mediatized by a VW [Lemer,2001], to make it easier to understand. We consider the act of communication as relevant if it isn't ambiguous and gives sufficient information for the comprehension of the message. The enrichment is orchestrated by an agent which we describe in section IV.A.

In this article, we present many aspects of human communication and its specificity in a VW. We study in detail the various methods of communication and their use. We finish by explaining the potential of VR in communication.

II. Communication

Communication comes under different forms and its oral form is the most used. The purpose of all communications is to transmit knowledge. Each communicator shares the information they have, and the information and the ideas are pooled.

A. Evolution of the Communication

The communication follows the development of technology which results in an improvement and an acceleration of communications. The evolution of the communication up to now went through following steps (illustrated below):

1. Face to face: gestures ...
2. Telegram, telephone ...
3. Teleconference, VR ...

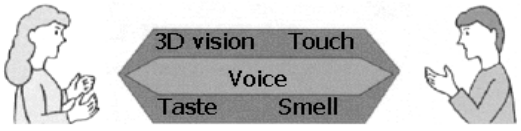

Figure 1 : Face to face communication♦

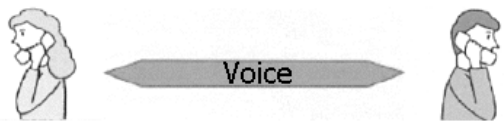

Figure 2 : Communication mediatized by the telephone♦

♦ Images from ATR MI&C Labs

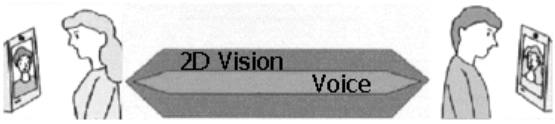

Figure 3 : Communication by the teleconference♦

Currently, VR offers possibilities of collaborative interactions – e.g. teleoperation and codiagnostic over distance. In this context, what are the new possibilities for communication in VR?

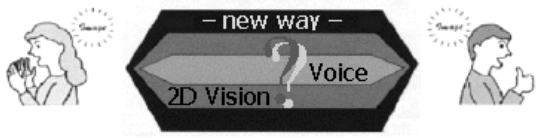

Figure 4 : The augmented communication♦

Let us look more precisely at the various components of communication which enable us to understand logical components of it.

B. Communication Components

The communication can be divided into 6 elements:

- **Intention**: *the origin of the act of communication.*
- **Emitter**.
- **Message**: *the information contained in the communication.*
- **Medium**: *the support used for the communication.*
- **Receiver**.
- **Impact**: *the effect obtained after the reception of the communication.*

An improvement of communication implies an augmentation of one (or more) of these components. Among these, we decided to focus on the message and the medium.

The message (the information, the contents):

"More words than one go to a bargain."
"A word to a wise is sufficient."

The message and the transmitted information are often incomplete or ambiguous. In speaking and in computer "chat", it is common to omit some details especially if they appear obvious to us. Consequently, to remove ambiguities in communication, we need to add extra information to allow a better understanding.

The medium (the support, the container):

"Language is the medium of thought."

The medium of communication can be regarded as an interface which is a group of processes destined to restitute the message. The interface formats the intentions and ideas before transmitting them. Consequently, the improvement of the medium comes down to enlarging the sphere of activities in the transporter and a diversifying of the medium.

For a long time, man has been using voice, gesture, gaze, etc. to communicate. Currently, VR is the most advanced tool with a virtualization of many acts of communication. We will study some VW in order to propose some contributions.

C. Communication in a Virtual World

Over the last years, we have seen the arrival of new more complex VW with limited tools and paradigms of interactions. New tools and paradigms introduce new ways of communication between man and machine. Their intentions are to develop the interactions and to assist the user in his evolution in VW.

Each VW has its advantages and its limitations. Our interest was directed towards certain possibilities of the VW enabling us to exceed reality. VR, through its simulation of reality has some important specificities: the knowledge of all the conditions and all the limitations of the interaction [Fraser,2000]. According to this principle, we understand that the improvement of the interaction means a better use of the information instilled in the VW. There is another important characteristic: all the actions are carried out by interaction. In other words, the operator perceives and acts on the world only by the intermediary of a hardware device. To make our communication more clever, we should use the information provided by this media as well as possible. All graphical conditions, the user's field of view (FOV) and all the information in VW help the user. The power of the VW is in the knowledge of the world.

