Employing Virtual Humans for Education and Training in X3D/VRML Worlds

Lucio Ieronutti
HCI Lab
Dept. of Math and Computer Science,
University of Udine
via delle Scienze, 206

ieronutt@dimi.uniud.it

ABSTRACT

Advances in computer graphics, improvements in hardware performance, and network technologies have enabled a new class of interactive applications involving *virtual humans*, three-dimensional simulations of human beings. Virtual humans are more and more used in multimedia learning environments, e.g. to explain maintenance operations, to train medical staff, or can be employed as virtual teachers. Nevertheless, in Web3D sites virtual humans are in practice rarely used, since they are complex to implement and their proper development requires knowledge of several disciplines (e.g., biomechanics, kinematics, artificial intelligence, computer graphics,…). Moreover, the programming effort to develop and re-use the virtual human into different learning environments can be considerable. In this paper, we propose a general architecture that allows Web3D content creators to easily integrate virtual humans into learning environments. The proposed solution can be used independently from the specific learning application domain (e.g. from a technical presentation to an history lesson). To test the applicability and effectiveness of our approach, we have applied it in a virtual museum of computer science.

Categories and Subject Descriptors

I.3.6 [**Computer Graphics**]: Methodology and Techniques – *interaction techniques.* I.3.7 [**Computer Graphics**]: Three dimensional Graphics and Realism – *Virtual Reality.* H.5.1 [**Information Interfaces and Presentation**]: Multimedia Information System – *Artificial, augmented, and virtual realities.*

General Terms

Design, Experimentation, Human Factors.

Keywords

Virtual Environment, Learning Environment, Virtual Humans, Embodied Agents, H-Anim.

1. INTRODUCTION

Virtual humans, i.e. three-dimensional simulations of human beings, are more and more used for different applications in multimedia learning environments: i) they are used to explain physical and procedural human tasks (e.g. maintenance operations) [1][13] by allowing users to be given less theoretical explanations which are more intuitive and engaging; ii) they are employed in medicine, from training applications designed to train civilian officers to recognize and interact with mentally ill people [8], to applications that range from simulations of emergencies in first aid [3] to simulations of surgical operations [4]; iii) virtual humans are used in military applications to train servicemen [15], by allowing one to reproduce very dangerous situations; iv) they are used as virtual personal teachers that complement their explanation by referring to virtual objects and places, in order to simplify the understanding of the lesson. The virtual teacher can lead users through the environment [5] and present different topics by following a logical order.

In general, virtual humans employ locomotion abilities and perform actions (such as deictic gestures) to focus students' attention on most important aspects and interact with virtual objects, while they can use facial expressions [14] in order to make the communication with users more realistic, effective and engaging. Virtual humans can communicate with users by exploiting both verbal and nonverbal communication; a virtual human can use gaze and gestures to focus the student's attention [11][12][13] during its explanation. For example, it can use gaze to regulate turn-taking in a mixed-initiative dialogue [3], while head nods and facial expressions can provide unobtrusive feedback on the student's utterances and actions without disrupting the student's train. Moreover, the presence of an anthropomorphic agent may increase the student's arousal and motivation [10]. From a human-computer interaction point of view, the interaction with virtual humans is based on metaphors consistent with the real world experience of users and then it suggests the possibility of more natural ways of communication.

There is a large amount of research in the field of virtual humans, and a large number of different approaches have been proposed in literature; each of them varies in appearance, function and autonomy according to the application field and the required detail and accuracy. Nevertheless, virtual humans are rarely used in Web3D sites, since they are complex to implement and there are no tools for supporting the development of virtual humans with complex behaviors in these Web sites. As a result, since the Web3D content creator has to implement virtual humans mostly by hand, and the programming effort to develop and re-use the virtual human into different learning environments can be considerable.

In this paper, we propose a general architecture for H-Anim characters [7], called Virtual Human Architecture (VHA), that allows Web3D content creators to easily employ virtual humans in Web-based applications for learning, education and training. To test benefits of our approach, we use it in a 3D Web site representing a Computer Science museum in which the virtual human leads users through the environment, it invites students to interact with devices and it provides technical explanations regarding the functioning of different devices.

This paper is structured as follows. In Section 2, we present problems related to the development of virtual humans and discuss the lack of supports for implementing virtual humans in Web3D sites. In Section 3, we present the Virtual Human Architecture, a general architecture that allows content creators to develop virtual humans in Web3D sites without worrying about low-level implementation aspects. In Section 4, we show how the proposed architecture can be used in a multimedia learning environment by providing a practical usage example. Finally, in Section 5, we discuss current limitations of the proposed approach and outline how we plan to overcome them.

