Adaptable Visual Presentation of 2D and 3D Learning Materials in Web-based Cyberworlds

Luca Chittaro, Lucio Ieronutti and Roberto Ranon

HCI Lab, Dept. of Math and Computer Science, University of Udine,

Via delle Scienze 206, 33100, Udine, Italy

{chittaro, ieronutt, ranon}@dimi.uniud.it

http://hcilab.uniud.it

Abstract

Educational Cyberworlds allow one to circumvent physical, safety, and cost constraints that often affect real-world training and learning scenarios, allowing students to study the learning content flexibly and individually. However, this does not eliminate the need for presenting the educational materials in an effective way, using the visualization, interaction and presentation techniques that are better suited to the learner's preferences, abilities, and available devices. The system we present in this paper is able, starting from a set of available educational resources (e.g., text, pictures, movies and 3D models) and lesson structure (e.g., suitable ordering of topics), to automatically build a set of alternative Web-based educational Cyberworlds characterized by different visualization (2D and 3D), interaction and presentation techniques, including virtual humans as instructors. The learner (or her instructor) then can choose the Cyberworld that best fits her preferences and needs.

Keywords: Educational Cyberworlds, Visual Presentation Techniques, Virtual Humans.

1. Introduction

Traditional teaching involves direct communication between learners and a teacher, where the teacher presents available educational materials choosing the presentation technique that best suits the needs of the learner or classroom. Web-based educational Cyberworlds are changing this traditional face-to-face relation, since it is possible for learners to access educational materials without needing the presence of the teacher within the same physical environment and at the same time. The advantage is that learners can progress on their own to study the content of the course flexibly and individually [1], choosing what to view, how long to view it and how many times to view it [2].

In recent years, a number of different visual presentation techniques for Web-based educational materials have been proposed. These techniques rely on various media (e.g., text, images, movies, 3D models and virtual environments, hereinafter VEs), and employ various forms of interaction and navigation support (such as embodied agents) to help the learner in accessing the learning content. However, studies performed so far do not allow one to establish which presentation techniques are more effective and when (in the case of animated agents, for example, results are mixed, as we discuss in the next Section). Moreover, the effectiveness of educational presentations might be different depending on the learner's preferences, abilities and available device.

To experiment with different possibilities, we have developed *PALE* (*Presentation of Adaptable Learning Materials*), a system that supports adaptable presentations of learning materials in Web-based Cyberworlds. Unlike the term *adaptive* (which means that the presentation can be customized by the system based on the user's profile), *adaptable* as defined in [3][4] means that the presentation can be customized by the user: in other words, the presentation is configurable. PALE integrates multimedia educational information using a range of visualization, presentation and interaction support techniques and lets the student (or her instructor) choose the ones that best fit her interaction context (e.g., device, available bandwidth) and preferences (e.g., familiarity with navigation in 3D VEs).

The paper is structured as follows. In Section 2, we illustrate some benefits and issues in using Cyberworlds for learning purposes. In Section 3, we illustrate PALE and the presentation techniques it supports. In Section 4, we describe the PALE architecture and finally, in Section 5, we discuss future work.

2. Educational Cyberworlds

Educational Cyberworlds have a great potential for learning and training purposes, since they can allow one to circumvent physical, safety, and cost constraints. Moreover, they allow one to integrate complex multimedia elements into the learning experience, using either simpler 2D visualizations (e.g., based on HTML pages) or more complex 3D visualizations (e.g., using Web3D open standards, such as X3D [5] and VRML [6]). One of the main advantages of 3D visualizations is that they give learners the possibility of analyzing the same subject or phenomenon from different point of views and create more complete and correct mental models to represent it. A thorough discussion on the pedagogical basis, motivations and current issues for 3D Cyberworlds is outside the scope of this paper and has been provided in [7].

One important issue is that there is no general consensus in the literature about the presentation and interaction techniques that are more effective for educational purposes. In the case of embodied agents, for example, the debate even concerns their usefulness, with arguments both for (e.g., [8]) and against (e.g., [9]) their use.

