Nonverbal Communication Interface for Collaborative Virtual Environments

Anthony Guye-Vuillème[1], Tolga K. Capin[1], Igor Sunday Pandzic[2], Nadia Magnenat Thalmann[2], Daniel Thalmann[1]

[1]Computer Graphics Laboratory
Swiss Federal Institute of Technology
CH1015 Lausanne, Switzerland
{capin,thalmann}@lig.di.epfl.ch
http://ligwww.epfl.ch

[2]MIRALAB-CUI
University of Geneva
24 rue de Général-Dufour
CH1211 Geneva 4, Switzerland
{Igor.Pandzic,Nadia.Thalmann}@cui.unige.ch

Abstract

Using virtual humans as participant embodiment increases the collaboration in Networked Virtual Environments, as it provides a direct relationship between how we interact with the real world and the virtual world representation. Nonverbal communication is an important aspect in real-life collaboration, therefore the means allowing the users to replicate it should be provided in the virtual world. In this paper, we present our category of nonverbal communication actions based on the psycho-sociological theory, and provide a solution that balances between practical limitations of input devices and sociological aspects. We introduce our implementation using our VLNET (Virtual Life Network) networked virtual environment.

Keywords: Embodiments, Networked Virtual Environments, Nonverbal Communication.

1. Introduction

Although a lot of research has been going on in the field of Networked Virtual Environments, most of the existing systems still use simple embodiments for the representation of participants in the environments. More complex virtual human embodiment increases the natural interaction within the environment. The users' more natural perception of each other (and of autonomous actors) increases their sense of being together, and thus the overall sense of shared presence in the environment.

What gives its real substance to face-to-face interaction in real life, beyond the speech, is the bodily activity of the interlocutors, the way they express their feelings or thoughts through the use of their body, facial expressions, tone of voice, etc. Some psychological researches have concluded that more than 65 percent of the information exchanged during a face-to-face interaction is expressed through nonverbal means [Argyl88]. A VR system that has the ambition to approach the fullness of real-world social interactions and to give to its participants the possibility to achieve a quality and realistic interpersonal communication has to address this point; and only realistic embodiment makes nonverbal communication possible [Capin97][Benford97].

In Networked Virtual Environments, virtual humans have various tasks to perform, in addition to nonverbal communication, such as navigation and interaction with virtual objects. These tasks should be merged realistically with the nonverbal communication within a *modular* approach. We exploit a flexible framework for the integration of virtual humans in the Networked Virtual Environments, called VLNET (Virtual Life Network).

This paper presents our solution to nonverbal communication in Networked Virtual Environments, with simple interfaces under constrained input conditions. We first survey embodiments in NVEs, present the theory developed for social sciences for nonverbal communication, and provide a solution that balances the sociological communication aspects and practical input limitations.

2. Embodiment in Networked Virtual Environments

The pace in computing, graphics and networking technologies together with the demand from real-life applications made it a requirement to develop more realistic virtual environments (VEs). Realism not only includes believable appearance and simulation of the virtual world, but also implies the natural representation of participants. This representation fulfills several functions: 1) the visual embodiment of the user, 2) means of interaction with the world, 3) means of sensing various attributes of the world.

The realism in participant representation involves two elements: believable appearance and realistic movements. This becomes even more important in multiuser Networked Virtual Environments (NVE), as participants' representation is used for communication. This avatar representation in NVEs has crucial functions in addition to those of single-user virtual environments: 1) perception (to see if anyone is around), 2) localization (to see where the other person is), 3) identification (to recognize the person), 4) visualization of others' interest focus (to see where the person's attention is directed), 5) visualization of other's actions (to see what the other person is doing and what is meant hrough gestures), 6) social representation of self through decoration of the avatar (to know what the other participants' task or status is). Using virtual human figures for avatar representation fulfills these functionalities with realism, as it provides the direct relationship between how we control our avatar in the virtual world and how our avatar moves related to this control.

Inserting virtual humans in the NVE is a complex task. The main issues are: 1) selecting a scalable architecture to combine these two complex systems, 2) modeling the virtual human with believable appearance for interactive manipulation, 3) animating it with minimal number of sensors to have maximal behavioral realism, 4) investigating different methods to decrease the networking requirements for exchanging complex virtual human information.

