Supporting Presentation Techniques based on Virtual Humans in Educational Virtual Worlds

Luca Chittaro Lucio Ieronutti Sara Rigutti
HCI Lab, Dept. of Math and Computer Science, University of Udine,
Via delle Scienze 206, 33100, Udine, Italy
{chittaro, ieronutt, rigutti}@dimi.uniud.it
http://hcilab.uniud.it

Abstract

Educational Virtual Worlds (EVWs) allow one to circumvent physical, safety, and cost constraints that often affect real-world training and learning scenarios. Virtual humans can improve the effectiveness of an EVW e.g., by explaining and demonstrating to the user procedural tasks and communicating with learners in a natural way by exploiting both verbal and nonverbal communication. However, choosing how the learning material has to be presented to the learner and how to employ the virtual human is influenced by different factors, ranging from available technology (e.g., bandwidth, graphics hardware, installed software) to learner's profile (e.g., physical and/or cognitive disabilities, spatial abilities, 2D/3D navigation expertise). The system we present in this paper flexibly supports different presentation techniques employing a virtual human into an EVW. The different capabilities of the system are illustrated, motivated and discussed.

1. Introduction and motivation

Teaching is a social process involving direct teacher/learner communication. Web-based technology is changing this traditional face-to-face relation since it allows one to benefit from educational materials without needing the presence of both the teacher and the learner within the same physical environment (e.g., a classroom) at the same time. Educational Virtual Worlds (EVWs) enrich the learning experience with features such as 3D simulations and/or virtual humans acting as teachers or pedagogical agents.

Constructivism theory [1][2] provides a motivation to EVWs, since it claims that a direct experience of the world activates a knowledge construction process that takes place when learners are intellectually engaged in "personally meaningful tasks" [3]. A similar consideration is the basis for the *learning by doing* principle that emerged within constructivism [4]: knowledge acquisition is improved by learner's active involvement in tasks that use that knowledge. From this point of view, the possibility of providing highly interactive experiences becomes one of the best-valued features of EVWs. As pointed out by Harper et al. [5], besides reality, the most appropriate way to generate a context based on authentic learner's activity may be through the use of EVWs.

Social agency theory [6] provides a motivation to the application of virtual humans to support learning since it claims that learning can be improved when learners live their relation with the computer as a social interaction [6][7]. The embedding of social cues within an EVW encourages learners to interpret human-computer interaction as more similar to human-human interaction. Any primed social interaction schema enhances the learners' motivation in understanding and drives learners to more deeply process the computer message. The processing of information with deep levels of understanding (*sense making process*) affects learner's ability in generalizing what she has learned to new situations (*transfer*). Virtual humans "life-like" features such as linguistic expressions, speech styles, accent, voice, gesture and facial expressions have been considered as factors affecting both interaction and learning [7][8][9][10][11].

Social learning theory [12] provides more motivations to the exploitation of virtual humans behaving as teachers, since it claims that individuals can learn by observing other individuals acting. Knowledge may be significantly modified by viewing the behaviors of a human model. Therefore, a virtual human should be very effective within educational contexts involving *learning by observation* (learning should be facilitated by the mere observation of the virtual human that performs tasks).

Turning the above theoretical considerations into real EVWs requires the availability of proper software tools that make it easier to employ virtual humans. This paper describes PETS (Presentation of Education and Training Subjects), a system we have developed to support different presentation techniques employing a virtual human into EVWs. In Section 2, we discuss related work on virtual humans for educational purposes. In Section 3, we illustrate PETS and the presentation techniques it supports. Finally, in Section 4, ongoing and future work is discussed.

2. Related work

In the context of EVWs, virtual humans are used for different purposes, such as explaining physical and procedural human tasks [13] and simulating dangerous situations [14]. Virtual humans can also be the main subject of study e.g., for training civilian officers to recognize and interact with mentally ill people [15] and in applications for medical staff training [16][17][18].

In this section, we first present an overview of some major systems, architectures and applications employing virtual humans in EVWs, and then we present some empirical proofs of the effectiveness of virtual humans in improving the learning process.