The work on a conversational agent [Nugues,1997] for an improvement of navigation in the VW showed an undeniable contribution. But this work showed its limitations with the use of only the textual modality, keyboard. The use of only the textual modality does not allow a full exploitation of the possibilities of navigation. With this agent, it is, for example, very difficult to interact correctly with a precise element if it doesn't have enough distinct characteristics – e.g. the selection of a pebble on a pebbled beach.

ULYSSE results from the work of Pierre Nugues [Nugues,1998] for the creation of an agent of navigation. The user communicates and dialogues with an agent which informs him or moves his avatar in the VW with only textual modality. This constraint on only using the textual method in a VW, where the possibilities of interacting are vast, seems quite restrictive. The idea to associate another modality allowing a designation and an

♦ Images from ATR MI&C Labs

improvement of the interaction is not recent. Many are the VW using several modalities, but only a few group them together. We can distinguish several types of VW. In some worlds, they use designation in fusion [Pfeiffer,2005] with methods to move and place objects. In others, they use modalities independently [Croquet,2003]. And some only use designation as a simple modality and it remains a labeling tool, a tool for describing an object, or a pointing tool – e.g. VREng [Dax,1997] or Dive [Andersson,1994].

Our approach for improved communication is to introduce a new modality: the deictic. Deictic is a designation method of a singular object through a visual medium. But the use of this new method is a way towards multimodal fusion. The introduction of this new method includes some problems and possibilities which we must study. In conclusion, the escalation of the communication necessitates an improvement of its modalities.

III. Modalities

Our research was directed towards the study of these media: a message typing device – keyboard, the textual modality – and a pointing graphic device – mouse, the deictic modality. A modality is defined as a process which respects the conditions defined by Jean-Claude Martin [Martin,1994].

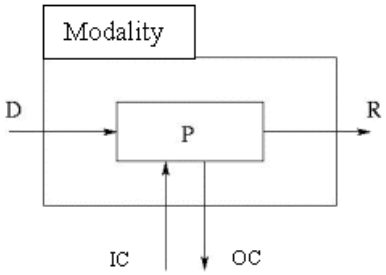

Figure 5 : Modality components

A modality is defined by : **data** {**D**} which is analyzed at a given moment and which respects a set of **incoming conditions** {**IC**}; a **process** {**P**} which analyzes the data; a set of results {**R**} which respects a set of **outgoing conditions** {**OC**}.

To increase the communication in the VW, we increase the modalities independently and their synergy.

A. Deictic Modality

If we observe the components of the deictic modality following this definition, we obtain:

- **Data**: *a set {X, Y, Z} representing the indicated point.*
- **Incoming conditions**: *object defined by the field of selection – visible object.*
- **Process**: *computation of the selected object.*
- **Outgoing conditions**: *object present in the {List of objects}.*
- **Result**: *an object "O" which corresponds to the selected object.*

According to this breakdown of the modality, we can observe a way to improve the input conditions. Indeed, the objects must be located in the FOV to be able to be selected. Nevertheless, we should be able to quickly select any element in the VW and not only the immediately visible objects. It will enable us to interact and to handle distant objects.

B. Textual Modality

If we observe the components of the textual modality, we obtain:

- **Data**: *a sentence in natural language.*
- **Incoming conditions**: *lexical and syntax recognition; the sentence must be recognized.*
- **Process**: *computation of the action to be performed*.
- **Outgoing conditions**: *semantic recognition; the sentence must have a meaning.*
- **Result**: *performing an action in the world.*

To increase the textual modality, there are a few possibilities. In the following example, we describe a 3D-scene with 3 objects: 2 cars and a radio represented by a white box. The first car is light gray and locked, the second dark gray and unlocked, and the radio is in the dark gray car. This scene could be described by the Figure 6/Figure 7:

Figure 6 : The XML description

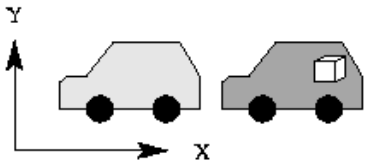

Figure 7 : The described scene

If we wish to communicate about these objects, we can find differences between the objects to differentiate them. We can, by using the VW, increase the textual communication in 3 possible ways.

Case 1: Differentiation Between Elements.
The first possibility to enhance the textual communication is a simple computation of the logical distances between the objects. We can distinguish the various elements present in the scene and in the communication, to indicate the obvious components of the object. When an act of communication indicates an object, we can, by using the description of the world, provide data

about it. In our example, the elements can be differentiated by giving some additional relative information about their color or size.

