Nonverbal Communication Interface for Collaborative Virtual Environments

A. Guye-Vuillème[1], T. K. Capin[1], I. S. Pandzic[2], N. Magnenat Thalmann[2], D. Thalmann[1]

[1]Computer Graphics Laboratory, Swiss Federal Institute of Technology, Lausanne
[2]MIRALAB-CUI, University of Geneva, Switzerland

Abstract: Nonverbal communication is an important aspect of real-life face-to-face interaction and one of the most efficient ways to convey emotions, therefore users should be provided the means to replicate it in the virtual world. Because articulated embodiments are well suited to provide body communication in virtual environments, this paper first reviews some of the advantages and disadvantages of complex embodiments. After a brief introduction to nonverbal communication theories, we present our solution, taking into account the practical limitations of input devices and social science aspects. We introduce our sample of actions and implementation using our VLNET (Virtual Life Network) networked virtual environment and discuss the results of an informal evaluation experiment.

Keywords: Nonverbal communication; Embodiments; Networked virtual environments; Social interaction in virtual environments; Emotional feedback

Introduction

Body postures and movements give substance to face-to-face interaction in real life. They augment spoken messages, by helping people express their feelings or thoughts through the use of their bodies, their facial expressions, their tone of voice and so on. Psychological studies have concluded that more than 65% of the information exchanged during a face-to-face interaction is expressed through nonverbal means [1]. Thus, a virtual reality (VR) system that is designed to approach the fullness of real-world social interactions and to give to its participants the possibility of achieving a quality interpersonal communication has to address this point [2,3].

Moreover, virtual environments are often referred to by users as cold, dehumanised places. Static avatars are also generally considered as lacking emotions [4]. Since nonverbal communication (NVC) has been identified by many authors as the most efficient way to communicate emotional content [5], we argue that the inclusion of this dimension of human communication could dramatically improve the comfort and quality of the experience lived by the participants.

A rather small number of developments concerning NVC have been made in Networked Virtual Environments (NVE). Some of them have focused on the automatic generation and scripting of nonverbal behaviours for autonomous agents [6,7], others on real-time interaction of human users [4]. Our primary goal was to offer a tool allowing a

human user to send basic emotional nonverbal messages. This is done by manipulating a small number of high-level parameters, so that the user need not know anything about the technical context. Our typical target application is a 3D chat system.

This paper presents our solution to NVC in NVEs with simple interfaces under constrained input conditions. The paper lays out our development process, starting with an outline of the requirements for the project, and ending with an initial evaluation of an implemented interface. We first survey embodiment in NVEs, present the basic theoretical background of the NVC field, briefly describe the implementation from a high-level point of view, and conclude with some observations from the evaluation experiment.

Embodiment in Networked Virtual Environments

In order to understand the 'body language' of other participants using his or her real-life decoding skills, the user clearly has to be able to identify a basic set of limbs on their embodiment. Moreover, the use of an articulated structure corresponding to a skeleton is well suited to and commonly used for body animation in 3D environments [8]. These elements have led us to use a complex avatar representation, which has the advantage of fulfilling several important functions that we shall now discuss, but also has some drawbacks.

Although a lot of research has been going on in the field of NVEs, most of the existing systems still use simple embodiments for the representation of participants in the environments. We consider that more complex embodiment is necessary for bodily communication and that it 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.

The avatar representation fulfils several important functions:

- the visual embodiment of the user,
- means of interaction with the world,
- means of sensing various attributes of the world.

It becomes even more important in multi-user NVEs, as participants' representation is used for communication. This avatar representation in NVEs has

crucial functions in addition to those of single-user virtual environments [2,3]:

- perception (to see if anyone is around),
- localisation (to see where the other person is),
- identification (to recognise the person),
- visualisation of others' interest focus (to see where the person's attention is directed),
- visualisation of others' actions (to see what the other person is doing and what is meant through gestures),
- social representation of self through decoration of the avatar (to know what the other participants' task or status is).

Using articulated models for avatar representation fulfils 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, allowing the user to use his or her real-world experience. We chose to use complex virtual human models aiming for a high level of realism, but articulated 'cartoon-like' characters could also be well suited to express ideas and feelings through the nonverbal channel in a more symbolic or metaphoric way (Fig. 1).