Supporters of embodied agents claim that they enhance several factors affecting the learning process, such as: (i) *motivation*, supporting cognitive functions such as problem solving, understanding and learning [10], (ii) *learnability*, facilitating interaction with the Educational Cyberworld by enabling people to apply social heuristics [11], (iii) *believability*, supporting the perceived plausibility of information provided by the system with expressiveness [12], (iv) *attention*, capturing attention with facial display [13], and (v) *communication efficiency*, providing non-verbal cues on the state of the dialogue [14]. Moreover, Lester et al. [8] state that the simple "presence" of a virtual agent in an Educational Cyberworld improves learner's perception of her educational experience (the *persona effect*). *Social learning theory* [15] provides more motivations to the exploitation of virtual humans behaving as teachers, since it claims that individuals can learn by observing other individuals acting. 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).

Opponents of embodied agents argue that humanization of an interface might have negative impact from a human-computer interaction point of view, since it may produce false mental models of the agent's capabilities [16]. For example, agents which exhibit human-like behavior may be perceived as more intelligent than they actually are, leading to incorrect expectations about the system's abilities. Another argument is that embodied agents can lead the learner to become easily distracted from the task they are performing.

Finally, the evaluations that have been carried out provide conflicting results. The studies are mainly focused on evaluating different presentation techniques employed by virtual humans, for example comparing the effectiveness of presenting information to the learner in a textual format or using narrations. Some studies (e.g., [17][18]) found that a narrating agent has a positive effect on learner's recall, transfer and interest, and measured better performance when learners were instructed by a talking virtual human than when they were just instructed by a virtual instructor and written text. However, recent studies demonstrated that such effectiveness depends both on the type of virtual human voice [19] (recorded human voice is better than synthesized voice) and on the type of virtual agent's speech [20] (e.g., colloquial speech is better than impersonal one e.g., third person). Additionally, some studies [21] show that, although the presence of a human face can provide important extra conversational cues, it requires more effort from the human interacting with the system and sometimes serves as a distraction.

Besides presentation and interaction techniques, there are also a number of *extrinsic* factors that can influence the effectiveness of an educational Cyberworld. Such factors range from the equipment available to the learner (e.g., device, available bandwidth), her experience, abilities and preferences (e.g., 3D navigation skill, physical disabilities), to the type and quality of available educational material (e.g., 3D models, images, movies).

The aim of this paper is to present a system that flexibly supports different presentation and interaction techniques, allowing learners and teacher to experiment with different settings and see which ones are best suited to the specific learning context.

3. The PALE System

The PALE system builds Web-based educational Cyberworlds using different presentation and interaction techniques. The instructor or courseware designer provides PALE with her available educational resources (e.g., text, pictures, movies and 3D models) and presentation structure (e.g., suitable ordering of topics) and the system is able to build a set of alternative educational Cyberworlds that, for example, may be based on 2D or 3D graphics and may include or not animated virtual instructors. The learner (or possibly her instructor) then chooses the Cyberworld that best fits her preferences and needs. In the following, we first motivate and discuss the available presentation and interaction techniques, and then we describe the architecture of the system and discuss implementation details.

We will use two examples of educational Cyberworlds we have recently developed. The first one (*Computer Science Museum*) is concerned with computer science history and presents computing devices from the seventies (more specifically, hardware from the Univac/Sperry 90/30 line), providing information about their features, functioning and differences with today's computers. The second example (*E-Dvara*) has been developed in the context of an Indian-European cross-cultural project, and is concerned with Indian art, presenting a collection of sculptures and their links with Indian astrology and religion.

3.1. Available Presentation and Interaction Techniques

The range of educational Cyberworlds that PALE can build is controlled by four parameters: *visualization* (3D or 2D), *navigation* (free or constrained), *virtual instructor* (present or absent), *explanation* provided to the learner (audio narration or written text). Table 1 lists the resulting sixteen possible combinations.

	visualization	navigation	virtual instructor	explanation
1	2D	free	absent	text
2	2D	free	absent	audio narration
3	2D	free	present	text
4	2D	free	present	audio narration
5	2D	constrained	absent	text
6	2D	constrained	absent	audio narration
7	2D	constrained	present	text
8	2D	constrained	present	audio narration
9	3D	free	absent	text
10	3D	free	absent	audio narration
11	3D	free	present	text
12	3D	free	present	audio narration
13	3D	constrained	absent	text
14	3D	constrained	absent	audio narration
15	3D	constrained	present	text
16	3D	constrained	present	audio narration

Table 1. Combinations of presentation and interaction techniques supported by PALE

In the following, we discuss separately each parameter.