2.1. Systems, architectures and applications

JUST-TALK [15] is a system designed to teach policemen basic techniques for managing encounters with mentally ill people through a series of one-to-one scenarios with a virtual human. Through the dialog with the virtual human, the learner has to decide whether to release or detain it. Virtual Medical Trainer (VMET) [16] is a system that combines medical procedures databases and virtual patients to produce an interactive environment that can mimic the cognitive pre-hospital assessment and care demands of a real emergency. In the same educational context, another interesting system has been presented by Cavazza and Simo [17]; the application simulates the cardio-vascular system to generate clinical syndromes and derives the consequences of the trainee's therapeutic interventions. The system employs the virtual patient's appearance as the most direct manifestation of the seriousness of the situation. VHD++ [19] is a real-time software framework for developing virtual and augmented reality applications employing advanced virtual character simulation. In the context of EVWs, VHD++ has been used for example to develop JUST [18], an health emergency decision training system. An example of how virtual humans can be used in educational contexts has been provided by Sims [20], who developed a system to teach aspects of the Iraqi dialect and culture (such as gestures and customs). Virtual humans can also be used as virtual companions within a collaborative EVW, asking questions if the user is too passive or performing tasks if the user does not feel ready to do something as well as making errors. For example, STEVE [13] is a virtual human that helps students in learning how to perform physical and procedural tasks. In particular, STEVE is able to demonstrate how to perform tasks and can also monitor students while they practice tasks, providing assistance when needed. Besides teaching students individual tasks, the system can also help them in learning how to perform multi-person team tasks: STEVE can either behave as a tutor for a student learning a particular role in the team, or play the role of a teammate when a human teammate is unavailable.

2.2. Empirical proofs

Virtual human effectiveness has been empirically shown by means of either direct or indirect measures [21], where the former refers to different types of preference ratings, while the latter refers to different indicators, such as knowledge transfer and recall. Experiments made by different researchers (e.g., [10][22][23][24][25]) support the idea that virtual humans enhance several factors affecting the learning process, such as: (i) *motivation*, supporting cognitive functions such as problem solving, understanding and learning [22], (ii) *learnability*, facilitating interaction with the EVW by enabling people to apply social heuristics [23], (iii) *believability*, supporting the perceived plausibility of information provided by the system with expressiveness [10], (iv) *attention*, capturing attention with facial display [24], and (v) *communication efficiency*, providing non-verbal cues on the state of the dialogue [25]. Lester et al. [26] state that the simple "presence" of a virtual agent in an EVW improves learner's perception of her educational experience (*persona effect*). Starting from this consideration, several intrinsic factors of virtual humans (i.e., voice, gestures and facial expressions) have been studied. In particular, empirical research has been focused mainly on different presentation techniques employed by virtual humans, e.g., comparing the effectiveness of presenting information to the learner in a textual format or using narrations [6][7][27]. Mayer et al. [6] found that a narrating agent had a positive effect on learner's recall, transfer and interest. Similarly, Atkinson [27] measured a better transfer performance when learners were instructed by a talking virtual human than when they were not (voice only or text only). Consistently with social agency theory, voice is interpreted as a social presence indicator improving the learning process. The relevance of voice is confirmed by other results: a virtual human answering by narration is more effective than a virtual human answering by printing words [7]. Following the theory of *multimedia learning* [28], the visual information processing channel may be overloaded when learners have to process on-screen graphics and on-screen text at the same time. On the other hand, a reduction of the cognitive load on the visual channel is achieved by presenting words in narrated form. Recent research showed that learning is affected both by the type of virtual human's voice [7] (with a preference for human recorded voice over synthesized voice) and by the type of virtual agent's speech [8] (with a preference for personalized speech over non-personalized one e.g., third person). Empirical research focused also on non-verbal communication as a crucial aspect in virtual human-learner dialogue. Gaze, gesture, intonation and body posture have conversational functions, e.g., indicating conversation start and ending, turn-talking, feedback and error correction [29]. Results by [25] and [11] suggest that non-verbal behaviors are more important during interaction than emotional feedback (e.g., smile and puzzlement) in supporting information exchange between the learner and virtual human.

3. The proposed system: PETS

The empirical proofs reported in previous sections suggest a positive effect of EVWs and virtual humans on learning. However, the effectiveness of a lesson setting could be affected not only by virtual human *intrinsic* factors (as those reported), but also by *extrinsic* ones (e.g., hardware/software available to the learner, learner's experience with navigation, type of educational materials). In particular, the available technology (e.g., bandwidth, graphics hardware, installed software), as well as individual learner's features (e.g., physical and/or cognitive disabilities, spatial ability, 2D/3D navigation expertise) can strongly affect the effectiveness of an EVW.