The use of complex models, virtual humans or cartoon characters, has a performance cost and can limit other aspects of the simulation. The main issues are the rendering speed and the network overload. Several techniques that we have implemented contribute to minimise the negative impact on both central processing unit (CPU) and network activity: levels of detail, predictive coding, dead reckoning, lossy and lossless data compression, etc. [8]. Taking into account the clear trade-off situation, we are also aiming for a high scalability of the articulated virtual human representation, allowing the customisation of the avatar depending on the application and the technical context (Fig. 2). Still, we think that
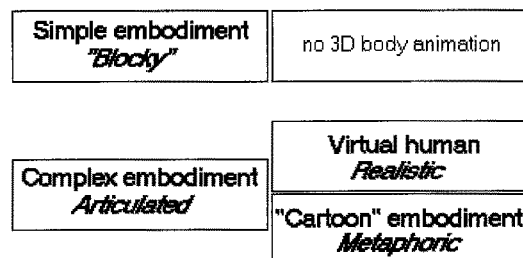
Simple embodiment "Blocky"	no 3D body animation

| Complex embodiment Articulated | Virtual human Realistic |
| | "Cartoon" embodiment Metaphoric |

Fig. 1. Embodiments typology for nonverbal communication (NVC).

A. Guye-Vuillème et al.

choosing the heuristic goal of handling highly complex and realistic embodiments is a good way to improve our techniques and knowledge.

Finally, we have to emphasise that controlling the articulated model 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 or her embodiment. In this paper, we survey these tools which help a user with a desktop VR configuration. We did not consider full tracking of the body using magnetic trackers, although our approach can be combined with limited tracking of the participant's arms.

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
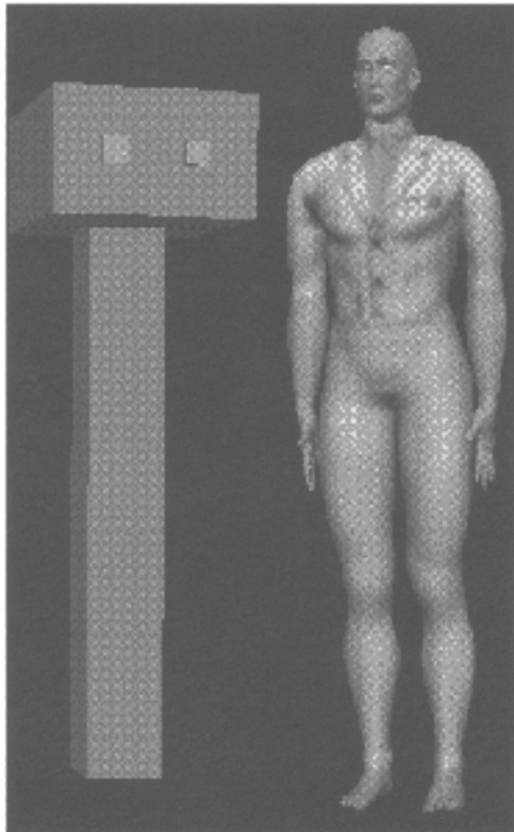

Fig. 2. Our 'blocky' Mister T and Peter, one of our complex models.

'nonverbal communication'. The definition of this field is based on an exclusion: one defines NVC as the whole set of means by which human beings communicate except for the human linguistic system and its derivatives (writing, sign language, etc.).

Corraze [5] proposed to distinguish between three types of information which are conveyed by NVC:

• information about the affective state of the sender,

• information about his or her identity,

• information about the external world.

To communicate this information, three main channels are used:

• the body and its movement,

• the artefacts linked to the body or to the environment,

• 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 analysis of the distance and angle chosen by the individuals before and during their interactions, the relationships associated with each distance, permission to touch and its circumstances, etc.

Kinesics includes gestures, postural shifts and movements of the hands, head, trunk, etc. and analyses what is sometimes called 'body language'. Three main types of bodily movements have been identified by several authors: the 'emblems', the 'illustrators' and the 'affect displays' (postures and facial expressions) [1,5]. The 'emblems' are gestures having a precise meaning that can be translated by one or two words: typically the nod meaning 'yes', thumb up for 'good', etc. Their knowledge is often specific to a group or subculture and their use is mostly conscious. The 'illustrators' are movements that are directly tied to speech, serving to illustrate what is being said verbally. They are difficult to describe but the fact that the amplitude of the gesture follows the loudness of the speech is typical of illustrators. Several authors [1,5] 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 one of its parts, compared with a determined

system of references. For example, the bodily attitude of prostration with the head bent and the shoulders falling is typical of a uneasy person. Other types of gestures are the 'regulators', which are used to regulate the conversation (e.g. showing the person who will talk next), and the 'adaptators', which are object or self manipulation related to individual needs or emotional states (e.g. scratching its head).