Case 2: Deductions of the Scene.
Elements have different characteristics , but there are also differences according to the context. Moreover, we can use the logical relations in 3D-topology to compute the space topology of the scene which allows us to describe the scene. We can provide important information between the various elements of the scene. In our example, space computation enables us to know the position of the first car in relation to the second in our FOV. We can also know the contents of the elements which maybe not-visible (the radio being inside the second car).

Case 3: Selection by Functions.
This last possibility is special; it uses the engine of the VW and comes from observation. When we communicate about an object, we can also indicate about its functionality. When an object is created, it has some properties describing its interaction with the user and the other objects in the VW. In other words, it is possible to differentiate and to increase the communication by providing the possible interactions with the object. Objects can have very similar characteristics but can have different interactive functions. In our example, the two cars can differ through their interaction functions, and the radio (represented by a box in Figure 7) is the only one object able to be turned on/off.

Each modality can be increased independently, but their fusion offers an improvement.

C. Multimodal Use

Multimodal fusion contributes to the improvement of virtual communication. The modality provides information and the pooling of these data seems necessary. We must study when and how multimodal fusion takes place.

There are several types of multimodal fusion. To understand our type of modality, we describe all the possibilities. Let us begin our specification with examples. At the time of a bimodal speech (textual and deictic), we observe all the possible cases of informational uses (we use “” and ** for textual and deictic modality, respectively):

a) Textual modality only: “`Go to the radio !”`
b) Deictic modality only: `*pointing to an object*`
c) ~~**Redundant**~~**:** `“Go to the radio!”, *pointing to the radio*`
d) **Complementary:** `“Go to this car !”, *pointing to the car C1*`
e) **Contradictory:** `“Go to the car !”, *pointing to the radio*`

Leaving aside the first 2 examples, we concentrate on the more interesting ones. To increase the synergy between these two modalities, we must maximize the use of the complementary case and avoid the others. We wish to obtain a "combined" use of these modalities and not an independent use of example a or b, which results in a limited act.

In this breakdown, we observe that the sphere of activity of the textual modality provides more information. The deictic modality cannot be used alone in our case. Instead, it is more advantageous to use the deictic as a support to the textual in order to "supplement" ambiguous textual information.

At which moment do we use these modalities and in which order? Let us take temporal examples to illustrate the various cases (A < B denotes that A takes place before B and A // B denotes that A occurs at the same time as B).

a) Deictic :
`*Pointing* < “Go to this car!”`
b) Textual :
`“Go to this car!” < *Pointing*`
c) ~~**Deictic-Textual**~~ **:**
`“Go to this car!” // *Pointing*`

We are not concerned in “c)” case because the interfaces of data acquisition do not allow us to indicate with the mouse when we use the keyboard. Moreover, this method is not natural and too heavy for a VW user on a standard screen.

There remains the other two cases. The “a)” case can be interpreted like an act of information. In this case, we assume that the user cannot make errors. He selects an object and confirms his selection. With this assumption, it is noticed that the operator can bring some more precise additional information to the VW itself and the object pointed at. If the user indicates with the deictic modality an object which type differs from the object indicated with the textual modality – e.g. the indicated object and the requested object are different. We are thus able to inform the user on the nature of the object and to remove any doubt about it.

The “b)” case is also interesting. In this case, the operator wishes to go towards a precise object. We know the nature of intention, but we need a designation because his request is ambiguous. Then, we can increase, between these two actions, the selection possibilities and remove the ambiguity. It is possible to help the designation by highlighting. In our example, after the textual request, all the potential objects (visible or not) are quickly accessible.

Through these steps, a solution appears. We must use the VW and its knowledge to augment the acts of communication between users. All acts can be increased to help the users if they request it. In order to clarify the ideas of each one, we present the use of the textual accrual:

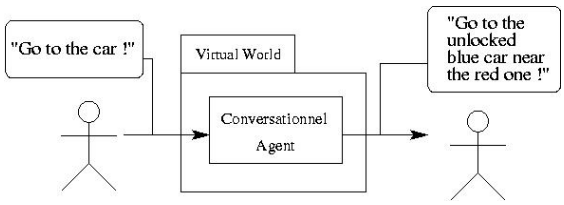

Figure 8 : Improvement of the message

In this example – described by the Figure 8 –, only the transmitted message is augmented. The contribution of designation makes it possible to increase and diversify the container types:

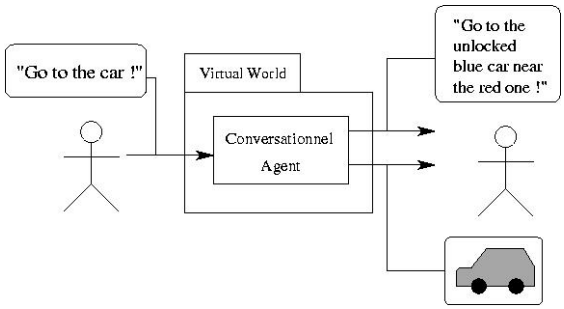

Figure 9 : Improvement of the medium

We offer a precise vision of the object and of his context. The user has all possible information available on the object. We have a textual description of the object and a good FOV. The support which was only textual is amplified by different FOV of the various objects; the deictic modality is used to provide extra information. Now, we explain an implementation of these protocols using a VW: "**V**irtual **R**eality **Eng**ine"[Dax,1997].

IV. Implementation

The VW can be created and modeled with various tools. A good tool for visualization and handling of the world is necessary to confirm the implementation of our ideas. After a preliminary study of the various methods of interactions, we observe a lot of differences on their language of description and their internal representations. We searched for a VW adapted to our centers of interests.

Among these VW, some are centered on the speed of creation [Pesce,1995], while others on the temporal processing [Richard,2001] where the whole scene is described in the shape of "agents". Others are centered on the communication between objects [Superscape,1997]. Finally, other VW mix all of these aspects : VREng, Dive [Andersson,1994].

Our choice was centered on the VREng developed in the ENST-Paris because it gives the easy way to create a VW using a XML structure. The 3D-engine uses the C++ language and the OpenGL library. It enables us to use the standard Gnome-LibXML library to study and analyze the structure of the world. Indeed, XML makes it possible to quickly seek elements of the VW with a hierarchical structure.

Our program operates between the acts of communication; it receives and increases them. This program is like an omniscient agent, which observes and analyzes the world to give all the desired information. This conversational agent is at the center of our research.

A. Conversational Agent

The agent must understand the communications of the users and try to augment them.

An agent dialogist is essential to show our approach. We created a linguistic analyzer in OCaml language able to understand basic orders. The possible actions are limited but they are sufficient for our needs. Here is a set of orders which are processed:

- `"Look behind the door near the plant!"`
- `"Take the bottle on the table!"`
- `"Go to the statue!"`
- `"Open!"`

Formally, we take into account only the conversations with this syntax :
Order(Verb(v),Position(p),Object(o), Complement(c))
Order(Verb(v),Position(p),Object(o))
Order(Verb(v),Object(o), Complement(c))
Order(Verb(v),Object(o))
Order(Verb(v))

Only the requests recognized by the lexical and syntactic analyzer are augmented. In the other cases, the messages are transmitted to the recipient without any augmentation.

The general diagram of our agent is illustrated in Figure 10. All the communications are transmitted to the agent by the interface. The agent processes and solves the problems of references. Thereafter, the message is transmitted to the recipient by the 3D engine.

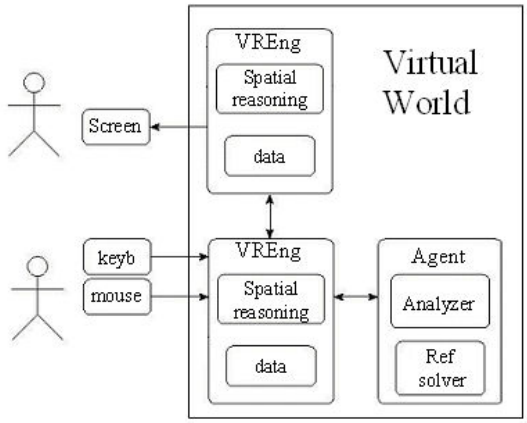

Figure 10 : Diagram of the Agent

B. Augmented Transmission

The elements *Object(o)* and *Complement(c)* are reused by the agent to find the object in the scene. The agent transforms the object and the complement into the *XPath (XML tool)* requested to obtain information. If the agent does not find the elements in the scene, it informs the operator of his failure. The latter can either supplement its request or just transmit it. The completion of the request can be carried out in a textual way – adding information – or in a deictic way – designating.

