Virtual Human Animations in a Collaborative Virtual Environment

Arnaud Bouguet, Dominique Pavy, Pascal Le Mer,
Stéphane Louis dit Picard, Laurence Perron, Grégory Saugis

France Télécom R&D Lannion
2 avenue Pierre-Marzin
22307 Lannion Cedex – France
{arnaud.bouguet, pavy.dominique, pascal.lemer,
stephane.louisditpicard, laurence.perron, gregory.saugis}@francetelecom.com

Abstract

In a Collaborative Virtual Environment, distant people are able to meet one another, to work, to communicate or to play together. France Telecom R&D and INRIA/LIFL have developed a Collaborative Virtual Environment named "Spin-3D" where each user is visible as a 3D virtual substitute of his/her physical representation: an avatar. We think that animated avatars could be a good way to improve interpersonal communication and enhance the telepresence in a collaborative task.

In our architecture, an intelligent control engine allows switching seamlessly from one animation to another according to remote users' actions or behaviours. We use real-time body and facial animations from pre-computed animations to context-dependent gestures.

This article describes the Spin-3D platform and our avatar animation architecture, designed to increase the effectiveness of virtual collaboration.

1. Introduction

Networks and Virtual Reality allow geographically distant people to meet one another, to communicate and to perform collaborative tasks in a Collaborative Virtual Environment (CVE). Historically, two main approaches of user representation have competed or have been combined: "Live" video image or synthetic avatar.

A few years ago, the lower bandwidth availability was the first argument to prefer 3D body representation instead of video camera for on-line communication. Today access to bandwidth necessary for teleconferencing is easily available. Nevertheless in a Collaborative Virtual Environment having 3 or 4 people simultaneously, the video solution can limit the diffusion on the network of other data. Also, the availability of automated tools allows modelling realistic anthropomorphic avatar (or clone) without the skills of infographists, and real-time animations tempt to reach realism as in video games.

In a synchronous collaborative task the communication is essential. Applications proposing multimodal communications by text, audio and video are widespread, and some of them include synthetic avatars. However, often the role of an avatar is limited to some "extras"; it could also be a good way to improve communication, especially nonverbal communication. The nonverbal part of the communication is very important in a face-to-face conversation. It should increase the quality of the collaborative task if it can be provided an animated feedback of the nonverbal aspects of human interaction and communication. A video-based solution offers a part of the nonverbal aspects, however, a video is not easy to manipulate and doesn't provide the same synthesis capacities as a synthetic avatar.

In this paper first, we introduce our CVE named "Spin-3D" and our work on user representation with animated avatars. This platform as we know it today is the result of research and development work by France Telecom R&D Lannion and INRIA/LIFL. Next, we present some previous work in nonverbal communication, user representation and the Spin-3D platform. In section 3 we describe the avatar animation architecture. In particular we explain the intelligent part of our system able to switch from one animation to another according to remote users' actions or behaviours. Finally we present some results and future work.

2. Previous work

2.1 The nonverbal communication

In the seventies many authors proved how important the nonverbal communication is. According to Morris et al [1] and Knapp [2], about 65 per cent of a face-to-face conversation takes place as nonverbal. Ekman & Friesen [3] showed that the face is the most significant indicator of the emotional state of a person. According to Argyle [4], the incoming nonverbal communication body codes are body contact, closeness, posture, physical appearance, facial and gesture, direction of gaze, aspects of speech. Most of this research leads to prove the great importance of the nonverbal communication.

2.2 User representation

In classic 3D multi-users environments user representation is made by a synthetic 3D avatar. There are stylized avatars (like the 3D "blockies" of Massive [5]), anthropomorphic avatars (Dive [6]) or realistic clones. Also, with new technologies, people are more and more used to see 3D virtual humans, for instance in animation movies.

What is best for the collaborative task to have realistic or stylized avatars is a question we may ask. In 1995, Benford & al [7], suspected that achieving realism was wasteful and non effective for the collaborative task. In particular in CVE, realism is probably not justified. It's why, in our approach we have chosen not to focus on the realism of the avatars.