With the PETS (Presentation of Education and Training Subjects) system, we aim at providing a Web-based solution supporting different presentation techniques (including the different options mentioned in section 2.2) for integrating educational materials into EVWs. The courseware designer provides PETS with the available resources (e.g., text, pictures, videos and 3D models), and the system is able to provide the learner with a set of alternative EVWs. The learner (or possibly her instructor) can choose the EVW that best fits her preferences and needs. Moreover, PETS includes a monitoring system that allows one to acquire and store information on learner's actions; this functionality enables the courseware designer and instructors to analyze collected data e.g., for monitoring the students' learning process. In the following, we first briefly describe the high-level architecture of PETS and then we illustrate, motivate and discuss presentation techniques supported by our system.

3.1. Architecture

The system is mainly composed by four modules (see Figure 1): the *Data Acquisition* module and the *Virtual Human Architecture* (VHA) that run on client side, the *Recorder* and the *Loader* modules that run on server side. Moreover, a database is used for learning materials and another database for usage data.

Data Acquisition acquires information on user's actions and sends it to the Recorder. The Recorder stores this data into the *Usage Data* database, together with proper identifiers for the considered user and session. More specifically, Data Acquisition acquires information on user's navigational behavior and interactions with the EVW. Navigational information is obtained by sampling the position and orientation of the learner into the EVW at short time intervals. Data Acquisition records any other user's action in terms of type of action (e.g., selection or manipulation), element involved in the action (e.g., link or 3D object) and timing information.

Recorded data can be later analyzed, e.g., to (i) evaluate how the student accessed the learning material (e.g., following a logic order), (ii) determine if the learner accomplished a procedural task successfully, or (iii) measure how much time the user spent in different parts of the EVW. Data analysis applications (e.g., spreadsheets and statistical tools) can be employed for studying actions data, while navigational data can be analyzed by using tools such as VU-Flow [30][31] that highlight navigability problems and interests or preferences of learners.

Data acquired by Data Acquisition is also sent to VHA [32], a Web-based software architecture that allows one to control H-Anim [34] characters in virtual worlds handling geometrical, kinematic, physical and behavioral aspects. More specifically, VHA enables the virtual human both to autonomously navigate through the environment avoiding collisions with obstacles (e.g., walls) and to interact with objects of the virtual world. The VHA allows one to define the virtual human behavior through scripting. In particular, behavior has to be defined in terms of *stimulus-action* associations: in response to user's actions (stimulus), VHA determines the appropriate (re)action, the associated animations and information to be presented to the user. As a result, depending on user's actions (detected by Data Acquisition), additional (both 2D

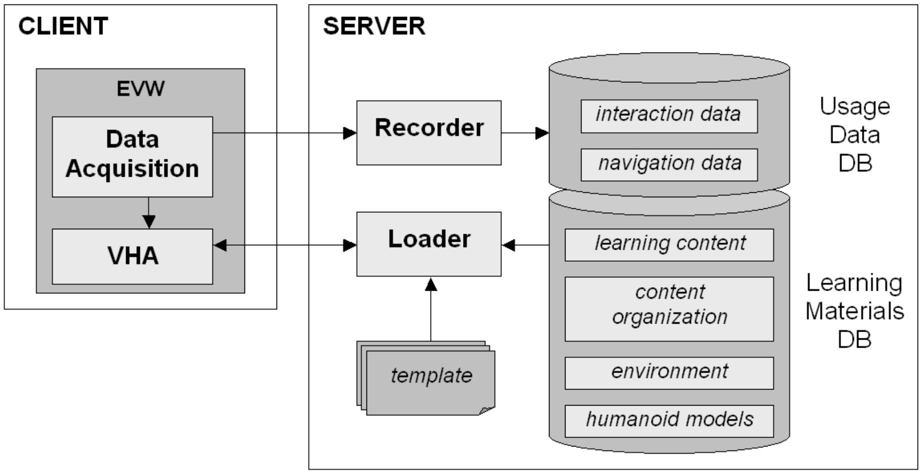


Figure 1. High-level architecture of PETS.

and 3D) elements and data (e.g., textual information or audio data) can be included into the presentation and specific virtual human behaviors activated. Technical aspects of VHA (e.g., details about path planning, collision detection and grasping) can be found in [32].