1. Designation Modality

Designation does not allow a great diversity of actions in our context. Nevertheless, this modality complements the textual modality. It supplements and discriminates the textual elements. It is important to be able to point at every element which composes the world. In our previous study of this modality, we observe a limitation of selectable objects. We can select only the visible objects. All objects of the scene can contribute to increase the information but only some of them can be pointed at – limited by the FOV.

Indeed, it is irrelevant to be unable to select a not-visible object directly. In the case of a multimodal use, the user gives a textual request before and some information in the request can be reused for designation. After a request like: "`Go to the car!`", if there are several cars in the world, we should be able to indicate every car of the world to help the designation. The textual modality must help and inform the deictic modality by an highlighting of the possible objects.

From this highlighting, we develop two designation paradigms: indirect designation – designation of occulted objects – and indirect designation with intelligent camera – designation in a set of occulted objects.

1.1 Indirect Designation

This paradigm designates the hidden objects. The method is easy. We use the buffer of selection implemented in OpenGL to let the operator see through the objects in his FOV.

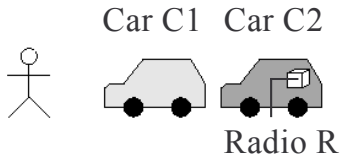

Figure 11 : Diagram of the scene

In Figure 14, the selection laser goes through – by using the *"see through"* button – the 1st car C1 and highlights the 2nd hidden car C2. By reusing the *"see through"* button, we can select the hidden object in the dark gray car.

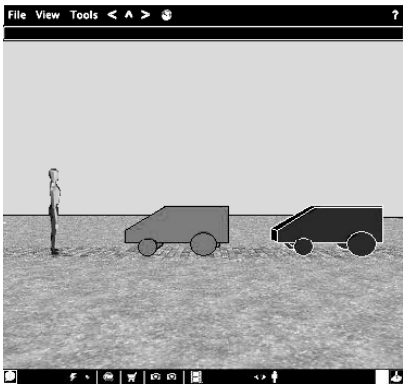

Figure 12 : The scene in VREng

Figure 13 : Indicating the car C1 in front of the user

Figure 14 : He sees through the car C1. It is made transparent and we select the second car C2.

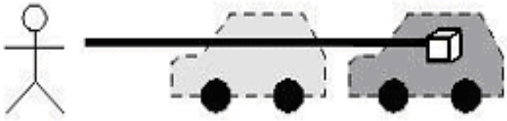

Figure 15 : He proceeds in the same way and he can select the radio R inside the car C2.

1.2 Intelligent Designation with Camera

But indirect designation is sometimes ambiguous when several objects appear next to each other. The

natural human interaction is to move nearer the object and to observe in detail the desired object. However, this paradigm of movement can be replaced easily by an intelligent camera which would move and would automatically take an appropriate position where the desired object is visible. This protocol of designation assisted by the camera allows a simplification of designation and can avoid visible ambiguities. Let us explain by an example. The operator wants to take the bottle B1 placed on the table. He sends: *"Take the bottle on the table!"*. Two bottles exist in this scene and the small bottle B2 is partially hidden by a big one B1. The agent informs the operator about the ambiguity of the scene and proposes to send a camera.

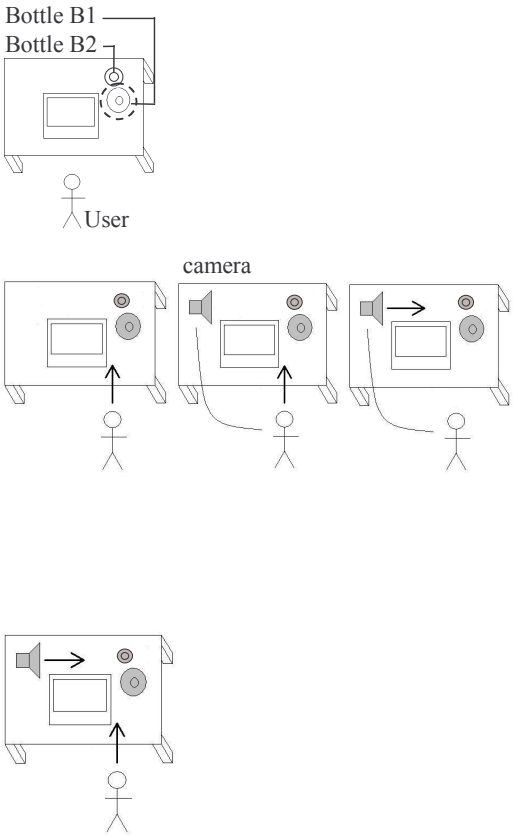

Figure 16 : Use-case of the intelligent camera

The sending of this camera gives us the possibility of removing the ambiguity. Thus, the operator has a clear sight of the objects and can easily select the desired object.