3.1.1. Visualization

The *visualization* parameter allows the learner to choose between a learning experience exploiting interactive 3D graphics (*3D*) and a more traditional one based on Web pages (*2D*).

In the case of 3D visualization, the learner is immersed into an X3D/VRML VE; this type of visualization is suited for first-person learning experiences, where real world-like situations can be reproduced using, for example, interactive 3D models, simulations, animations, and environmental sounds.

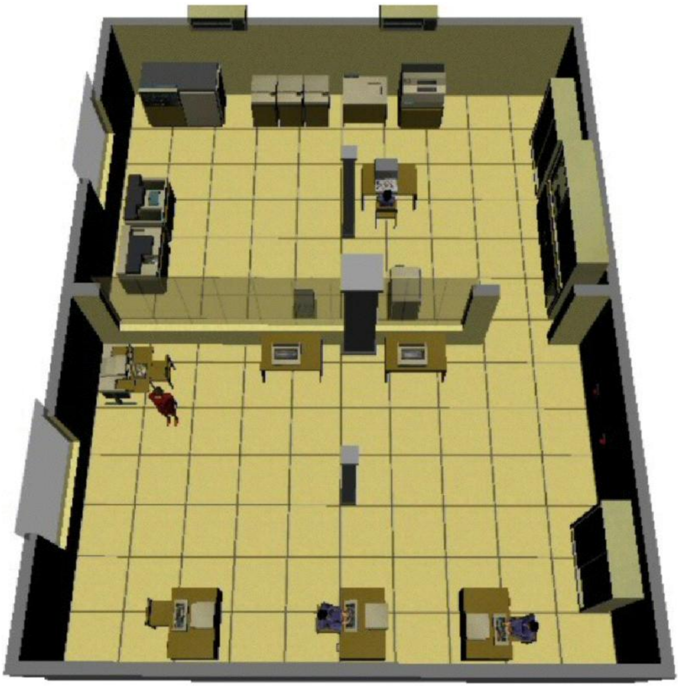

Figure 1. The VE for the Computer Science Museum used in the 3D visualization.

For example, Figure 1 shows the 3D VE for the Computer Science Museum reproducing a data processing center from the seventies where interactive 3D models of computing devices, such as terminals, printers, disk drives (see Figure 2) and computers, are presented to the learner.

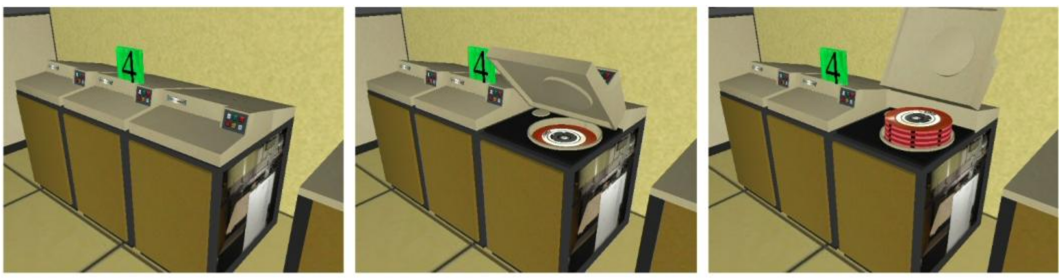

Figure 2. An example of interactive object in the Computer Science Museum: the hard disk cabinet can be opened and the disk extracted.

In the case of 2D visualization, the learner is presented with a Web page where topics are presented using text, images and movies. For example, Figure 3 shows the 2D visualization of E-Dvara, where sculptures are visualized using pictures, while Figure 4 shows the 2D visualization of the Computer Science Museum.

While simpler, 2D presentation is the best choice in situations such as the following: (i) when a full 3D VE is not available, because it is too costly to build (ii) when the VE is available, but is too complex for the learner's device (e.g., in the case of PDAs) and/or available bandwidth; (iii) when the learner is not familiar or has interaction difficulties with 3D VEs.