Particularly, controlling the virtual human with limited input information is one of the main problems. For example, a person using a mouse will need extra input techniques or tools to exploit the functionalities of his embodiment. In this paper, we survey these tools that help a user with desktop VR configuration, we did not consider full tracking of the body using magnetic trackers, although this approach can be combined with limited tracking of the participant's arms.

3. The Field of Nonverbal Communication in Social Sciences

The use of the body in interpersonal communication has been studied in psychology under the name of « nonverbal communication » (NVC). The definition of this field is based on an exclusion : one defines nonverbal communication as the whole set of means by which human beings communicate except for the human linguistic system and its derivatives (writings, sign language, etc.). Communication is to be seen in this definition, in short, as the sending by a sender of stimuli which causes a change in the state or behaviour of the recipient. So, for Argyl [Argyl88] : « Nonverbal communication, or bodily communication, takes place whenever one person influences another by means of facial expression, tone of voice, or any of the other channels [... except linguistic]. This may be intentional or it may not be. ».

In most situations, NVC is not used alone but jointly to verbal communication. It is not to be seen as a concurrent channel but rather as a mean for people to strengthen what they say, explicit it, etc. But according to Ekman and Friesen [Ekman67], there is a type of signal which is still independent from language : the affective expression. It seems that nonverbal communication doesn't need any verbal expression in the task of communicating emotional messages, and that it is able to express things that would be very difficult to express using the linguistic system, even in a very powerful way.

[Corraze80] proposes to distinguish between three types of information which are convoyed by NVC : information about the affective state of the sender, as we have just seen, information about its identity and information about the external world. To communicate this information, three main channels are used : 1) the body and its moves, 2) the artefacts linked to the body or to the environment, 3) the distribution of the individuals in space. Each of these channels has its own field within the psychological study of NVC, the most important ones being the study of Proxemics and the study of Kinesics. The study of Proxemics analyses the way people handle the space around their body and situate themselves next to other people in space. Proxemic research focuses on the analyse of the distance and angle chosen by the individuals before and during their interactions, the relationships associated to each distance, the permission to touch and its circumstances, etc. Other attitudes like the way people choose a place in a restaurant or the distribution of the interlocutors around a table for example, are part of nonverbal communication and inform the others, gestures, postural shifts, and movements of the hands, head, trunk, etc.; their study analyses what is sometimes called the « body language». Three main types of bodily movements have been identified by several authors : the « emblems », the « illustrators » and the postural shifts [Argyle88][Corraze80]. The « emblems » are gestures having a precise meaning and that can be translated by one or two words. Their knowledge is often specific to a group or subculture and their use is mostly conscious. The « illustrators » are movements which are directly tied to speech, serving to illustrate what is being said verbally.

Several authors [Argyle88][Corraze80] have stated that, together with facial expressions, postures are the best way to communicate emotions and states-of-mind. A *posture* is a precise position of the body or some of its parts, compared to a determined system of references.

4. Introducing Social Science Theories in NVEs

The navigation functionalities within NVE systems allow the user to handle in a convenient enough way its « proxemic » type signals, but no functionalities exist for the kinesic aspect. For the “kinesics”, we decided to develop a 2D interface allowing the user to select predefined actions. The use of predefined actions clearly raises some problems, for example because it postulates a conscious choice by the user of its posture, but it seemed to us the best compromise between practical constraints and the will to take in account what the social sciences have taught to us about nonverbal communication

For the beginning of the project, we only wanted a small number of gestures and postures (less than 30), so we decided to try to identify a basic « palette » of actions, which is a difficult task because NVC does not work as a linguistic system and because no psychological study has been able to build such a classification for the interactions in the « real world ». At best, such attempts have been made in more limited fields, like the facial expressions, and we have tried to use them. The following criteria have been used to select the actions :

- documented in scientific papers
- basic action, commonly used, expresses simple idea
- different enough to compose a « palette » of actions
- can be understood in many places/cultures
- can be performed in the standing position
- a graphical representation of the action was available

Table 1 Chosen actions, classified by posture/gesture and part of the body

Postures / Expressions		Gestures / Mimics		
Face	**Body**	**Head/Face**	**Body**	**Hand / Arm**
Neutral	Neutral	Yes	Incomprehension	Salute
Happy	Attentive	No	Rejection	Mockery
Caring	Determined	Nod	Welcoming	Alert
Unhappy	Relaxed	Wink	Anger	Insult
Sad	Unsecure	Smile	Joy	Good
Angry	Puzzled		Bow	Bad

5. Implementation

This section introduces the implementation decisions to integrate nonverbal communication with the networked virtual environment VLNET.