2.3 The "Spin-3D" platform

Often 3D multi-users worlds are realistic representations of our real world. Even if the result may seem pretty, it can be difficult or boring in the longer run. It's difficult for users to focus on a collaborative task if at the same time they have to manage movements and objects manipulation. In 1998, works of Dumas et al [8] brings a new approach to CVE with new ways of interaction and communication: "Spin-3D" (Figure 1).

This platform [9] uses a meeting-room metaphor and allows a small number of people (up to five) to work together around a virtual table, interacting with shared 3D objects or tools (Louis Dit Picard et al [10]). The focus is on collaboration and not on introducing new hardware, thus we opted for simple devices: a classic way is pointing and selecting by a mouse (represented by a 3D pointer) and manipulation with a Spacemouse[1] or keyboard. The Figure 1 presents a classical collaborative session between three users. Each user is represented by his/her name, a colour, a 3D animated avatar and a 3D telepointer (the distant representation of the local 3D pointer). A person in the session will not see his/her own avatar, only those of remote users.

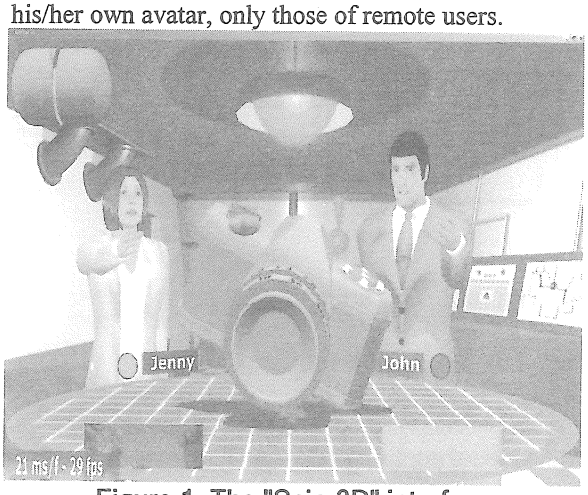

Figure 1. The "Spin-3D" interface

The voice is carried and spatialized in our application, and we have early integrated the results of the nonverbal communication (Le Mer et al [11]). In the collaborative tasks, user embodiment through animated avatar provides information to easily understand remote users' actions and behaviours and permits to answer questions like "who is who?", "who is doing what?", "who is talking?" just to mention a few.

Using services for the virtual meeting management, the Spin-3D platform offers the possibility to define synchronous collaborative applications like cards games, CAD reviewing or 3D Medical Imagery. In addition to 3D objects (geometry and behaviour), interactions or specialized display plug-ins, the application builder has to define rules to animate automatically avatars of his/her application.

3. The avatar animation architecture

Figure 2 presents the local user and his/her distant representation. In the next sections we will see more in details the available *animation modes*, the *User Engine* and its animation *rules*, the *Shared User State Vector*, the *Remote User Engine* and the *Animation Engine*.

[1] http://www.3dconnexion.com/

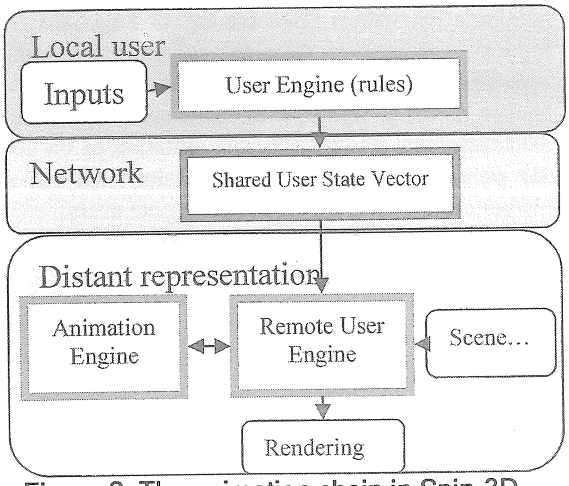

Figure 2. The animation chain in Spin-3D

3.1. The animation modes

Our system provides several animation modes. Main modes are listed below:

- Playing a pre-computed body animation, body posture or automatic body life signs.
- Looking at something or someone.
- Showing something or someone with right or left arm.
- Playing a facial expression, a facial behaviour or automatic facial life signs.
- Animating separately eyelids, eyes, neck or lips.
- Interpreting motion capture data by video analysis.