Elements and information needed for the presentation are stored into the *Learning Materials* database. At the beginning of a session, the Loader, based on available learning materials, provides the learner with a choice of alternative EVWs; the learner can choose what best fits her preferences and needs. Alternatively, an instructor can make this choice in advance for the learner. Then, the Loader uses pre-defined templates (e.g., specifying the Web page layout) to generate the EVW integrating required 2D and/or 3D elements. Moreover, during the learning session, the Loader is able to retrieve additional materials from the database and send them to VHA. In particular, elements that can be included into the presentation are:

- *content*. The learning content is represented by text, images, videos, sound and (interactive or static) 3D models. For example, images, videos and/or 3D models can be used to complement textual and/or speech audio explanations. Moreover, additional data (e.g., 2D and 3D coordinates) has to be specified for allowing the virtual human to interact and/or point toward a particular part of an image, video or 3D model.
- *content organization*. Data specifying the order in which the learning content has to be presented to the learner. In particular, the content is organized in a graph structure that defines what information has to be presented to the user, (objects and virtual human) animations to be executed, and elements (e.g., 3D objects) to be included into the presentation in response to a particular user's action (e.g., link selection or object manipulation).
- *environment*. Geometric model of the virtual world (e.g., walls, windows and doors) and information on the topology of the virtual world (e.g., areas that can be traveled by the virtual human, possibly derived using automatic methods such as [33]).
- *humanoid models*. Models of different virtual humans based on the H-Anim standard [34].

3.2. Composing presentation techniques

PETS supports multiple presentation techniques, according to the setting of the following variables: *type of visualization* (freely navigable 3D, 3D with constrained navigation, 2D), *presence of the virtual human* (present, absent), *type of explanation* provided to the learner (audio narration, text, both). In the following, we discuss each variable separately.

3.2.1. Type of visualization. In the first type of visualization, the learner freely navigates inside a 3D world by controlling the position and orientation of her viewpoint (typically through mouse and arrow-keys). Although complete control of the viewpoint is a desirable feature for users that are wed to 3D navigation, navigation in 3D worlds is often difficult for novice users and the effort required to control the viewpoint can affect the effectiveness of the learning experience (e.g., the user might spend considerable time in understanding where she is, how she can reach a specific place, how to obtain a desired viewpoint on 3D objects). In the second type of visualization, user's freedom of movement is reduced to minimize navigation effort. The onus of controlling the viewpoint is on the system that sets the position and

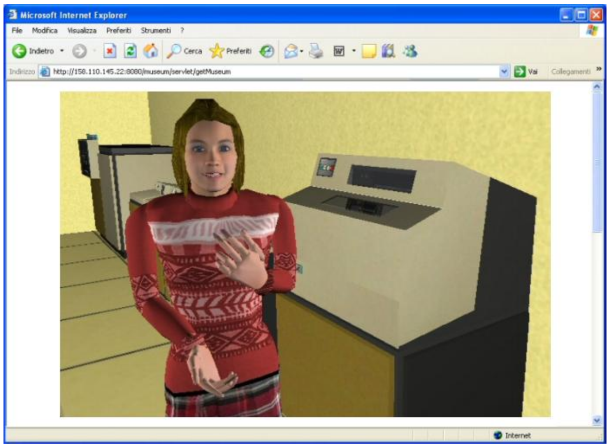


Figure 2. Virtual teacher inside 3D world.