If a full description of the practices and context of use of these actions would go beyond this paper's purpose, it is interesting to underscore their relationship with speech and the degree of intention and awareness of their performers.

Often NVC is not used alone but jointly with verbal communication. In this case, it can be used by people as a means to signal importance or that the speech is finished, for example. The 'illustrators' and the 'regulators' are the types of gestures which are not used without speech and which are highly synchronised and combined with it. The concept of 'interactional synchrony' accounts for that characteristic, and several of Kendon's [9] studies give good examples of the high level of intricacy between the speaker's speech and actions, and the listener's nonverbal behaviours [10]. But according to Ekman and Friesen [11], there is a type of signal that is still independent from language: affective expression. It seems that NVC doesn't need any verbal expression in the task of communicating emotional messages, and that it is able to express in a powerful way things that would be very difficult to express using the linguistic system [5]. Postures and facial expressions are broadly independent from speech in the sense that they don't need it to convey emotions.

The intentionality of nonverbal actions, whether someone intends to send a specific message or not, is an important point. Some authors talk about 'communicative behaviour' and 'informative behaviour' to introduce this distinction [9]. Listed below are the different degrees of intentionality and awareness for the types we have identified:

- The use of emblems is intentional and the person is aware of what he or she is doing.
- The person using 'illustrators' is slightly less aware of what he or she is doing than with emblems.
- We are usually aware of our facial expressions and postures, but they may occur with or without a deliberate intention to communicate.
- Regulators and adaptators are on the 'periphery of awareness' [9].

This distinction is especially important when trying to include NVC in virtual environments, because standard user interfaces are a lot more appropriate for intended actions. Making the user responsible for handling the normally unconscious actions forces him or her to regularly analyse his or her feelings, which can be experienced as an unnatural task and necessitates a great deal of the user's attention. Full tracking of the body or multimodal interfaces are appropriate ways to handle the 'informative behaviours' but are not compatible with a standard desktop configuration.

Description of the Solution

Because no functionalities exist in the VLNET core system to handle the kinesic aspect of NVC, we decided to focus primarily on it. For the first stage of the project, we chose to give priority to the 'affect displays' (facial expressions, postures), requested by the users, and to the 'intended' actions ('emblems'), well suited to a 2D interface. The gestures needing a high synchronisation with the speech ('illustrators', 'regulators') have been temporarily put aside because of technical issues regarding synchronisation and also because they would have necessitated much attention from the user in a desktop configuration. It was decided to handle the other gestures ('adaptators': deep breathing, head-scratching, small movements of the hand and of the wrist) automatically and use them as a way to increase the sense of shared presence in the environment.

Cassel and Thórisson [12] argue that envelope feedback, constituted of nonverbal actions surrounding the conversion ('regulators'), is more useful than emotional feedback. We think that both aspects are important and must be implemented. The choice of using 'affect displays' is consistent with our goal of providing more 'friendly' virtual environments with emotional content. Moreover, studies have outlined the importance of emotions to ground social interaction [13]. Cassel and Thórisson [12] have based their claim on a specific speech-oriented application with very little emotional content (description of the solar system) and agree that emotional expressions may be effective in systems where the transmission of emotions is more central, e.g. a 3D chat application.

Because we wanted our solution to be usable with a desktop configuration, we decided to develop a 2D interface allowing the user to select predefined

actions. As formerly discussed, this approach is less appropriate for actions that are not always under conscious control, e.g. postures, but it seemed to us the best compromise between practical constraints and the will to include this aspect of human communication in a desktop environment

Selected Action

For the beginning of the project, we wanted a small number of gestures and postures (less than 30), so we decided to try to identify a basic 'palette' of actions, (see Table 1) which was a difficult task because NVC does not work as a linguistic system. The following criteria were 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.