In Figure 16, the operator requests to take a bottle but there are 2 bottles on the table. The agent informs him of this ambiguity and he proposes a point of view to help him. The operator can thus indicate precisely the object.

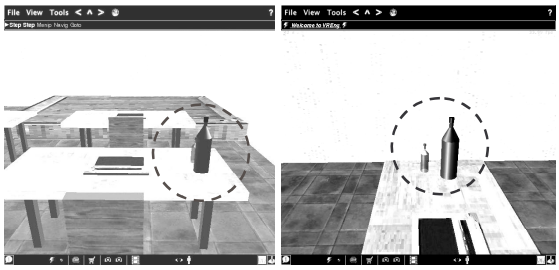

Figure 17 : The same use-case in VREng

After the selection, the user's point of view is reused. This FOV supposes an unambiguous vision of the object. Now, it is possible to understand the message and to observe the object. The transmission of the point of view is an improvement in communication. With the assistance of the graphic library *"Ubit"* [Lecolinet,1999], the improvement of the request with an image of the scene is simplified :

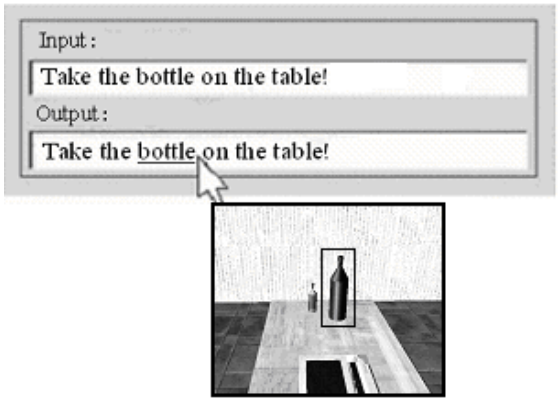

Figure 18 : An Augmented Communication

In Figure 18, the operator wants to transmit "`Take the bottle on the table`". The intelligent agent increases the message and transmits it with an image of the designated bottle. Thus, all possible ambiguities are removed.

2. Textual Modality

We use textual modality to transmit the main ideas. But some ideas remain ambiguous and require additional information to remove the ambiguities. We saw that it was possible to use another modality for removing ambiguities. Now, we see the possibilities of increasing the textual modality. It brings additional information concerning the entity. We use the properties of the VW: the complete description of the world and its internal interactions. To explain the various possibilities, we use the example of Figure 6 and Figure 7, where we observe two cars side by side.

2.1 Deductions of the Scene

A more interesting enhancement is the deduction of the scene. The scene is seen in a subjective way by different operators. The description of the scene is related to the position of the user and his FOV. From the position and dimension of the various

elements of the world, we "describe" the existing relations between the entities according to a simplified space topology. But we cannot process all the objects of the world. We locally describe the scene with the required entity and the FOV of the user. We define an “Area of Interest”(AOI) where we apply our methods. The area is using a vicinity list to compute the information about the object.

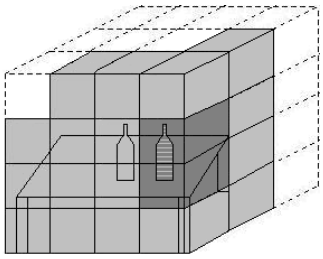

Figure 19 : An example of the “Area of Interest”

We use the visual information of the world and the OpenGL access functions to determine our needs. We use the division of the world in regular subsections given by the collision-engine. Then, we restrict the analysis to the elements in the neighborhood of the entity understudied.

For example, for Figure 6, if we wish to obtain more information concerning the car on the left, we can generate a simple explanatory text. We analyze the objects with our point of view and we obtain a list of short descriptions such as:

We obtain a textual description of the relations between the surrounding objects. We can help the operator to understand the context of the entity and to allow him to locate it.

2.2 Selection by Functions

Here, the augmentation comes from a process of reasoning. In the acts of communication, we observe the people who seek an element by its functionality. For example, a person arrives in front of a door and he says : “Open!”. The object concerned by the action is the door. The action of opening is guessed.
Unfortunately, the functionalities are not in the VW scene description but specified in the programming code in VREng and Dive. This is contrary to what is indicated in Figure 6.