5.1 VLNET

Typically, the virtual environment simulation systems are complex software systems, therefore a modular design imposes itself. It is appropriate to design the runtime system as a collection of cooperating processes, each of which is responsible for its particular task. This also allows easier portability and better performance of the overall system through decoupling of different tasks and their execution over different processors on a multiprocessor machine. In VLNET, we use this multiprocess approach.

Within the VLNET core, the main process executes the main simulation and provides services for the basic elements of VEs to the external programs, called drivers [Capin97]. The VLNET core consists of logical units, called engines. The role of the engine is to separate one main function in the VE to an independent module, and provide an orderly and controlled allocation of VE elements. Moreover, the engine manages this resource among various programs which are competing for the same object.

Drivers provide the simple and flexible means to access and control all the complex functionalities of VLNET. Each engine provides a shared memory interface to which a driver can connect. The drivers are spawned by the VLNET Main Process on the beginning of the session, based on the command line where all combinations of drivers can be specified. From the VLNET system point of view, The NVC application is a Facial Expression Driver, using the MPA (Minimal Perceptible Actions) format which provides a complete set of basic facial actions allowing the definition of any facial expression, and also a Body Posture Driver which controls the motion of the user's body.

Virtual Human Animation Parameters

For the control of the virtual human body posture animation, an articulated structure corresponding to the human skeleton is needed. Structures representing the body shape have to be attached to the skeleton, and clothes may be wrapped around the body shape. We use the HUMANOID articulated human body model with 75 degrees of freedom without the hands, with additional 30 degrees of freedom for each hand [Boulic95]. The skeleton is represented by a 3D articulated hierarchy of joints, each with realistic maximum and minimum limits. Attached to the skeleton, is a second layer that consists of blobs (metaballs) to represent muscle and skin. During runtime the skin contour is attached to the skeleton, and at each step is interpolated around the link depending on the joint angles. From this interpolated skin contour the deformation component creates the new body triangle mesh. Thus, the body information in one frame can be represented as the rotational joint angle values.

5.2 The nonverbal communication application

In order to fulfil the need for an easy and fast usage, intuitive and easy to learn user interface, it has been decided to use image buttons, displaying a snapshot of the actual move and a textual label describing the idea or state of mind expressed by the action.

We decided to work with three windows : the posture, gesture panel and control panels. The posture panel offers a global view of all postures available, with clickable image buttons. It is constituted of two sections containing the postures classified by part of the body, one section for the face, and one for the body. The actions are ordered by « emotional impact », and the sections are divided into three columns : positive, negative, neutral. The gesture panel is organised in the same way as the posture panel but with one section for the head/face, one for the hand/arm and one for the body. This configuration has the advantage of offering a global view of the postures/gestures available. The high degree of organisation of the actions (posture/gesture, part of the body, positive/neutral/negative) combined to the fact that all actions are immediately activable, allow the user to find and execute the action that fits best the situation as fast as possible.