The body postures and gestures come from a classic and commonly used sample of nonverbal actions, first developed by Rosenberg and Langer [14]. The postures we have selected illustrate very well the four fundamental postural attitudes described by W. James, in which the positions of head and trunk are essential: attitude of approach with the body bent forward ('Attentive'), attitude of rejection with the body turned away ('Rejection'), attitude of pride with the expansion of head, trunk and shoulders ('Determined'), attitude of prostration

with the head bent and the shoulders falling ('Insecure') [5]. The hand gestures were chosen because their cultural and geographical distribution has been intensively studied, e.g. Morris [15]. Finally, the sources of facial expressions are Miller's [16] and Ekman's [11] work.

This is only the starting 'palette' of actions we used for the evaluation experiment. These actions have the advantage of being well known by psychologists but cannot be considered sufficient. The application is 'open' and new actions can be easily added by users without programming knowledge and without recompilation.

The User Interface

In order to fulfil the need for an intuitive and easy to learn user interface, it was 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 (Fig. 3) and control panels. The panels offer a global view of all actions available, with clickable image buttons. They are constituted of several sections containing the actions classified by part of the body and 'emotional impact' (positive, negative, neutral). The user can also parameter the speed of execution of the action, and use keyboard shortcuts to run them. The high degree of organisation of the actions, combined with the fact that all actions are immediately activable, allow the user rapidly to find and execute the action that best fits the situation (Fig. 4).

The panels can be automatically attached and scaled with the VLNET view window for convenience. A 'mood setting' (cool, normal, stressed)

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	Insecure	Smile	Joy	Good
Angry	Puzzled		Bow	Bad

modifying the speed and frequency of gestures, and the possibility to watch and automatically follow other participants, have also been added (Fig. 5).

Integration in VLNET

For this project, we exploit our flexible framework for the integration of virtual humans in the NVEs, called VLNET (Virtual Life Network). The main VLNET process executes the main simulation and provides services for the basic elements of VEs to the external programs, called drivers [2]. The VLNET core consists of logical units, called engines. The role of the engine is to encapsulate one main function of the VE in an independent module, and provide an orderly and controlled allocation of VE elements.

Drivers provide the simple and flexible means to access and control all the complex functionalities of VLNET. Each engine provides a shared memory
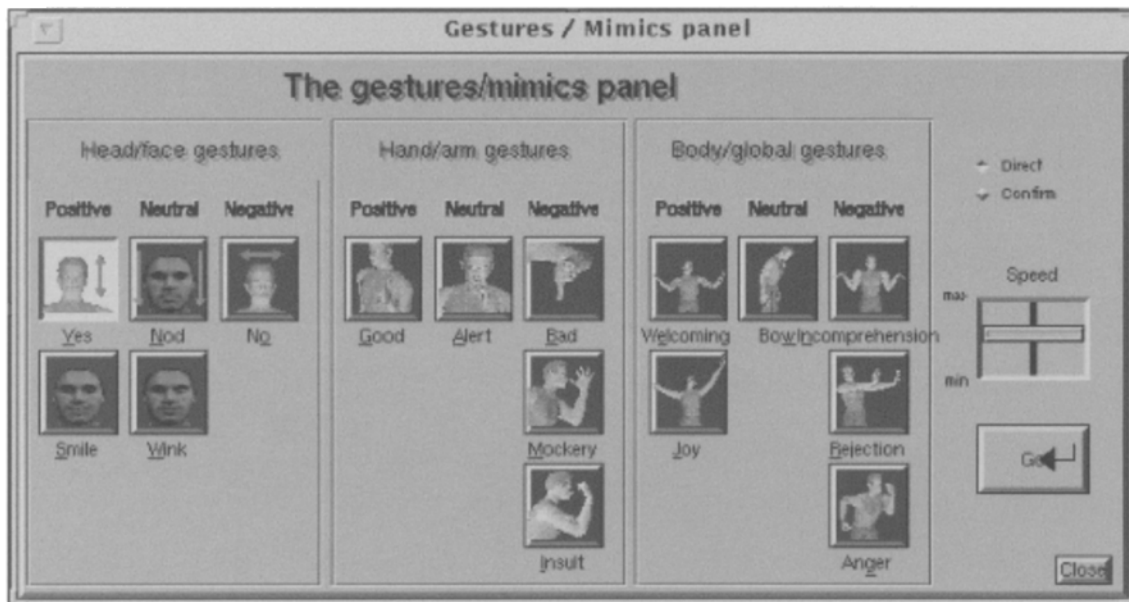

Fig. 3. The NVC application interface: the gesture panel.