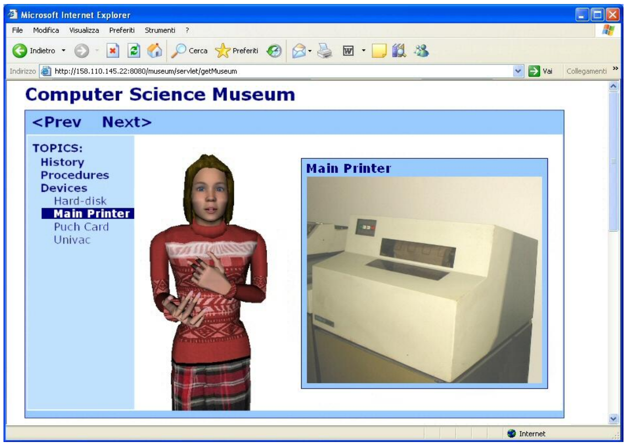
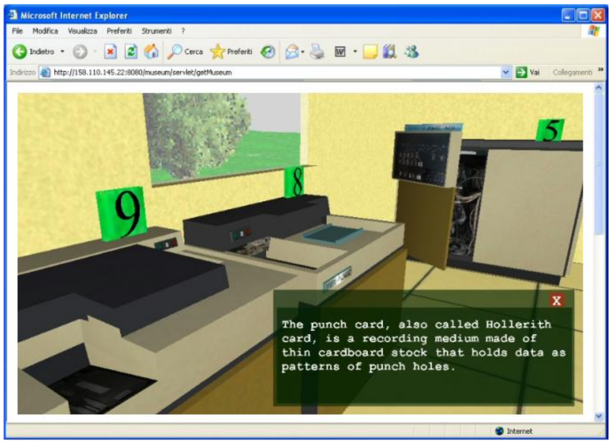


Figure 3. Virtual teacher explaining images.

orientation of the viewpoint to focus user's attention towards particular aspects of the virtual world and, at the same time, to limit occlusion problems (e.g., the virtual human stands between the user and the object it is presenting). However, it is important to note that (i) modeling 3D contents can be a time-consuming activity (often 3D models have to be built manually), and (ii) visualizing complex 3D models can require graphics hardware that could not be available to the learner. In the third type of visualization, content is presented to the learner through images and/or videos. The learner can access a specific topic by selecting the link concerning it as in a traditional Web site. However, this type of visualization has drawbacks e.g., interaction with the learning material becomes more limited. In the case of videos, a good bandwidth is also needed.

3.2.2. Presence of the virtual human. Regardless of the setting of the *type of visualization* variable, PETS is able to integrate (and control) the virtual human into the EVW. The virtual human provides both pedagogical and navigational assistance. In the two types of visualization employing a 3D world, the virtual human leads the learner through the world and explains the educational content (e.g., in Figure 2 a realistic humanoid guide is explaining the functioning of the main printer in a 3D reconstruction of a data processing center of the '70s), by providing the learner with both declarative (e.g., object description) and procedural (e.g., showing how to perform physical tasks) information. In the case of 3D with constrained navigation, the viewpoint position and orientation is automatically set to follow the virtual human during its movement through the virtual world. In the case of freely navigable 3D, the learner can decide to follow the virtual human or move freely through the world. In the case of 2D visualization, the virtual human explains the learning content by drawing (when needed) the user's attention to specific features of the image or video (e.g., by indicating a specific part of an object). With this presentation technique, the virtual human resembles an instructor who teaches by using a projector. If one chooses not to have the virtual human in the 3D world, the logical order in which contents should be presented to the learner is marked by numbered labels displayed upon each object of interest (see Figure 4). In particular, in the case of constrained navigation, PETS moves the viewpoint from one labeled object to the next one (following the defined sequence) while, in the case of free navigation, the learner can explore the virtual world and obtain information on objects by clicking on their labels. In both cases, object animations can be used for complementing textual/narrated explanations. In the case of 2D visualization, the learning content without the virtual human is presented similarly to a traditional Web site (HTML pages possibly including images and/or videos).

3.2.3. Type of explanation. Explanations can be provided to the learner through audio narration, text or both. The audio required for the narration can be either pre-recorded (from a human voice) or generated from text by a Text-to-Speech (TTS) engine. Audio recording is a time consuming activity and can be unpractical for long presentations. Moreover, the time required to download the audio files from the Web can be considerable. The solution based on a TTS engine allows one to face this problem, but requires the learner to download and install the engine. Moreover, the synthesized voice can be perceived by learners as unexpressive, unnatural and impersonal. If the text explanation is chosen in the two types of visualization employing a 3D world, textual data is displayed on a semi-transparent On Screen Display (OSD) in the



The punch card, also called Hollerith card, is a recording medium made of thin cardboard stock that holds data as patterns of punch holes.