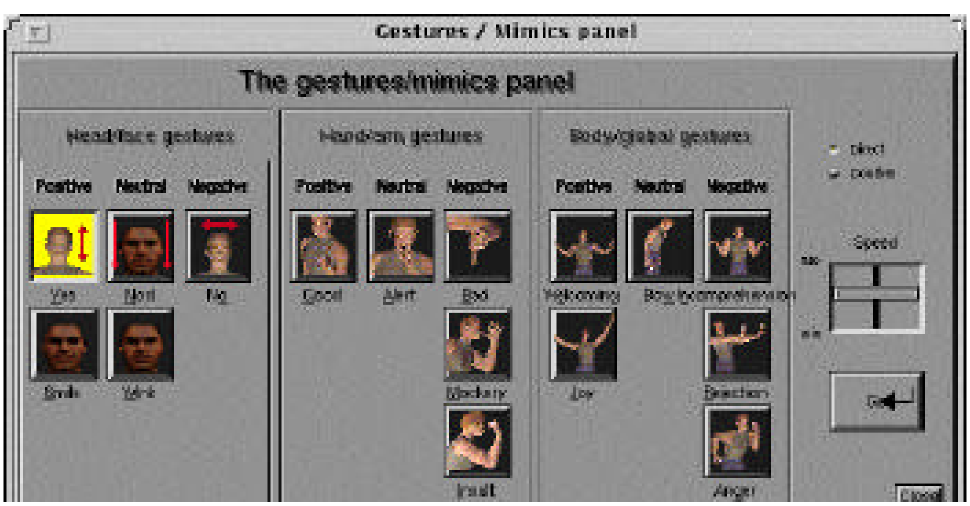

Figure 1 The NVC application interface: the gesture panel

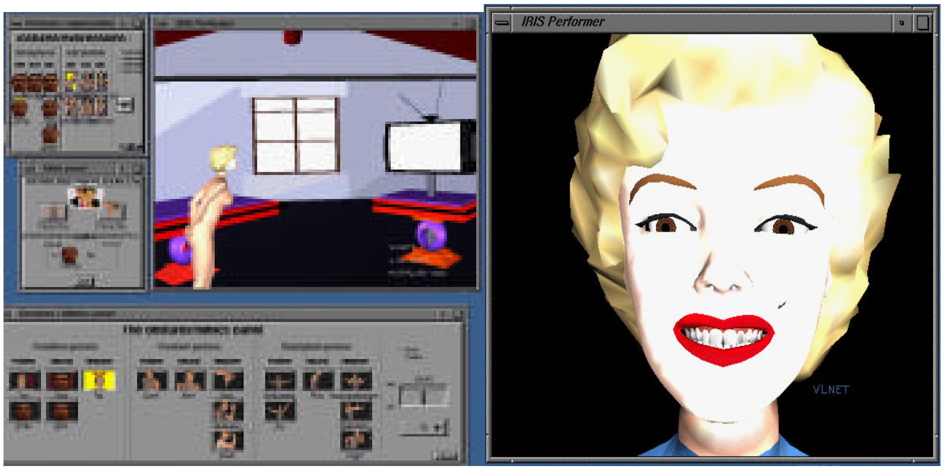

Figure 2 Example of use of the NVC application

5.3 Integration of Nonverbal Communication and Other Actions

As discussed above, the nonverbal communication program is visualized as driver from VLNET point of view. Normally, more than one external driver should be able to connect to the human posture interface, and the task of the engine is to resolve conflicts among the external drivers. For example, while the walking motor updates the lower body, the NVE application will control the upper part of the body. The human posture engine should convert these motions' effects to a single virtual body posture.

Our solution for combining motions is transparent to the drivers. The drivers put their values to the shared memory segment associated with them, and the body animation engine combines these values. In addition to joint values, each driver also is associated with a priority, between 1 and 10, and weight on each joint angle. The default values are 5 for priorities and 1 for weights. The drivers are given right to change their priorities and weight for each joint angle.

For example, the walking and arm motion are combined in the following way. For walking motion driver, all the joints belonging to the upper part of the body, have the default priority:5. The lower body joints have higher priority:6. Therefore, the walking motor's effects are visible on the lower part of the body. Similarly, the arm motion driver has the priority 6 for the right arm joints. The gesture driver has default priorities for all the body. Therefore, the effects of the gesture driver are added equally with the arm motion and walking motion for the parts of the body other than walking motor and right arm.

This motion combination achieved the initial goal: to provide a solution that is transparent to each driver, i.e. the communication is necessary among drivers except the indirect priorities and weights. In this way, while developing the nonverbal communication application, we did not consider walking and grasping motion implementation.