Fig. 4. Example of use of the NVC application.
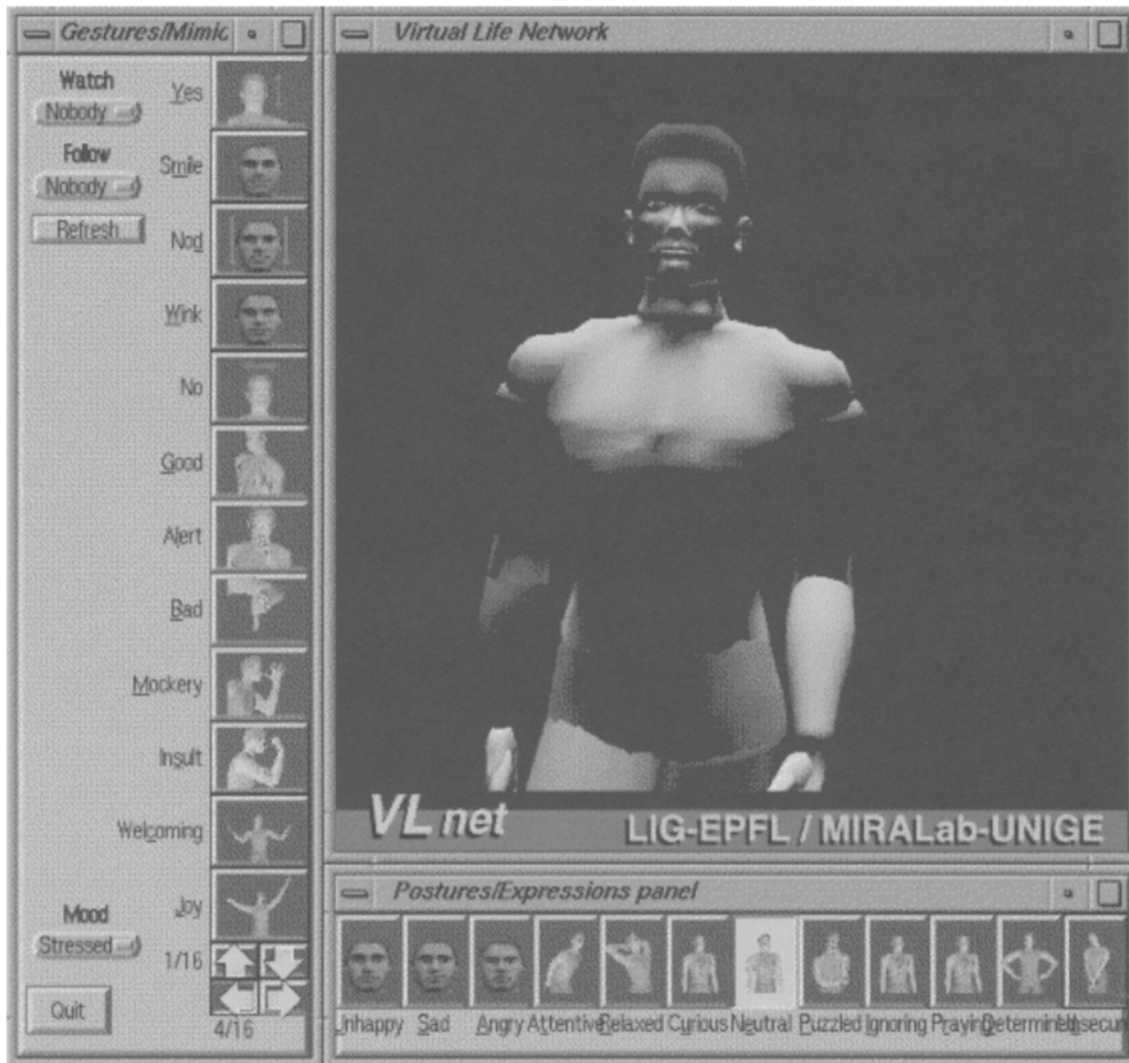
A. Guye-Vuillème et al.

Fig. 5. The compact NVC interface.

interface to which a driver can connect. The drivers are spawned by the VLNET Main Process at the beginning of the session. 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.

For the control of the virtual human body posture animation, an articulated structure corresponding to the human skeleton is used. Structures representing the body shape are 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 [17]. 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.