Of course it is possible to add modes to this list. Also modes combination and blending animations are allowed: as an example playing a pre-computed body animation, playing a facial behaviour and showing something in the same time is possible.

We work on interpolation algorithms to naturally switch from one animation to another without visual jump. Sometimes, it could be more intelligible for user to set neutral position to the avatar before moving on the next animation. Even if we don't necessarily search to do realistic or naturalness movements we try to respect existing model of human psychomotor behaviour such as Fitts' Law. Future experiment with users should prove the efficiency of our choices.

3.2. The User Engine

Thus the application builder has several animation techniques. Now, the challenge is how to intelligently switch from one animation to another. The User Engine, considered as the brain of the local system, drives automatically the distant avatar animation according to rules defined by the application builder. Rules depend on:

- Hardware devices
- Voice
- Camera, tracking
- Spin-3D, plug-ins, application events
- Scene, data, context, historic
- Others sources of information

Rules have to make information selection, manage conflicting actions and define relevant animations for the collaborative application to easily understand remote users' actions and behaviours. Table 1 shows some examples of local actions/behaviours and corresponding interpretation and remote representation.

Table 1. Examples of Observations/Interpretations/Representations

Information from application events	
Observation	The user has selected the designation tool and he moves his/her 3D pointer
Interpretation	The user wants to show us something
Representation	Making a deictic gesture with right arm, head and gaze orientation
Information from Spin-3D platform	
Observation	The sound module detects a sound in the microphone
Interpretation	The user is speaking
Representation	Loading facial animation, automatic life-signs
Information from Spin-3D platform	
Observation	The application is just opening
Interpretation	The user is entering to the session
Representation	Playing a symbolic pre-computed animation of "hello"

Thanks to observations in real situation, we actually try to define a set of generic rules for all kinds of collaborative application. However rules also depend on the application and it should be possible to make specialization of the User engine according to the application.

3.3. The Shared User State Vector

Each user is associated with an object named: a Shared User State Vector. As an example, for a session with 3 users, there are 3 Shared User State Vectors. It represents the state of a user at a given time and it is accessible through the network by others users of the collaborative session. Locally all User Engines (from each user) automatically updates his Shared User State Vector according to his observation and interpretation (rules). The Shared User State Vector has to contain the appropriate information to permit a good representation. Some examples of information stored are given below:

- The absolute location of the 3D pointer.
- The relative location of the 3D pointer (when it is in an object).

- The animation mode.
- Identifier of the selected object.
- Identifier of the pointed object.
- The location target for right arm.
- The location target for the left arm.
- The location target for head orientation.
- Identifier of the pre-computed body animation.
- Identifier of the facial expressions or behaviour.
- A flag speech or not speech.

According to the rules and the application, all fields of the Shared User State Vector are not required and updated. All remote users consult the Shared User State Vector in the Remote User Engine.

3.4. The Remote User Engine

On remote computers, the Remote User Engine reads the data contained on the Shared User State Vector. The information from each user's animation is treated by the Remote User Engine which authorizes changes according to scene, data or context. For example, the Remote User Engine can convert an identifier of an object to a 3D space position (spatial conversion due to the different environment configurations of each user). The Remote User Engine displays and animates the avatar and the 3D telepointor thanks to the Animation Engine.

