VIRSTORY: a Collaborative Virtual Storytelling

Atman Kendira, Laurence Perron

France Telecom R&D Lannion
2 avenue Pierre Marzin
22307 Lannion - France
{atman.kendira, laurence.perron}@orange-ft.com

Abstract. From the results of two Human behavior studies in small group interactions we constructed iteratively a Collaborative Virtual Environment named "VIRSTORY". This system is a digital storytelling with speech and 3D gesture recognition technologies using like input devices. It includes several modules: multimodal interaction module, behavior module to animate autonomous expressive characters, etc.

1 Introduction

Our aim is to carry out some experiences and use the results to design a multimodal Collaborative Virtual Environment (CVE) and particularly storytelling environments. Our recent studies are focused on virtual collaborative games into small groups. In these particular games, the users are represented by characters in narrative digital environment and it is primordial to animate the characters in the way that they are assisted by the Human communication especially non-verbal behavior: gestures, gaze, life signs, social attitudes, emotion, etc. Voice and gesture are both the basis of natural dialogue between players and the bimodal interaction support of the digital storytelling setup. This double function is very interesting for the users but remains a Human Computer Interaction research field and more specifically multimodal interfaces [1,2]. In fact, we are focusing on voice and gesture interaction because these modalities give the user more freedom to create and participate in the story narration [3]. In the first part of this paper, we investigate through two human behavior studies a way to permit the "natural" mediated collaboration between users in creative CVE. In the second part, we present the improvements of CVE called VIRSTORY that to take into account the results of studies.

2 Storytelling experimentations and CVE design

At first, we imagine a narrative game for everybody. This game uses cards to illustrate a piece of story e.g.: princess, wolf, etc. Each user must play with her/his pieces of story and with the other users to elaborate a nice story. This game was the starting point of VIRSTORY design.

If the primary aim is to conceive VIRSTORY, the secondary aim is to understand Human behavior especially the non-verbal behavior like gaze, facial expressions, and gestures of hand to animate a character in CVE [4]. Existing character animation techniques in CVE prompted us to study sensorial deprivation along two lines of reflection. We wished to establish how subjects interact in a group, and what they do when they are deprived of a form of expression (eyes, hand, and face) as the case with the intrusive device in CVE (cf. Figure 1). The idea was that subjects compensate for the absence of one of these forms of expression either verbally or by means of other non-verbal behaviors not under constraint in the experimental situation. Observations of this experiment showed that characters can already be animated with recurring self-contact gestures and these gestures can be used by example to create animations library. Moreover, these gestures contribute to the feeling of presence that users expect when facing a character without speech analysis.

Fig. 1. Four experimental conditions. 1: without deprivation (normal) - 2: with masks to deprive subjects of face expressions - 3: with very dark sunglass to deprive subjects of gaze modality - 4: with their hands under the table to deprive subjects of gesture modality

In VIRSTORY game (cf. Figure 2) when the user takes a cube, the distant user can see the character with the cube and the colour feedback on the cube. If the character is red then the cube is red, etc. On the bottom of a user's screen is displayed the "*private space*", consisting of the cubes the user has not yet used in the story (the own cubes of the distant user are not visible). Only cubes already used by either user are visible by both users. These cubes lie on the yellow points that represent the common storyline. The first point is the beginning and the last point corresponds to the end of the story. When a user selects a cube from his private space and moves it to the storyline, this action is seen by both users. As a user moves a cube on the beginning of the storyline and tells "once upon a time", the second user finds a cube to continue the story and so on.

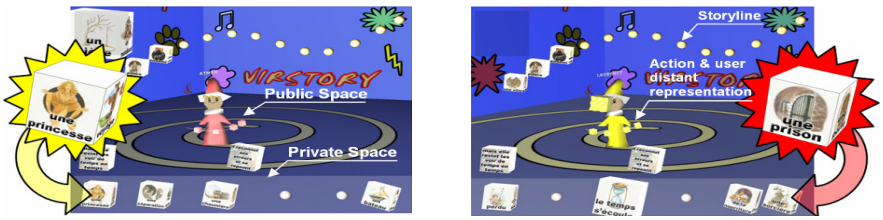

Fig. 2. The VIRSTORY CVE: on the left, the first player – on the right, the second player

We realized a “Wizard of Oz” experimentation (WOZ) with 20 subjects (9 males and 11 females, 16 adults and 4 children) (cf. Figure 3). This experimentation allows understanding the organization of the oral and gesture modalities between themselves, as a function of the habits and competences of the subject (adult/child difference for example). One of the interests of multimodal interfaces is to let a user choose the way to interact as she/he prefers. If pointing with a laser is efficient when this tool is given to a user, the question remains open whether such efficiency is kept without a physical pointing device, for example with computer vision pointing gesture recognition. The experimenter only reproduced gestures performed on objects of the storytelling setup, while other gestures such as wrist or arm rotation to turn a cube were interpreted.