The system, VLNET core and NVC driver with complex embodiments, is designed for SGI workstations: low-end models (e.g. O2) are sufficient for three or less participants, but more powerful workstations (e.g. OCTANE, ONYX) are necessary for a higher number of users. Detailed performance data can be found in Capin [8].

Evaluation of the Nonverbal Communication Application in VLNET

The idea of evaluating the immersive aspect of the VLNET system and the contribution of the NVC application in realistic situations with external people, was present from the beginning. We needed a usability evaluation of the solution but were especially interested in observing how users would be able to handle social interactions using it

Organisation of the Experiment and Methodological Choices

The first decades of research in the NVC field have seen a wide use of laboratory experiments. Nowadays there is an increasing preference among psychologists for observing real and spontaneous behaviour [1]. In the Collaborative Virtual Environments field, the use of ethnographic methodology has given good results for evaluating applications and identifying typical practices [18]. These considerations motivated our choice of evaluation method. To encourage spontaneous behaviours and to reduce the impact of the researcher on the results, participants were free to act and interact as they chose. A small number of participants took part in the study, and 2 hours of interaction were recorded and analysed. Our analyses were qualitative in nature; careful observations of their interactions were taken, and their impressions were gleaned from a survey conducted at the end of the study. Because we wanted to have results sufficiently fast to guide us in the next developments and improvements in our solution, we didn't try to 'prove' a hypothesis but to identify crucial issues and behaviours. The hypothesis built on this small-scale experiment can then be verified on a larger sample and quantitative analysis can be done.

One of our main interests in carrying out this study was to establish whether the users, using the nonverbal tools at their disposal, could replicate their relationship with other participants. Thus, the degree of intimacy with each other has been our main criterion in selecting the subjects for the experiment. We chose six participants, none of whom were computer scientists: two were female (R and L) and four male (J, J2, T and R), two were very familiar with each other, two were acquaintances, and two were strangers to each other. After an introduction to the system, they were given total freedom of action, being allowed to talk with each other or stay silent, explore the scene or stay at the same place, use NVC or not. Three systems were at their disposal for interacting: a navigation system allowing their avatar to walk freely in the environment, rotate, etc.; the NVC application with its 30 actions; and a microphone and headphones for verbal communication. The scene we used represented a square with a bar at its centre and was chosen for its public and socially oriented characteristics (Fig. 6). SGI OCTANE/SI (175 MHz) and OCTANE/MXI (195 MHz) workstations on a 100baseT network were used for the experiment.

Main Observations

Use of the Nonverbal Interface

Users had no difficulty in using the interface. After a couple of minutes, all were used to it and J and R even started to 'play' with it, trying to run several actions simultaneously. The participants used some actions a lot more than others, the main point being that they used many more gestures than postures. A posture was very often chosen at the beginning of the interaction, but it stayed a long time as the participants didn't think of changing it. This can be explained by the fact that postures are often chosen unconsciously, as formerly discussed. In the survey, all users had difficulties in identifying what useful gesture or posture was missing. Their method was to examine what was at their disposal and use it, rather than searching what would be best suited and check if the application had it. What has been

Fig. 6. Two subjects interacting at the bar.

A. Guye-Vuillème et al.

strongly requested is the ability to touch the other avatars, tap, punch or simply shake hands. This suggests we should add new actions that involve physical touch.

The study was divided into sessions during which the NVC application was active or inactive. In the survey, the periods without NVC were rated as 'boring' by the majority of users. Typical expressions used by the participants to describe the influence of the application on their experience are: 'it was funny' (R, L and J), 'added something' (T), 'the whole scene seemed more life-like'(J). The inclusion of emotional content was rated by all users as 'useful' or 'pleasant'. This is an encouragement for us to keep on working in this direction.

Because we chose actions that don't need to be highly synchronised with speech, the users had few problems of this kind. It is mainly the 'yes' and 'no' gestures that they wanted to run at the same time as the corresponding words. If the delay of the nonverbal signal was long at the beginning of the experiment (several seconds), it significantly reduced when they got used to the interface (approximately 1 second). The 'attentive' posture was frequently used as a way to indicate to the speaker that one was listening to him, and the 'puzzled' posture when questions were asked. This confirms that regulating speech is also a very important function of NVC.