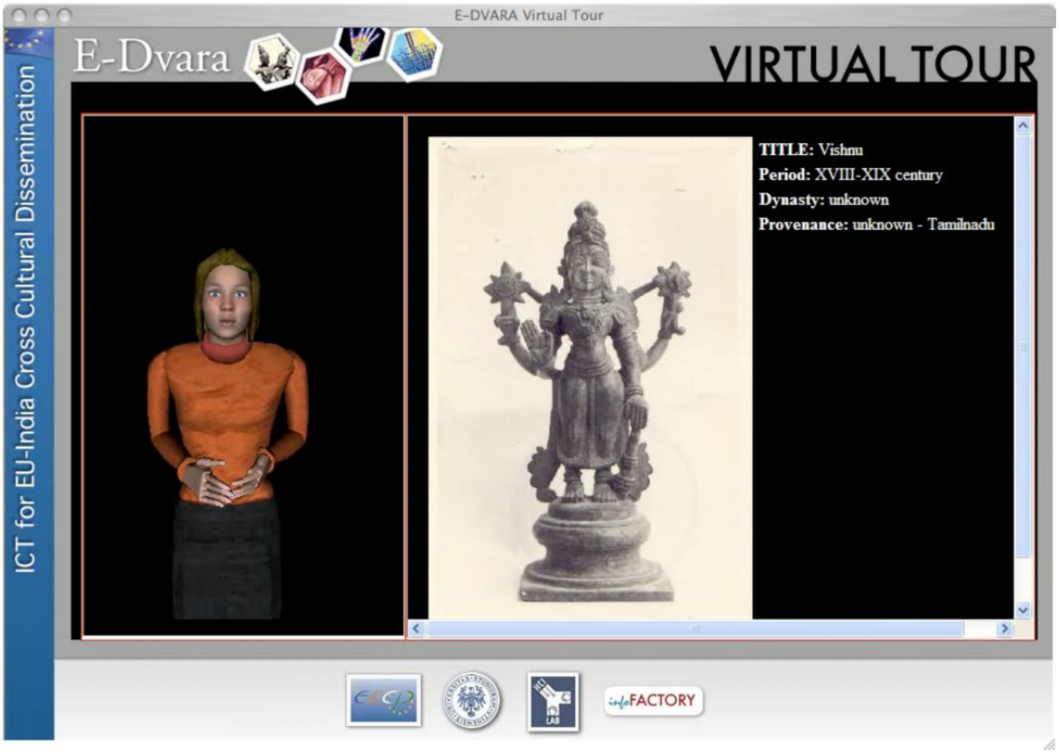

Figure 3. The 2D visualization of the E-Dvara application using a virtual instructor, constrained navigation and audio narration.

3.1.2. Navigation

PALE supports two different types of navigation of learning content: *free* and *constrained*. The first one is more suited to learners that prefer to choose the order in which learning material is presented, and, in the case of 3D visualization, have the ability to navigate on their own. The second one is more suited when it is mandatory that content is presented in a suitable order (e.g., at the beginning of a new course) or when the learner has not the necessary ability to navigate on her own.

The idea is that the instructor (or courseware designer) specifies a suitable ordering of topics, which is only suggested by PALE when free navigation is selected, while it is enforced in the case of constrained navigation.

When free navigation is selected in the 3D visualization, the learner navigates inside the 3D VE by controlling the position and orientation of her viewpoint (e.g., through mouse or arrow keys). In the 2D visualization, the learner can access a *topics* menu from which any specific content can be chosen (for example, in Figure 4, the learner has chosen to access the topic describing the main printer).

When constrained navigation is selected in the 3D visualization, viewpoint movements are controlled by the system to focus the learner's attention towards parts of the 3D VE (in the given order) and avoid occlusion problems. In the case of 2D visualization, learning content is presented sequentially following the given order, as in Figure 3.