Figure 4. Labeled objects and On-Screen-Display.

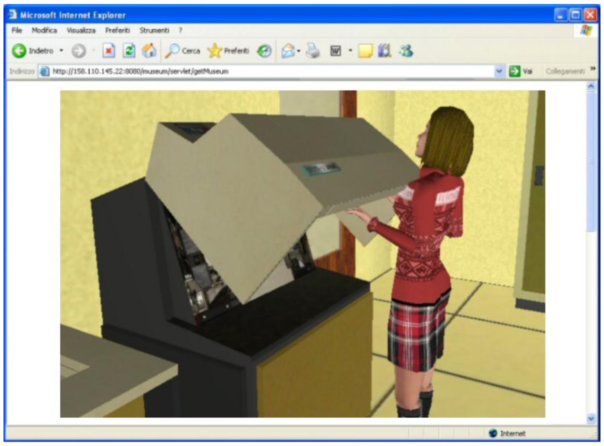


Figure 5. Virtual human interacting with objects.

virtual world (see Figure 4). Text in the OSD is automatically formatted according to display size and the learner can move and resize the OSD as needed. A study of nonverbal communication has been conducted on a group of four teachers to identify the gestures the virtual human should perform. To this purpose, we video-recorded short lessons given by the teachers based on materials provided by us to describe a processing center of the ‘70s. A post-hoc gesture analysis (co-occurrence) allowed us to identify *illustrators* [35][36] i.e. general movements that accompany and punctuate speech and accent in specific educational contexts.

3.3. Examples of presentation techniques

To test PETS, we have developed an educational application aimed at pointing out the several differences existing between data processing centers in the '70s (in particular, we have considered hardware from the Univac/Sperry 90/30 line) and current computers, and providing information about device features and functioning. In particular, lesson contents are presented to the learner by (i) following a logic order (e.g. explaining what memory devices were used in the 70s before providing the detailed punch card description), (ii) considering what information has been already presented to the learner during the visit, (iii) providing comparisons among different devices (e.g., punch card and card reader) and (iv) illustrating how to perform procedural tasks (e.g, how to extract the internal disk). For the virtual world, we have considered a typical structure of data processing centers of the ‘70s, organized into two rooms: a computer room and a terminals room. The environment contains devices such as terminals, printers, disk drives, computers, and the needed generic furniture (e.g., desks and chairs). In this section, we briefly show the effects of three different settings of the variables described in section 3.2; each setting produces a different presentation of the educational materials.

The first setting consists in freely navigable 3D visualization, presence of a realistic human model and pre-recorded audio for the explanation. With this setting, the learner has the possibility to look at the different devices in their original context of operation and manipulate them to examine their internal parts. The virtual human leads the learner through the virtual world, it presents different devices, and it provides the learner with procedural information by interacting with different devices (e.g., Figure 5 shows an interaction with the main printer). The pre-recorded audio for the virtual human employs an informal language.

The second setting consists in 2D visualization, presence of virtual human, and synthetic voice for the explanation. This setting will run more easily on low bandwidth connections with respect to the previous one. The devices of the ‘70s are now shown through pictures (e.g., the digital photo of the main printer in Figure 3). Since this does not allow the virtual human to interact with devices, procedural information is given to the learner with sequences of images automatically taken during the virtual human-object interaction (of the previously described presentation technique) from given viewpoints. Although the virtual human explains the content by following a logic order, the 2D visualization includes a set of links corresponding to the different topics of the lesson; the learner can access a specific topic by selecting the corresponding link.

Finally, the third setting consists in 3D with constrained navigation, absence of virtual human, and text for the explanation. Each device is now marked by a numbered label that defines the order in which it is

Figure 6. Animating the hard disk: the upper panel is opened and the internal disk is extracted.

presented to the learner. The system provides the learner with textual information (through the OSD, see Figure 4) and activates animations (e.g., to show the internal parts of devices, as shown in Figure 6 for an hard disk). At the end of each object presentation, the learner is automatically led to the next device.

4. Conclusions

In this paper, we have presented PETS, a system that is able to flexibly support different presentation techniques for EVW. Such a flexibility allows the learner to benefit from different presentations of the same educational materials (e.g., considering hardware/software available to the learner and the individual learner's needs and preferences). Moreover, PETS can be configured according to particular instructor's strategies or considering available learning material (e.g., availability of 3D models or videos).