Another fundamental need emerged from the participants' impressions collection, the presence of bodily feed-back. Without being able to 'feel' the posture of their avatar, J, L, R and J2 strongly asked for the possibility to view their own body during the experiment. But this solution could take away some of the immersion feeling because the user can see himself or herself as totally exterior to the situation. A strategy they used was to ask other participants about their own appearance. The simulation of proprioception is a difficult challenge for VR researchers, but crucial for a quality immersion in the virtual environment and control of the avatar.

The caricatured aspect of many gestures and postures was also emphasised in the survey. We are thinking of using a mime or an actor to produce more realistic actions. But the probability is high that any predefined action would be considered caricatured, 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. But according to the users report, this caricatured aspect of actions was disconcerting only at the beginning of the experiment.

Then, the users got used to it, and used these actions for their symbolic meaning.

The Importance of 'Agreement'

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, if it was not really their interlocutor, could at least '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 said 'your representation', 'your character', etc. but always used the 'You' pronoun 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. A sentence used by R shows very well the particular relationship that showed up between the individual and his or her avatar: 'Look how I'm smiling!'. One can see very well here the acceptance of the avatar as a representation of the 'self' but also some distance because such a sentence clearly cannot be heard in real life.

We think that this 'agreement' is crucial for the quality of interactions in virtual environments. The obtaining of this 'agreement' depends partly on the participant, on his or her desire to interact or on his or her familiarity with technology for example, but also on what is 'offered' to him or her, e.g. the quality of embodiment. At the VR technology current level of development, it would be hazardous to try to mislead the participants and make them believe against their will that they are physically in another place. What must be done is to try and obtain their active collaboration by the inclusion of such mechanisms as gestures for example, so that they can 'play the game' of interaction and thus participate in the building of a rich virtual reality. Confronted to the meaningful behaviours of his interlocutors, the participant can finally partly forget the specificity of the situation and act in a natural way.

Reproduction of the Real-World Social Relationships

It is interesting to notice that the users have been able to reproduce, through the mechanisms of NVC, their relationship of the real world in the virtual environment. We observed that the subjects who didn't know each other before the experiment (R and T) 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 aggressive gestures while the other participants (who knew each other) used several times the 'mockery' gesture or the forearm jerk.

At another level, the NVC application allowed them also to respect the formal structure of social interactions. At the beginning of the interaction, they all used one of the actions to greet the other one ('Bow', 'Welcoming') and signal that they were ready to begin the exchange. The end of the interaction followed the same logic and was always confirmed by nonverbal means. The normative sanction produced when someone doesn't respect these rules in real life showed up as follows. R was speaking with J. R suddenly decided to explore the world and abruptly left J. J became angry and used verbal and nonverbal means (anger and insult gestures) to express it. R came back and they left to explore the world together.

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 off the experience. Later, the avatar of J2 (male) and L (female) were very close, nearly touching each other in a position that could have been interpreted as very intimate. A strong emotion was noticed on the participants, first in the form of uneasiness and then laughter. This behaviour was typical of the relationship between J2 and L: they had different gender identities and didn't know each other very well. The movements and positions of their avatar weren't 'free' because they had real consequences and this scene had nearly the same effect as if it had happened in real life.

A final example illustrates this 'real' effect of 'virtual' interactions. During the experiment, J became really angry because R wanted him to do something that he didn't want to. R refused to speak for a moment but used the 'forearm jerk' gesture in a totally sincere way.

Conclusion

Finally, we have to recognise that, beyond these encouraging results, the quantity of nonverbal information that the user can provide with our solution and the subtlety of the proposed actions, should be much higher. The sentence 'It's not funny, you're not moving!' (R) is typical of this record. In real life, you cannot stop communicating. During the experiment, the subjects always wanted to decode

signals that were not present or just suggested. This is because several mechanisms are still missing: 'illustrators' should be available, lips movements should follow the speech, orientation of the eyes should be properly controlled. 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.

Future Evaluation

We are planning to continue the evaluation of the system, and now briefly discuss the ideal evaluation. The best would be to build a representative sample of the population that would allow us to test with a large-scale experiment our preliminary results, and confirm or invalidate our hypotheses. It is a very exciting task for the future, since it is likely that with the current societal context and technological development, the number of social interactions in virtual environments will slowly grow and finally involve a significant part of the population. But it is a huge task. From another point of view, it could also be very interesting to evaluate our solution with a representative sample of the probable short-term users: the typical users of teleconferencing, for example, which would require the organisation of a lower number of experiments.