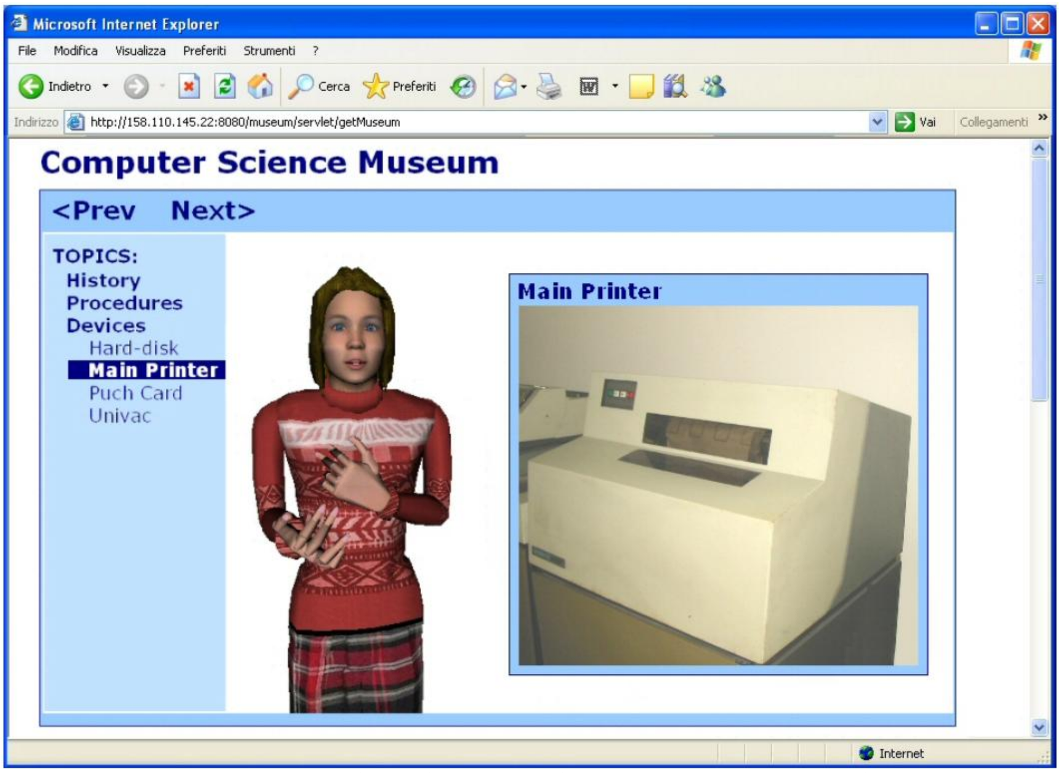

Figure 4. 2D Visualization of the Computer Science Museum using free navigation, a virtual instructor and audio narration.

3.1.3. Virtual Instructor

PALE allows the courseware designer to include (and control) a virtual human that acts as instructor, providing both pedagogical and navigational assistance in the learning experience.

With 3D visualization, the virtual human leads the learner through a tour of the environment and explains the educational content with both declarative (e.g., object descriptions) and procedural (e.g., showing how to perform physical tasks, as in Figure 5 where it shows how to open the main printer) information. In the case of constrained navigation, the viewpoint position and orientation is automatically set to follow the virtual instructor during its movement through the VE, while, in the case of free navigation, the learner can decide whether to follow the tour proposed by the virtual instructor or not (in this case, the virtual instructor can follow the user and provide explanations on what the user is focusing on).