2. MODELING VIRTUAL HUMANS

The development of virtual humans requires to acquire the knowledge of different disciplines, such as computational geometry, kinematics, artificial intelligence, computer graphics, and bio-mechanics. The complexity of building an embodied agent requires to subdivide the problem; this can be done in a hierarchical way [6], as shown in Figure 1.

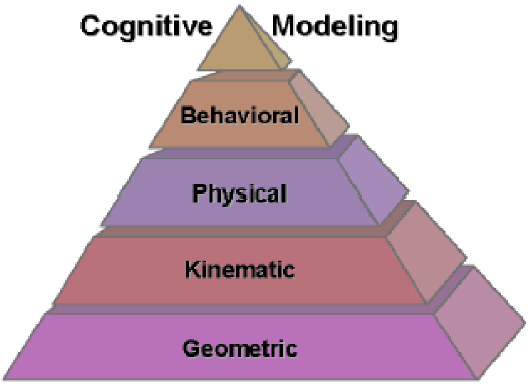


Figure 1– The modeling hierarchy proposed by [6]

At the base of the pyramid there is the *geometric* layer that concerns the definitions of the virtual human model and its appearance. In the *kinematic layer*, the virtual human is represented as a set of rigid bodies, called *segments*, organized hierarchically and connected by *joints*. From this point of view, an animation can be defined in two ways: i) by specifying joints rotations, or ii) by defining (or automatically computing) positions of segments extremities in time. The latter approach uses inverse kinematics, technique that allows one to compute the joints configuration (in terms of rotation values) needed to reach the specified position; this technique is especially suitable to control hands and feet movements. In the *physical layer* the animation is obtained by applying physical laws to different parts of the body; this method is used to compute complex animations, such as skin deformations or the hairs movement. The *behavioral layer* represents the instinctive behavior of the virtual human (e.g. in terms of stimulus-action associations). *The cognitive modeling* instead binds various stimuli with reasoning process that allow the virtual human to search for the most suitable reaction; cognitive models go beyond behavioral models in that they govern what the virtual human knows, how that knowledge is acquired, and how it can be used to plan actions.

2.1 Modeling Virtual Humans in X3D/VRML

Although to implement virtual humans on the Web different technologies and approaches can be employed (see [9] for an overview), we focus our attention on H-Anim [7], the standard for implementing humanoids by using X3D/VRML technologies.

H-Anim defines the virtual human body as a set of *segments* organized hierarchically and connected by *joints*; each joint is defined by its position and its own rotation value. A humanoid animation is defined in X3D/VRML by specifying different rotation values of joints in time; the resulting motion is generated by smoothly interpolating specified rotation values. Moreover, since H-Anim defines the name of different joints, it allows one to apply the same animation to different virtual humans. This kind of animations are called *pre-stored animations*, since the complete description of the movement is specified in advance. Actually H-Anim standard supports only this kind of animations.

To animate a virtual human it is more convenient to use another kind of animation, called *parametrized animations*, that use a small set of parameters to generate at execution-time an animation as a function of these parameters. Parametrized animations are more general and flexible than pre-stored ones, since they can generate a variety of movements by changing animation parameters. Usually parametrized animations use inverse kinematics to control end-effectors movements (e.g. feet and hands) and employ path planning algorithms to generate collision-free motions. A typical example of parametrized animation is the walking motion; by starting from an high-level description of the movement (e.g. specified the initial and final humanoid position and by defining the set of parameters that characterizes the movement, such as the length of a single step), the corresponding animation is generated by using a path planning algorithm to compute a collision-free trajectory, and by employing inverse kinematics to derive a legs movement that avoids compenetration with the walking surface. Unfortunately H-Anim does not support parametrized animations.

Moreover, H-Anim does not specify the way for describing the high-level behavior of the virtual human. As a result, the implementation of an animated virtual human in learning and training environments is a trial-and-error, time-consuming activity for the Web3D content creator.

3. The proposed VIRTUAL HUMAN ARCHITECTURE (VHA)

In this section we briefly describe the high-level architecture we propose (illustrated in Figure 1). The main internal module of the architecture are i) the *Behavioral Engine* and ii) the *Animation Engine*. It is important to note that the proposed architecture is a good compromise between the required realism (or at least believability) of the representation (at each level of the modeling hierarchy [6]) and the efficiency of the simulation (the simulation has to be carried out in real-time on common home computers).