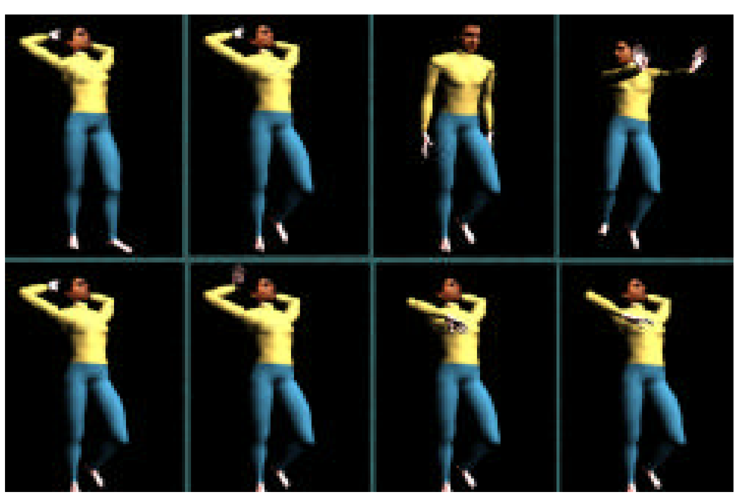

Figure 3 Example results of motion combination

6. Evaluation of the nonverbal communication in VLNET

The idea to evaluate the immersive aspect of the VLNET system and the contribution of the nonverbal communication application in realistic situations with people that didn't know it before, was present from the beginning. After some tests, it appeared to be a rather difficult but promising task.

To avoid a too important impact of the researcher on the results and because there is an increasing preference among psychologists for observing real and spontaneous behaviour, we have decided to give a total freedom of action to the subjects, carefully observe what happens and collect their impressions with a survey at the end of the experiment. In this kind of experiment, the researchers try to identify crucial issues or behaviours, but it would not be accurate to give any number as a result because of the small sample used. The hypothesis built on this small scale experiment can then be verified on a greater sample and quantitative analysis can be done.

We have chosen six subjects : two females and four males, two very familiar with each other, two having already seen each other some times, and two complete unknowns. After an introduction to the system, they have been given a total freedom of action, being allowed to talk with each other or stay silent, to explore the scene or stay at the same place, to use nonverbal communication or not. The experiment lasted for two hours. Three systems were at their disposal for interacting: - a navigation system allowing their avatar to walk freely in the environment, rotate, etc. - the nonverbal interaction application with its thirty actions - a microphone and headphones for verbal communication. The scene we used represented a square with a bar in its centre and was chosen for its public and socially oriented characteristics. Here are some of the interesting points that have been observed :

First, a very important point for the immersive quality of the system that has been noticed, is that the users agreed that the avatar they saw on their screen, maybe was not really their interlocutor, but at least could « work » as the real person and was a credible representation of the other. This is very clear in the words chosen by the subjects : they never spoke about « your representation », « you character », etc. but always used the « You » particle as in « I can see you », « Why don't you move », « You look funny » etc. The same thing is true with their own avatar : « I'm coming in front of you », etc.

Then, it is interesting to notice that the users have been able to reproduce, through the mechanisms of nonverbal communication, their relationship of the real world in the virtual environment. The subjects that didn't know each other before the experiment situated themselves at a bigger interactional distance than the ones who were familiar, and this is typical of what the study of proxemics has showed. Moreover, they carefully avoided all the aggressive gestures when the other ones (who knew each other) used several times the « mockery » gesture or the forearm jerk.

Figure 4 Two subjects interacting at the bar

At another level, the nonverbal communication application allowed them to also respect the formal structure of the social interactions. At the beginning of the interaction, they all used one of the actions to greet the other (« Bow », « Welcoming ») and signal that they were ready to begin the exchange. The end of the interaction followed that logic too, and has always been confirmed by nonverbal means. The normative sanction present when someone doesn't respect these rules in real life has showed up : R was speaking with J. R suddenly decided to explore the world and left J abruptly. J became angry and used verbal and nonverbal means to express it. R came back and they left to explore the world together. This funny example shows very well that, even in virtual environments, when minimal means of control of the interaction are provided, the individuals must respect some rules or the other ones don't appreciate it.

Many other elements confirm this point. During the experiment, the avatars of J and L collided with each other. They naturally apologised and then laughed of the experience. Later, the avatar of J (who is male) and L (who is female) were very close, nearly touching each other that could have been interpreted as very intimate. A strong emotion has been then noticed on the participants, first under the form of uneasiness and then laugh. This behaviour is typical of the relationship of J and L : they have different sexual identities and know each other a bit. The movements and positions of their avatar weren't "free" because they had real consequences and this had nearly the same effect as if it had happened in real life.