4. The Animation Engine

In our applications based on the Spin-3D platform, 3D objects are described in VRML97[2] format. Logically our avatars follow the VRML97 and H-Anim[5] description, but only on the hierarchy, not for the animation. We consider that H-Anim animations are not easy to manipulate and not flexible enough. We have decided to develop our own solution of animation to control all elements of the animation chain.

The animation engine permits to animate the head and the body of the 3D avatar. It supports blending context-dependent gestures and pre-computed animations in real-time. Actual capacities of the animation engine correspond to main animation modes listed above in the section 3.1.

4.1. Pre-computed animations

Observations in real situation of collaboration reveal that some significant gestures or actions are recurring (Perron[12]). It is interesting to create a library of relevant pre-computed poses and animations that can be replayed automatically, randomly or voluntarily during the collaborative session. At the time of our first works, market solutions were not suitable because they were often incomplete, not flexible, not open and not easy to use by non specialist in computer like ergonomists. Thus, we have decided to develop an authoring-tool named "Vestibule". Figure 3 shows the Vestibule interface:

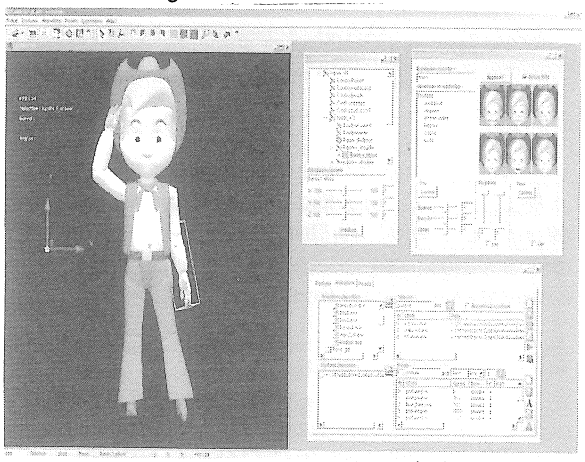

Figure 3. The Vestibule interface

In changing the local position and orientation of each joint of the hierarchical tree we define poses in the same way we can see in modellers. An animation is defined as a sequence of poses associated with a time. Our animation engine is able to go from one pose to another with several means of interpolations (linear, accelerated, and decelerated). Furthermore, using the same structure (joints and segments), an animation realized with an avatar can be applied to another. Vestibule allows functionalities such as animations blending or time deformation.

Figure 4 presents an example of use: a remote user is manipulating a part of the shared object. The User Engine interprets this action and asks to the Remote Engine to start a symbolic action of manipulation. This animation has been created with Vestibule and replayed during the session.

Figure 4. An example of pre-computed animation, replayed in real-time in a collaborative session

[2] http://www.web3d.org/x3d/specifications/vrml/

4.2. Context-dependent gestures

We are able to dynamically create animations according to the scene configuration. Here are some examples of context-dependent gestures:

- To touch an object.
- To look at something or someone in the scene.
- To cast a glance at an object without moving the head.
- To do a deictic gesture.
- To orientate the arm to an object.

We develop specific algorithms to interpolate movements and we use Inverse Kinematics to orientate the arm according to a target (Inverse Kinematics algorithm from Bertel [13]).

4.3. Facial animation

We use a real-time facial animation engine named "FaceEngine" (Breton et al [14]). FaceEngine is an animation system using both muscular and parametric animation. Based on this library we defined a simple interface to control the user head. The main functions include eyes, eyelids and neck movements, facial expressions (anger, disgust, fear, happiness, sadness, surprise, etc), behaviours and automatic life signs.

During a session when a user speaks on his/her microphone, we animate the lips of his/her distant avatar. The main difficulties are the synchronization of lips with the speech and the necessity to make full phoneme segmentation in real-time. As it has been previously mentioned we have chosen not to focus on the realism, for that reason we actually use a simple solution measuring the level of the sound signal to deduce the information "speech" or "not speech". When speech is detected we play automatic life signs and random visemes.

When several users are in the environment, this simple lips animation provides sufficient information to know who is talking. Even if facial animation is not precise enough to read the lips, it still provides a strong, essential nonverbal communication feedback.