The Behavioral Engine, by sensing user's interactions and depending on the defined behavior of the virtual human, identifies both the information that has to be provided to the user and animations that the virtual human has to perform.

Information are taken from a database and are at the same time both displayed on a semi-transparent On Screen Display (OSD for short, see Figure 3) and presented by using a synthesized voice (e.g. using the Microsoft Text-to-Speech engine).

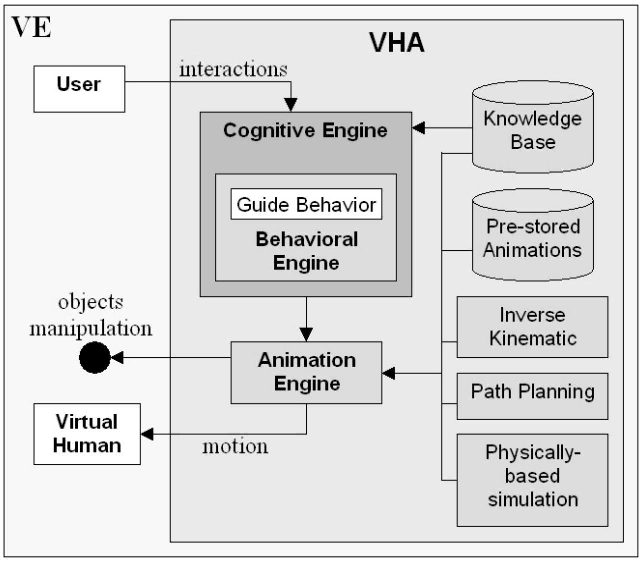
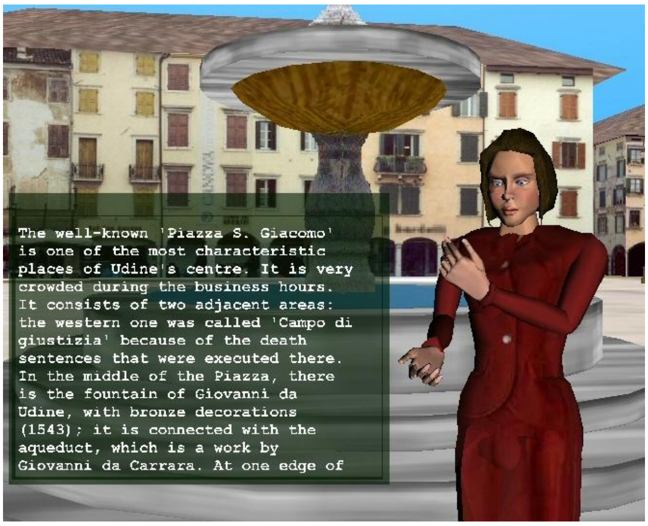


Figure 2 – The Virtual Human Architecture (VHA)

By using the inverse kinematics, path planning algorithms, pre-store animations and by exploiting objects information stored into a database (e.g. their location and orientation into the virtual environment), the Animation Engine generates at run-time the required animation.

In the following we describe the two modules in more detail.



The well-known 'Piazza S. Giacomo' is one of the most characteristic places of Udine's centre. It is very crowded during the business hours. It consists of two adjacent areas: the western one was called 'Campo di giustizia' because of the death sentences that were executed there. In the middle of the Piazza, there is the fountain of Giovanni da Udine, with bronze decorations (1543); it is connected with the aqueduct, which is a work by Giovanni da Carrara. At one edge of

Figure 3 – The required presentation text is read by a text-to-speech engine and displayed on a OSD.

3.1 Animation Engine

The Animation engine allows the virtual human to perform parametrized animations such as walking, stairs-climbing, grasping, and the ability to make facial expressions.

The walking animation is computed by using inverse kinematic techniques; this approach allows the VHA to generate realistic and flexible walking animations. Given the position of footprints on the desired path, VHA automatically creates the walking animation by computing trajectories of the humanoid pelvis and feet. Since VHA takes into account both the virtual human model and the morphology of the walking surface to generate the animation, the Animation Engine is able also to generate stair-climbing animations only by specifying the walking surface parameters.