With respect to future goals of our research, we intend to test our system in additional educational contexts to elicit new possible requirements. Moreover, we plan to evaluate the learning benefits (e.g., by comparing different free recall performances concerning both declarative and procedural knowledge) provided by the different presentation techniques supported by the system.

While information on the topology of the virtual world can be automatically derived by using the method described in [33], information needed to enable the virtual human to interact with objects has to be manually specified. Since this work can be a time consuming task, we intend to develop an authoring tool that makes it easier to specify such information (e.g., enabling the user to directly select parts of interest on the 3D model, image or video).

5. References

[1] J. Piaget, "Development and Learning", *Research in Science Teaching, 2*, 1964, pp. 176-186.

[2] L. Vygotsky, Thought and Language, Cambridge, MIT Press, 1962.

[3] S. Conceição-Runlee, and B.J. Daley, "Constructivist Learning Theory to Web-based Course Design: An Instructional Design Approach", *Annual Midwest Research-to-Practice Conference in Adult, Continuing and Community Education*, Muncie, Indiana, U.S.A, 1998.

[4] E. Bredo, Reconsidering Social Constructivism: the Relevance of George Herbert Mead's Interactionism, Chapter 5 in Phillips, D.C. (ed.) *Constructivism in Education: Opinions and Second Opinions on controversial issues*, University of Chicago Press, 2000.

[5] B. Harper, G. Hedberg, and R. Wright, "Who benefits from Virtuality?", *Computers and Education*, *34* , 2000, pp. 163-176.

[6] R. Moreno, R.E. Mayer, H. Spires, and J. Lester, "The case for social agency in computer-based teaching: Do students learn more deeply when they interact with animated agent?", *Cognition and Instruction*, *19*, 2001, pp. 177-213.

[7] R.E. Mayer, K. Sobko, and P.D. Mautone, "Social Cues in Multimedia Learning: Role of Speaker's Voice", *Educational Psychology*, *95*(2), 2003, pp. 419-425.

[8] R. Moreno, and R.E. Mayer, "Personalized Messages That Promote Science Learning in Virtual Environments", *Educational Psychology*, *96*(1), 2004, pp. 165-173.

[9] R.K. Atkinson, "Optimizing Learning From Examples Using Animated Pedagogical Agents", *Educational Psychology*, *94*(2), 2005, pp. 416-427.

[10] J. Cassell, T. Bickmore, M. Billinghurst, L. Campbell, K. Chang, H. Vilhjàlmsson, and H. Yan, "Embodiment in Conversational Interfaces: REA", *Proc. CHI 99*, ACM Press, New York, 1999, pp. 520-527.

[11] K.E. Link, R.J. Kreuz, and A.C. Graesser, "Factors that Influence the Perception of Feedback Delivered by a Pedagogical Agent", *Speech Technology*, *4*, 2001, pp. 145-153.

[12] A. Bandura, *Social Learning Theory,* General Learning Press, New York, 1977.

[13] J. Rickel, and W. Johnson, "Animated Agents for Procedural Training in Virtual Reality: Perception, Cognition, and Motor Control", *Applied Artificial Intelligence*, *13*, 1999, pp. 343-382.

[14] D. Traum, and J. Rickel, "Embodied agents for multi-party dialogue in immersive virtual worlds", *Proc. AAMAS 02*, ACM Press, New York, 2002, pp. 766-773.

[15] R.C. Hubal, G.A. Frank, and C.I. Guinn, “Lessons learned in modeling schizophrenic and depressed responsive virtual humans for training”, *Proc. IUI 03*, ACM Press, New York, 2003, pp. 85-92.

[16] P.N. Kizakevich, M.L. McCartney, D.B. Nissman, K. Starko, and N.T. Smith, “Virtual Medical Trainer: Patient Assessment and Trauma Care Simulator”, in J. D. Westwood, H. M. Hoffman, D. Stredney, and S. J. Weghorst, *Medicine Meets Virtual Reality*, IOS Press, Amsterdam, 1998, pp. 309-315).

[17] M. Cavazza, and A. Simo, “A virtual patient based on qualitative simulation”, *Proc. IUI 03: International Conference on Intelligent User Interfaces,* ACM Press, New York, 2003, pp. 19-25.