Fig. 3. On the left: a child with a laser pointer. On the right: The WOZ experimenter carrying the child’s command and the partner in the background playing with the child

To conclude, such an experiment allows defining a spoken vocabulary, to observe, depending on the context, the importance of a given modality over other modalities. The most important lesson learnt from these experiments is that modality influences the cooperation. The more the subjects pointed and the lesser the story was built in a cooperative manner. When the subjects are centred on the cubes manipulation task, they couldn’t manage an efficient narrative communication. The story was build by successive elements without real integration of the distant user.

3 VIRSTORY

From previous human studies, we improved VIRSTORY. We included two news modules: one hand, speech and visual gesture recognition and the other hand, a behavior module that is under way in our laboratory (cf. Figure 4). The gesture and speech input modalities are interpreted by MOWGLI (Multimodal Oral With Gesture Large display Interface) module [3] and used like synchronous or not synchronous interface commands in order to interpret and sometimes disambiguate the user's commands. The choice about sequential gesture or speech commands is a result from the WoZ experimentation. In fact, the user points the cube on the screen and after, he uses the oral command (e.g.: “turn”) to turn the cube and to see the other faces. The Animation module contains a library of non-verbal behavior like face expressions, postures, animations bodies and some inverse kinematics algorithms.
The first step about Behavior module concerns the relations between the interpreted

data (head and arms position, speech detection, speaking time, words recognition, etc.) from MOWGLI module and the running of appropriate behaviors by the Animation module on a 3D articulated character through an agent based system.

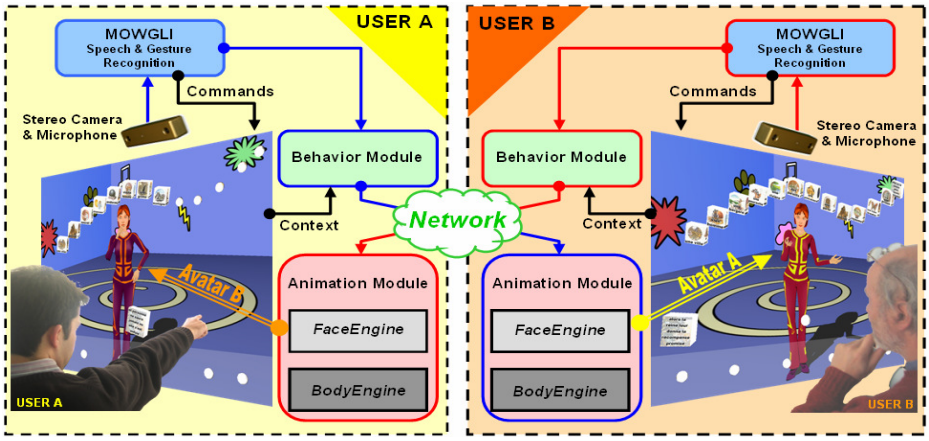

Fig. 4. The new VIRSTORY architecture: User A (at left) see user B's character and vice versa

4 Conclusion

To conclude, the analysis that we carry out here points out the necessity to work on human studies particularly in situ or in ecological conditions, in order to discriminate transpositions from the real world into the virtual world. Our objectives are also to explore step by step: the multimodal interaction in relation to character animation and strategies of collaboration, the improvement of behavioral module according to the creative task.

References

1. Buisine, S., Martin, J.C., Bernsen, N.O.: Children's gesture and speech in Conversation with 3D Characters. International Conference of Human Computer Interaction, Las Vegas (2005)
2. Vilhjalmsson, H.: Animating Conversation in Online Games. Entertainment Computing, Lecture Notes in Computer Science, ICEC 2004, (2004) 139-150
3. Carbini, S., Delphin-Poulat, L., Perron, L., Viallet, J.E.: From a wizard of Oz experiment to a real time speech and gesture multimodal interface. Signal Processing (2006)
4. Perron, L.: An avatar with your own gestures. INTERACT'05 Workshop "Appropriate Methodology for Empirical Studies of Privacy", Communicating Naturally through Computers, Rome (2005)