VHA adopts inverse kinematic techniques also to generate grasping animations; these animations can be used for example in virtual training environments to demonstrate how to repair a broken engine or to explain how to carry out maintenance procedures.

Moreover, VHA supports expressive virtual humans (humanoids that are able to display facial expressions); from a Human-Computer Interaction point of view, interacting with an expressive virtual human makes the conversional more believable, realistic and engaging for users. Among different techniques proposed in literature for this purpose, the proposed architecture uses the physics-based one, i.e. technique that controls facial expressions by acting on muscles contraction and computing the resulting face expression according to physic laws (e.g. using the mass-spring-model).

3.2 Behavioral Engine

Models of virtual human behavior are integrated into the architecture; these models specify how the virtual human responds to user input in terms of actions performed and information presented. Each model of the virtual human behavior is represented by a Finite-State Machine (FSM for short); this structure specifies how the virtual human acts given the current state and considered student interactions.

Each FSM is represented with the FSM G = (V, E), where V are the set of nodes n_i , while E are the set of oriented edges (n_i, n_j). Each node corresponds to a particular state of the virtual human, while each edge corresponds to a transition that allows the virtual human to change its internal state. Each transition (n_i, n_j) is characterized by the couple conditions-actions (c_{ij}, a_{ij}): c_{ij} is the set of conditions that determine the applicability of the transition, while a_{ij} is the set of actions that the virtual human performs if the corresponding transition is activated by the Behavioral Engine. A transition can be activated if and only if corresponding conditions are satisfied.

Given a current state of the virtual human, the Behavioral Engine senses user's interactions and determines what conditions are satisfied, identifies applicable transitions, activates one of them and returns the set of actions associated to the chosen transition.

Each FSM represents a different behavior and, in a didactical application, can correspond to the structure of an interactive lesson. From this point of view, nodes can represent different concepts (or group of concepts), while edges can represent relations between concepts. This way, given the concept c_k (that is associated to the state n_k), all nodes belonging paths connecting the initial state with n_k can be considered concepts necessary for the understanding of c_k. As a result, the user can actively participate to the lesson by influencing the order in which different concepts are presented, since the user behavior (e.g. its interaction with objects and the virtual human) determines the way in which the graph is explored.

4. An example of a learning environment: a 3D COMPUTER SCIENCE MUSEUM

The proposed architecture can be used in different learning environments independently from the application domain. To test the effectiveness of our solution, we considered two different case studies. First, we employed a virtual human into a learning environment aimed to explain the functioning of computer devices of the 70's. Second, we used the same virtual human as a guide into an architectural virtual reconstruction of cultural heritage. Although the considered Web3D sites differ in purposes and contents, only few modifications have been required to shift from the first to the second application. While in the first learning environment the virtual human provides technical information by demonstrating how different devices worked, in the second application the same virtual human is used for the promotion of cultural heritage, since its main function is to tell the history of different buildings by highlighting main architectural differences.

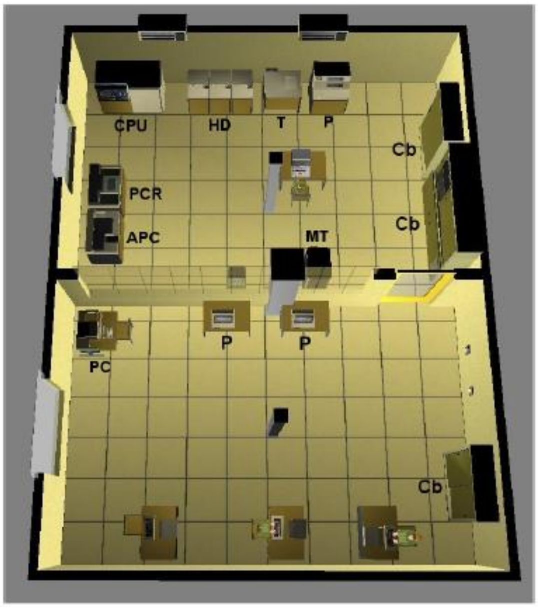


Figure 4 – Top view of 3D Computer Science Museum, with the different devices on display

Due to the limited available space, in this paper we focus only on the first mentioned learning environment. A screenshot of the second learning environment is provided in Figure 3.

The 3D Computer Science museum is based on the virtual reconstruction (developed using VRML and Java) of a typical data processing centre of the '70s, reproducing hardware from the Univac/Sperry 90/30 line. The main pedagogical goals for this virtual museum are concerned with pointing out the large differences between data processing centers in the '70s and current computers, e.g. by illustrating the mainframe – terminals architecture and the interaction based on text video terminals or (more often) punch cards.