[18] M. Ponder, B. Herbelin, T. Molet, S. Schertenlieb, B. Ulicny, G. Papagiannakis, N. Magnenat-Thalmann, and D. Thalmann, “Immersive VR Decision Training: Telling Interactive Stories Featuring Advanced Virtual Human Simulation Technologies”, *Workshop on Virtual Environments* (EGVE), ACM Press, New York, 2003, pp. 97-106.

[19] M. Ponder, G. Papagiannakis, T. Molet, N. Magnenat-Thalmann, and D. Thalmann, “VHD++ Development Framework: Towards Extendible, Component Based VR/AR Simulation Engine Featuring Advanced Virtual Character Technologies”, *Proc. CGI 03*, IEEE Computer Society Press, 2003, pp. 96-104.

[20] E. M. Sims, “Reusable, Lifelike Virtual Humans for Mentoring and Role-Playing”, *Computers and Education*, 2006, in press.

[21] D.M. Dehn, and S. van Mulken, “The impact of animated interface agents: a review of empirical research”, *International Journal of Human-Computer Studies*, *52*, 2000, pp. 1-22.

[22] S. van Mulken, E. Andrè, and J. Muller, “The persona effect: how substantial is it?”, *Proc. CHI 98*, ACM Press, New York, 1998, pp. 56-66.

[23] J.C. Lester, and B.A Stone, “Increasing believability in animated pedagogical agents”, *Proc. Autonomous Agent*, ACM Press, New York, 1997, pp. 16-21.

[24] L. Sproull, M. Subramani, S. Kiesler, J.H. Walker, and K. Water, “When the interface is a face”, *Human-Computer Interaction*, *11*, 1996, pp. 97-124

[25] J. Cassell, and K.R. Thòrisson, “The power of a nod and a glance: envelope vs. emotional feedback in animated conversational agents”, *Applied Artificial Intelligence*, *13*, 1999, pp. 519-538.

[26] J.C. Lester, S.A. Converse, S.E. Kahler, S.T. Barlow, B.A. Stone, and R.S Bhoga, “The Persona Effect: Affective Impact of Animated Pedagogical Agents”, *Proc. CHI 97*, ACM Press, 1997, pp. 359-366.

[27] R.K. Atkinson, “Optimizing Learning from examples using animated pedagogical agents”, *Educational Psychology*, *94*(2), 2002, pp. 416-427.

[28] R.E. Mayer, *Multimedia learning*, Cambridge University Press, New York, 2001.

[29] H. Noot, and Z. Ruttkay, “Variations in gesturing and speech by GESTYLE”, *International Journal of Human-Computer Studies*, 2005, *62*(2), 2005, pp. 211-229.

[30] L. Chittaro, and L. Ieronutti, “A Visual Tool for Tracing Behaviors of Users in Virtual Environments”, *Proc. AVI 04*, ACM Press, New York, 2004, pp. 40-47.

[31] L. Ieronutti, R. Ranon, and L. Chittaro, “High-Level Visualization of Users’ Navigation in Virtual Environments”, *Proc. INTERACT 05*, Springer-Verlag, Berlin, 2005, in press.

[32] L. Ieronutti, and L. Chittaro, “A Virtual Human Architecture that Integrates Kinematic, Physical and Behavioral Aspects to Control H-Anim Characters”, *Proc. of Web3D 05*, ACM Press, New York, 2005, pp. 75-83.

[33] L. Ieronutti, R. Ranon, and L. Chittaro, “Automatic Derivation of Electronic Maps from X3D/VRML Worlds”, *Proc. Web3D 04*, ACM Press, New York, 2004, pp. 61-70.

[34] H-Anim (2005). ISO/IEC FCD 19774. http://www.h-anim.org/Specifications/H-Anim200x/ISO_IEC_FCD_19774/ (last access on September 2005)

[35] P. Ekman, W.V. Friesen, and D. Lutzker, “Psychological reactions to infantry basic training”, *Journal of Consulting Psychology*, *1*, 1962, pp. 103-104.

[36] P. Ekman, W.V. Friesen, and P. Ellsworth, *Emotion in the human face: Guidelines for research and an integration of findings*, Pergamon Press, New York, 1972.