4.4. Motion capture by video analysis

Video analysis permits to not disturb the user with constraining technologies. Thanks to the work of Bernier et al [15] we are able to track the face and the hands of a person in near real-time (12Hz) using two cameras and a single workstation. Information of 3D locations (hands and head) can be read on remote computers to animate the avatar. Actually the system is limited by the conditions imposed on the background and clothes colour. Future objectives are a system using a simple camera and including more 3D information like the arms and the torso of the person.

This system provides a mimetic animation and his utilization is limited within the context of collaborative activities. Nevertheless it could be interesting in collaborative games or communications like 3D chat with an avatar in real size.

5. First results

As we saw before, many applications could be developed on the platform: collaborative cards games or shared CAD reviewing, just to mention a few. We have developed, in partnership with IRCAD[3] a medical 3D collaborative application: "Argonaute 3D" (Figure 5), based on the Spin-3D platform.

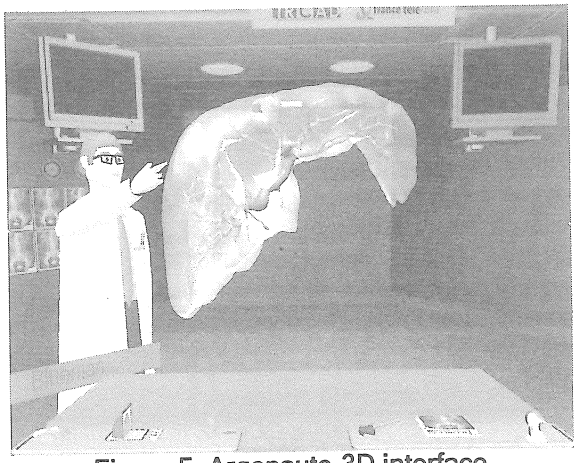

Figure 5. Argonaute-3D interface.

This application allows medical practitioners located in different places to work in real-time on a 3D representation of a patient's body suffering from liver cancer. The goal of this application is to create a diagnosis and surgical intervention strategy for the patient, using shared tools. The interface provides means to manipulate organs, visualize, navigate in blood vessels or simulate resection or laparoscopy. This application has been successfully tested between Paris, Strasbourg, Brest and Lannion in France on classical ADSL networks or between Lannion and Tokyo on internal networks. In this first version avatars were reduced to a simple static presence because we wanted to emphasize the medical aspects. Communication was essentially done by voice transmission.

In 2004 at the Brest Hospital in France in an experiment led by Tapie[16], different groups of practitioners tested the application. Some used the "Argonaute-3D" application without avatars, others with static avatars and others with animated avatars. The purpose was to improve avatar role and test some hypothesis, in particular: could avatars improve the communication and the collaborative task?

[3] Institute for Research into Cancer of the Digestive System

We modified the application so that, when a user uses his/her mouse, a deictic movement of his/her arm and his/her head was associated to the avatar, giving him an impression to show something in the scene, for example tumours in the liver. When a user does nothing for a given time, his/her arm and head returns to the original position with an interpolated movement. In Figure 5 the remote practitioner has selected a manipulation tool. He shows something in the liver. His/her avatar makes a deictic gesture (arm and head movement) according to the target.

First results indicate that presence of the avatars reduced the time used by the subjects to present them and to explain what they were doing. Nevertheless users seem not really conscious of that.

6. Conclusion & future work

As it has been previously demonstrated nonverbal communication is important in face-to-face conversation. Compared to classical audio and video systems, avatar-based communication can represent nonverbal aspects of human interaction and communication in a CVE.

According to this observation we have developed avatars animation chain, in our Spin-3D platform, able to increase the effectiveness of the virtual collaboration.