The virtual data processing center is divided into two main rooms (as shown in Figure 4): a computer room, containing the main system, devoted to data processing under the control of technical staff, and a terminal room, containing punch card units and video terminals, devoted to activities that are preparatory to real data processing. Museum visitors have the possibility to: i) observe the different devices in their original context of operation; ii) obtain information on the devices, by clicking on them and reading and/or listening to a description of their features and functioning; iii) manipulate the devices to observe their internal parts. For example, the user can open cabinet units to examine their internal details and working (Figure 5 illustrates the case of hard disks).

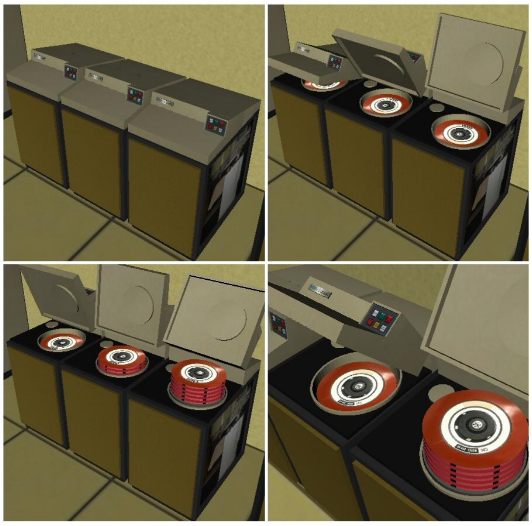
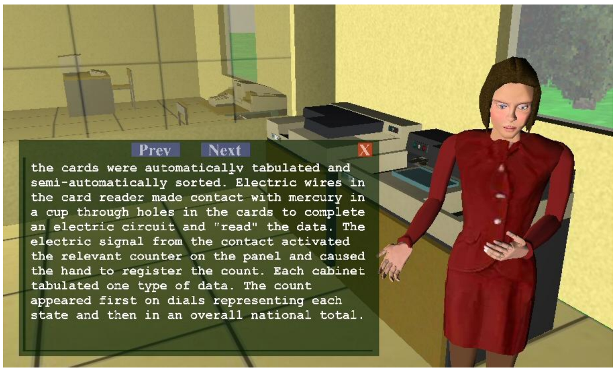


Figure 5 - Interactive components: hard disks

The virtual human leads student through the environment, it presents and describes different devices by following a logic order and by considering what information has been already presented to the user during the visit. If needed, the virtual human provides additional information (see Figure 6), provides comparisons between different devices (e.g. punch card and card reader), by organizing logically the lesson, e.g. the virtual human explains what memory devices were used in the 70's before providing the detailed punch card description.

To increase the realism of the user experience, we added the necessary furniture and included typical working people. Moreover, the audio channel is used to add typical noise and sounds of objects and human actions (for instance, printers, operator's typing, etc).



the cards were automatically tabulated and semi-automatically sorted. Electric wires in the card reader made contact with mercury in a cup through holes in the cards to complete an electric circuit and "read" the data. The electric signal from the contact activated the relevant counter on the panel and caused the hand to register the count. Each cabinet tabulated one type of data. The count appeared first on dials representing each state and then in an overall national total.

Figure 6– The virtual human while it is explaining the functioning of the card punch

We define different FSM that correspond to different structured information, each one designed to highlight a different aspect of the Computer Science Museum (e.g. an high-level introduction to the overall environment, an explanation of the hardware architecture, a description of work activities). The user can choose the lesson he intends to follow before the visit.

5. CONCLUSION AND FUTURE WORK

In this paper, we have presented a general architecture to develop virtual humans in Web3D sites. In particular, the architecture is used for learning and training applications in which virtual humans are used as virtual teacher and assistants. The proposed solution allows Web3D content creators to implement learning environments in different application domains. We have also presented a practical example that shows how virtual humans can be used to teach technical topics by using a 3D Computer Science Museum as a case of study.

We plan to extend the proposed approach by considering the user model in order to take into account needs, preferences and the individual knowledge of users. From this point of view, we plan to integrate our architecture with a tutoring system that dynamically generates Web content: this way, it is possible to provide users with personalized information, allowing the virtual human to present tailored lessons to different students. Users' information could be acquired by using an initial form that asks for typical information, exploiting a stereotypical knowledge, and by dynamically updating the student's profile by considering her interaction both with the virtual human and the environment.