Globally, the participants used more the gestures than the postures. A posture was always chosen at the beginning of the interaction, but often it stayed a long time as the participants didn't think of changing it. This can be easily explained by the fact that gestures are mostly conscious actions, while postures are often chosen unconsciously. The users had also difficulties in identifying what useful gesture or posture was missing. Their method was to mainly examine what was at their disposal and use it, rather than searching what would be best suited and look if the application had it. What has been strongly asked but is technically difficult, is to be able to touch the other avatars, for a « tap », a « punch » or simply « shaking hands ». This suggests to add new actions that involve physical touch.

The subjects used the scene in different ways. The bar has been used by R and J in a proxemic approach. They tried to lean on the bar next to each other as if they were having a drink, and used, in this position, mainly facial expressions. We must notice that they were trying to use the bar as a nonverbal communication artefact and they were able to adapt the channel used to the position of the avatar.

A fundamental need emerged from the collection of impressions of the participants : the presence of a visual feed-back. Without being able to « feel » the posture of their avatar, they strongly asked for the possibility to view their own body. The caricature-like aspect of the gestures and postures has also been emphasised. But the probability is high that any predefined action would be considered so, or would not be understood easily enough if the visual clues were to be weakened. The main point is that predefined actions cannot, by definition, be finely adjusted to the specific ongoing interaction.

Finally, we have to recognise that, beyond these encouraging results, the quantity of nonverbal information that the user can provide with our current solution and the subtlety of the proposed actions, should be much more. The subjects always wanted to decode signals that were not present or just suggested. That's because several mechanisms are still missing : « illustrators » should be present, lips movements should follow the speech, the orientation of the eyes and of the head should be properly controlled. The most used posture during the experiment is by far the « attentive » posture. The users explained to us that it was their only way to express their interest in what the other was saying, as they could not use their look to do so. They accepted the fact that this attitude was largely beyond what they wanted to express, because they felt that this feed-back given to the speaker was crucial for the continuing of the interaction. This example illustrates very well the current state of our work: we have given the users the possibility to send important messages to their interlocutors that they couldn't send before, but in a rather raw and limited way.

7. Conclusion

In this paper, we have discussed the importance of realistic embodiment and nonverbal communication for the Networked Virtual Environments. The inclusion in our work of a social sciences concern, has allowed us to better take in account this fact and helped us make decisions about it. Our development fulfilling the need of the inclusion of nonverbal communication in the VR systems, is only one of the possible solutions but we think it has interesting technical advantages and has allowed us to test our work and ideas. The evaluation of our solution has raised interesting points that we are planning to develop more in the future. A larger scale experiment would hopefully allow us to confirm our current conclusions and could give other valuable results.

We think that the path leading to a natural and realistic inclusion of nonverbal communication in Networked Virtual Environments is long and challenging, but crucial for the quality of the face-to-face interactions within these environments.

Acknowledgments

We are grateful to Beatriz Dias for her observations during the experiments, Luc Emering for his help in using the Agentlib library for playing gestures, Mireille Clavien for designing the gesture motions, and Ronan Boulic for his walking model. We also thank the assistants at LIG and MiraLAB for their contributions in the human models. This work is partially supported by European ACTS COVEN and VIDAS projects, and the Swiss SPP Fund.

References

[Argyle88] Michael Argyle, *Bodily Communication*, New York: Methuen & Co., 1988.

[Benford97] S. D. Benford et al., « Embodiments, Avatars, Clones and Agents for Multi-user, Multi-sensory Virtual Worlds », *Multimedia Systems,* Berlin, Germany: Springer-Verlag, 1997.

[Boulic95] R. Boulic et al., « The HUMANOID environment for Interactive Animation of Multiple Deformable Human Characters », *Proceedings of Eurographics'95*, 1995.

[Capin97] T. K. Capin et al., « Virtual Human Representation and Communication in VLNET Networked Virtual Environment », *IEEE Computer Graphics and Applications*, March 1997.

[Corraze80] Jacques Corraze, *Les communications nonverbales*, Paris: Presses Universitaires de France, 1980.

[Ekman67] P. Ekman, W.V. Friesen, « Head and Body Cues in the Judgement of Emotion: a Reformulation », *Perceptual Motor Skills*, No. 24, 711-724, 1967.

[Miller76] Gerald R. Miller, *Explorations in interpersonal communication*, London: Sage, 1976.