Whichever solution is chosen, there are some important correlation variables that must be taken into account. 'Age', for example, is certainly very important for the usability evaluation of the system, since many studies have shown that it was highly correlated to the familiarity with using technical interfaces. In the same way, the level of comfort when using technology is probably very important at the moment to achieve a quality interaction in virtual environments. It could also be very interesting to test the system with a multicultural sample of users, so that we can check if the selected actions are really widely understood, and if they are understood in the same way. Finally, it would be helpful to use different embodiments, articulated and non-articulated, realistic and 'cartoon-like', and compare their relative influence on the interaction.

Conclusion

In this paper, we have discussed the importance of nonverbal communication for NVEs. The inclusion in our work of a social sciences aspect has allowed

us to make better decisions about representation of social behaviour. Our development, in fulfilling the need for the inclusion of nonverbal communication in 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 further in the future. A larger-scale experiment would hopefully allow us to confirm our current conclusions and could give other valuable results.

We are now improving some aspects of our solution, the need for which has been emphasised by the evaluation experiment. We think that the path leading to a natural and realistic inclusion of nonverbal communication in NVEs is long and challenging, but crucial for the quality of face-to-face interactions within these environments.

Acknowledgements

References

1. Argyle M. Bodily communication. Methuen, London, 1988

2. Capin TK, Pandzic IS, Noser H, Magnenat-Thalmann N, Thalmann D. Virtual human representation and communication in VLNET networked virtual environment. IEEE Computer Graphics and Applications 1997; 17(2):42–53

3. Benford SD, Bowers JM, Fahlén LE, Greenhalgh CM, Snowdon D. Embodiments, avatars, clones and agents for multi-user, multi-sensory virtual worlds. Multimedia Systems 1997; 5(2):93–104

4. Vilhjálmsson HH, Cassell J. BodyChat: autonomous communicative behaviors in avatars. In: Proceedings of the 2nd International Conference on Autonomous Agents ACM Press. 1998: 269–276

5. Corraze J. Les communications nonverbales. Presses Universitaires de France, Paris, 1980

6. Cassell J, Pelachaud C, Badler N. Animated conversation: rule-based generation of facial expression, gesture and spoken intonation for multiple conversational agents. In: Proceedings of SIGGRAPH'94 1994:413–420

7. Perlin K, Goldberg A. Improv: a system for scripting interactive actors in virtual worlds. In: Proceedings of SIGGRAPH'96 1996: 205–216

8. Capin TK. Virtual human representation in networked virtual environments. EPFL, Lausanne, 1998

9. Kendon A. Nonverbal communication, interaction, and gesture. Mouton, The Hague, 1981

10. Weitz S. Nonverbal communication: readings with commentary. Oxford University Press, New York, 1974

11. Ekman P, Friesen WV. Head and body cues in the judgement of emotion: a reformulation. Perceptual Motor Skills 1967; 24:711–724

12. Cassell J, Thórisson KR. The power of a nod and a glance: envelope vs. emotional feedback in animated conversational agents. Journal of Applied Artificial Intelligence 1999; 13:519–538

13. Cañamero D, Van de Velde W. Socially emotional: using emotions to ground social interaction. In: Papers from the 1997 AAAI Fall Symposium 1997: 10–15

14. Rosenberg BG, Langer J. A study of postural-gestural communication. Journal of Personality and Social Psychology 1965; 2(4): 593–597

15. Morris D, Collet P, Marsh P, O'Shaughnessy M. Gestures. Jonathan Cape, London, 1979

16. Miller G. Explorations in interpersonal communication. Sage, New York, 1976

17. Boulic R, Capin T, Huang Z, Morozet L, Molet T, Katra P, et al. The HUMANOID environment for interactive animation of multiple deformable human characters. In: Proceedings of Eurographics '95 1995: 337–348

18. Bowers J, Pycock J, O'Brien J. Talk and embodiment in collaborative virtual environments. In: Proceedings of the ACM CHI'96, ACM Press. 1996: 58–65

Correspondence and offprint requests to: *Anthony Guye-Vuillème, Computer Graphics Laboratory, Swiss Federal Institute of Technology,CH1015 Lausanne, Switzerland. email: aguye@lig.di.epfl.ch*