Moreover, we intend to make the user-virtual human interaction more natural and intuitive; to achieve this goal, we plan to integrate into the proposed architecture a speech recognition engine; this solution allows users to communicate with virtual humans by using a well-known metaphor.

6. ACKNOWLEDGMENTS

This work is partially supported by the MIUR COFIN 2003 program and by the Friuli Venezia Giulia region (Regional Law 3/98, project "3D Web Sites for the Promotion of Tourism and Cultural Heritage").

7. REFERENCES

[1] Badler, N., Erignac, C., Liu, Y. Virtual humans for validating maintenance procedures. *Communications of the ACM*, 2002, 45(7), pp. 56-63.

[2] Badler, N., Webber, B., Clarke, J., Chi, D., Hollick, M., Foster, N., Kokkevis, E., Ogunyemi, O., Metaxas, D., Kaye, J., and Bindiganavale, R. MediSim: simulated medical corpsmen and casualties for medical forces planning and training. In *Proceedings of the National Forum: Military Telemedicine On-Line Today Research, Practice, and Opportunities*, IEEE Computer Society Press, 1996, pp. 21-28.

[3] Cassell, J., Pelachaud, C., Badler, N., Steedman, M., Achorn, B., Becket, T., Douville, B., Prevost, S., and Stone, M. Animated conversation: Rule-based generation of facial expression, gesture and spoken intonation for multiple conversational agents. In *Proceedings of SIGGRAPH '94*, ACM Press, New York, 1994, pp. 413-420.

[4] Cavazza, M., and Simo, A. A virtual patient based on qualitative simulation. In *Proceedings of the 8th international conference on Intelligent user interfaces*, ACM Press, New York, 2003, pp. 19-25.

[5] Chittaro L., Ranon R., and Ieronutti L. Guiding Visitors of Web3D Worlds through Automatically Generated Tours. In *Proceedings of Web3D 2003: 8th International Conference on 3D Web Technology*, ACM Press, New York, 2003, pp. 27-38.

[6] Funge, J., Tu, X., and Terzopoulos, D. Cognitive Modeling: Knowledge, Reasoning, and Planning for Intelligent Characters. In *Proceedings of SIGGRAPH '99*, ACM Press, New York, 1999, pp. 29-38.

[7] H-Anim web site: http://www.h-anim.org/.

[8] Hubal, R. C., Frank, G. A., and Guinn, C. I. Lessons Learned in Modeling Schizophrenic and Depressed Responsive Virtual Humans for Training. In *Proceedings of the Intelligent User Interface Conference*, ACM Press, New York, 2003, pp. 85-92.

[9] Lau, R.W.H., Li, F., Kunii, T.L., Guo, B., Zhang, B., Magnenat-Thalmann, N., Gutierrez, M. *Emerging Web Graphics Standards and Technologies*, IEEE Computer Graphics and Applications, 2003, pp. 66-75.

[10] Lester, J. C.,Converse, S. A., Kahler, S. E., Barlow, S. T., Stone, B. A., and Bhogal, R. S. The persona effect: Affective impact of animated pedagogical agents. In *Proceedings of CHI '97*, ACM Press, New York, 1997, pp. 359-366.

[11] Lester, J. C., Voerman, J. L., Towns, S. G., and Callaway, C. B. Deictic believability: Coordinating gesture, locomotion, and speech in lifelike pedagogical agents. Applied Artificial Intelligence, 1999, 13(4-5), pp. 383-414.

[12] Noma, T., and Badler, N. I. A virtual human presenter. In *Proceedings of the IJCAI Workshop on Animated Interface Agents: Making Them Intelligent*, 1997, pp. 45-51.

[13] Rickel, J., and Johnson, W. L. Integrating pedagogical capabilities in a virtual environment agent. In *Proceedings of the First International Conference on Autonomous Agents*, ACM Press, New York, 1999, pp. 30-38.

[14] Terzopoulos, D., and Waters, K. Physically Based Facial Modelling, Analysis and Animation, Journal of visualization and Computer Animation, 1990, 1(2), pp. 73-90.

[15] Traum, D., and Rickel, J. Embodied agents for multi-party dialogue in immersive virtual worlds. In *Proceedings of the first international joint conference on Autonomous agents and multiagent systems*, ACM Press, New York, 2002, pp. 766-773.