Based on first promising results we will focus our future works to two main points. First we would like to upgrade our user engine, the "brain" of the animation chain to take more into account the nonverbal communication studies, gestures classification or observations in real situation. Especially we are interested by the inputs combination between devices and scène events according to one situation or application. Secondly we need an improvement of our animations techniques (pre-computed animations, context-dependent gestures), including smooth skinning and constraint animations to enhance the "behavioural" realism of avatars.

7. References

[1]. D. Morris, P. Collett, P. Marsh and M. O'Shaughnessy, "Gestures: their Origin and Distribution", New York: Stein & Day, 1979.

[2]. M.L. Knapp "Nonverbal Communication in Human Interaction", (2nd ed.) Holt, Rinehart and Winston Inc., New York, 1978.

[3]. P. Ekman and W.V. Friesen "Unmasking the face", New Jersey, Prentice-Hall Inc, 1975.

[4]. M. Argyle, "Non-verbal communication in human social interaction" in Hinde, R. (ed.) Non-Verbal Communication, Cambridge, 1972.

[5]. C. Greenhalgh and S. Benford, "MASSIVE: a Distributed Virtual Reality System Incorporating Spatial Trading," in Proc. IEEE 15th International Conference on Distributed Computing Systems (DCS'95), Vancouver, Canada, May 30 - June 2, 1995, IEEE Computer Society.

[6]. C. Carlsson and O. Hagsand, "DIVE - A Platform for Multi-User Virtual Environments", Computers and Graphics 17(6), 1993

[7]. S. Benford, J. Bowers, L. E. Fahlén, C. Greenhalgh, and D. Snowdon. "User embodiment in collaborative virtual environments." In CHI '95: Proceedings of the SIGCHI conference on Human factors in computing systems, pages 242–249. ACM Press/Addison-Wesley Publishing Co., 1995.

[8]. C. Dumas, S. Degrande, G. Saugis, C. Chaillou, M.L. Viaud and P. Plénacoste, "Spin: A 3-D Interface for Cooperative Work", Virtual Reality, Springer-Verlag, 4, p. 15-25, 1999.

[9]. D. Pavy, A. Bouguet, P. L. Mer, S. L. D. Picard, L. Perron and G. Saugis (from France Telecom R&D), C. Chaillou, and S. Louis Dit Picard (from LIFL USTL/CNRS/INRIA), "Spin-3d: a VR-platform on internet ADSL network for synchronous collaborative work". In e-Challenges 2004: Proceedings of "eAdoption and the Knowledge Economy" Issues, Applications, Case Studies, volume 2, page 1486. IOS Press, 2004.

[10]. S. Louis Dit Picard, S. Degrande, C. Gransart, G. Saugis and C. Chaillou, "A CORBA Based Platform as Communication Support for Synchronous Collaborative Virtual Environments", International Multimedia Middleware Workshop (M3W), 2001.

[11]. P. Le Mer, L. Perron, C. Chaillou, S. Degrande and G. Saugis, "Collaborating with Virtual Humans", People and Computer XV – Interaction without frontiers: Proceedings of HCI 2001, Springer, pp.83-103, September the 12th 2001, Lille (France).

[12]. L.Perron, "What kind of gestures to animate an avatar?", Lannion, France, France Télécom R&D, Internal Report, 2004.

[13]. F. Bertel, "Humanoid animation in a conversational context involving verbal and non-verbal dialog", PhD thesis, University of Rennes 1, France, 2003.

[14]. G. Breton, C. Bouville, and D. Pelé. "FaceEngine a 3d Facial Animation Engine for Real Time Applications". In Web3D, pages 5–22, 2001.

[15]. O. Bernier and D. Collobert, "Head and hands 3d tracking in real time by the EM algorithm". In RATFG-RTS '01: Proceedings of the IEEE ICCV Workshop on Recognition, Analysis, and Tracking of Faces and Gestures in Real-Time Systems (RATFG-RTS'01), page 75. IEEE Computer Society, 2001.

[16]. J. Tapie, "Feedbacks about Argonaute-3D experiment", Lannion, France, France Télécom R&D, Internal Report, 2004.