The idea is to seek, in the 3D-engine, the functionalities of the various elements to express them in a literal way and to give an outline of the entities. In our example, we have 2 cars but only the car C2 on the right side has the “handle” function. The knowledge of this fact provides additional information on each car.

V. Conclusion

In the virtual collaborative work, communication is essential. A good communication permits to appreciate a situation without ambiguities and allows an effective transmission of information. In our study, a virtual world supports all acts of communication between the users. Its various tools (3D-engine) offer many possibilities of assisting the communication. Ambiguities are dissipated and thus, communications provide relevant information. We concentrated this role in **an omniscient conversational agent** which aims to supervise and to support the user-to-user acts of communication. This agent uses the various possibilities offered **by different modalities and multimodal fusion** to assist the user in his activities. As a result, we found many ways to augment the communication by studying in detail the different components of communication.

VI. Acknowledgments

The authors are very grateful to Peter Weyer-Brown (ENST), S. Nandini Nadar for their careful reading of the manuscript for this article.

References

[Andersson,1994] Magnus Andersson, Christer Carlsson, Olof Hagsand, and Olov Stahl, *"Dive, the distributed interactive virtual environment"*, Technical reference, Swedish Institute of Computer Science, Kista, 1994.

[Dax,1997] Philippe Dax, Denis Arnaud, Fabrice Bellard, Stéphane Belmon, Samuel Orzan, Lionel Ulmer, *"Spécifications globales de VREng : Virtual Reality Engine"*, France, February 1997.

[Fraser,2000] Mike Fraser, Tony Glover, Ivan Vaghi, Steve Benford, Chris Greenhalgh, Jon Hindmarsh, Christian Heath, *"Revealing the Realities of Collaborative Virtual Reality"*, CVE 2000, San Francisco, CA USA.

[Grudin,1994] Jonathan Grudin, *"Computer-Supported Cooperative Work: History and Focus",* Journal Computer, volume 27, number 5, year 1994, ISSN 0018-9162, pages 19 – 26, IEEE Computer Society Press, Los Alamitos, CA, USA.

[Lecolinet,1999] Eric Lecolinet, *"A Brick Construction Game Model for Creating Graphical User Interfaces: The Ubit Toolkit"*, Ecole Nationale Supérieure des Télécommunications, CNRS URA 820, Paris, France, 1999.

[Lemer,2001] Pascal Lemer, PhD Thesis, *"Modèle de communication Homme-clone-Homme pour les Environnements Virtuels Collaboratifs non-immerssifs"*, Université des Sciences et technologies de Lille, UFR d'IEERA, Lille, 2001.

[Martin,1994] Jean-Claude Martin, *"Coopérations entre modalités et liage par synchronie dans les interfaces multimodales",* Ecole Nationale Supérieure des Télécommunications, France, 1994.

[Nugues,1997] Matthias Ludwig, Alexandre Lenoir, Pierre Nugues, *"A Conversational Agent to navigate into MRI Brain Images"*, ISMRA, Caen, France, 1997.

[Nugues,1998] Olivier Bersot, Pierre-Olivier El Guedj, Christophe Godereaux, and Pierre Nugues, *"A conversationnal agent to help navigation and collaboration in virtual worlds"*, Virtual Reality, 3:71– 82, 1998.

[Pesce,1995] Mark Pesce. *VRML {Browsing and Building Cyberspace*. New Riders Publishing, 1995.

[Pfeiffer,2005] Thies Pfeiffer, Marc Erich Latoschik, *"Resolving Object References in Multimodal Dialogues for Immersive Virtual Environments",* University of Bielefield, 33594 Bielefeld, Germany, IEEE Virtual Reality 2004 March 27-31, computer society, Chicago, IL USA.

[Richard,2001] Nadine Richard, PhD Thesis, *"Description sur les comportements autonomes évoluant dans les mondes virtuels"*, Ecole Nationale Supérieure des Télécommunications, Paris, France, Octobre 2001.

[Smith,2003] Smith David A., Andreas Raab, David P. Reed, David P.Reed, "*Croquet: A Collaboration System Architecture"*, technical report, 2003.

[SuperScape,1997] SuperScape, *"3D Webmaster : User Guide"*, chapter 16, pages 207–227, Santa Clara, CA 95054, USA, november, 1997.