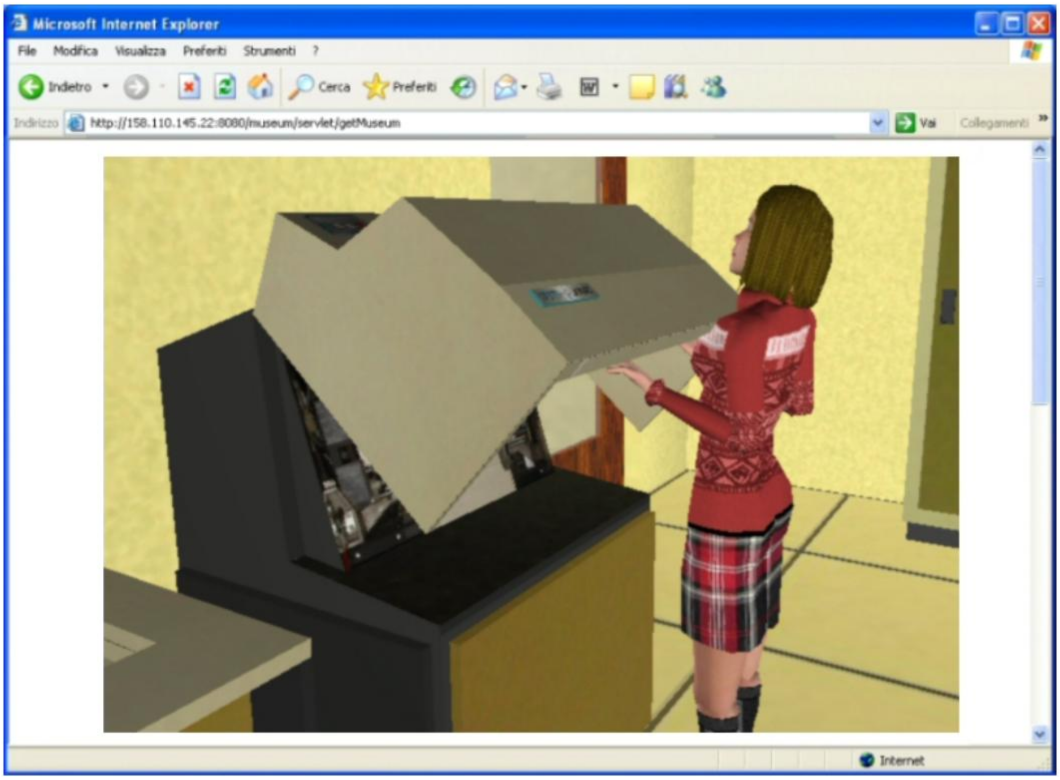

Figure 5. Virtual instructor interacting with objects.

With 2D visualization, regardless of the value of the navigation parameter, the virtual instructor is displayed on the Web page (as in Figures 3 and 4) and explains the learning content, also drawing the learner's attention to specific parts of images and movies (e.g., by pointing at them) as a real lecturer would do with slides.

When the virtual instructor is absent, a set of labels (as in Figure 6) are inserted in the 3D VE and can be clicked to allow the learner to get more information on specific objects or places. Labels are marked with numbers to suggest the visiting order.

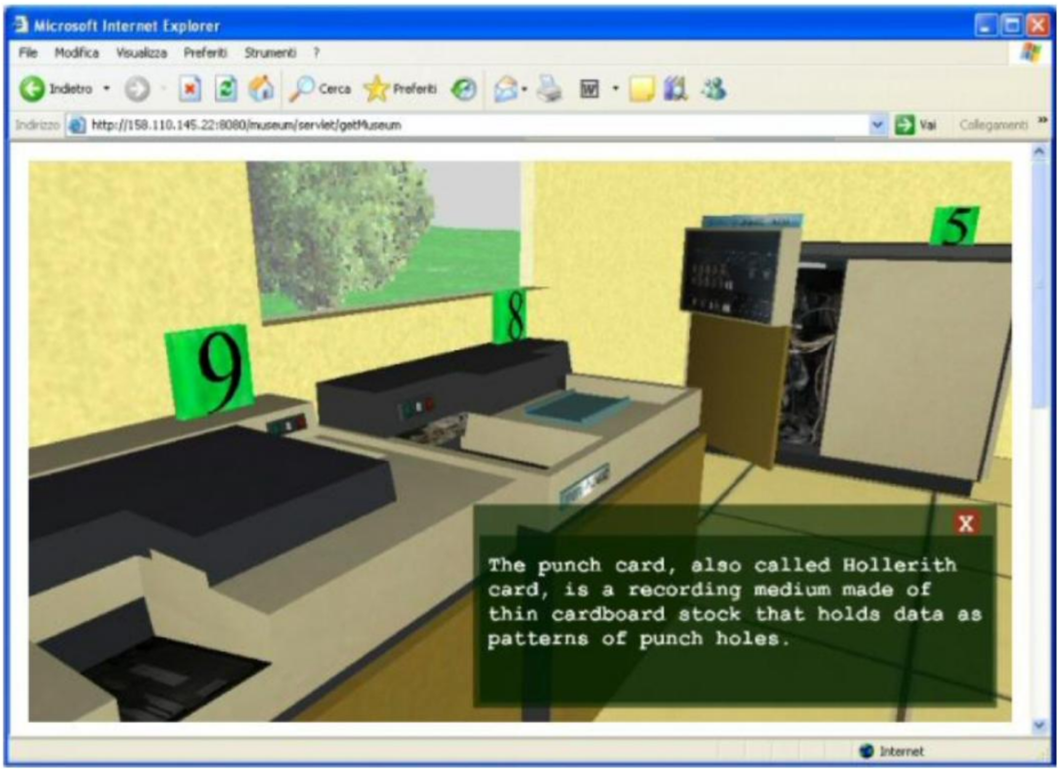

Figure 6. The 3D visualization of the Computer Science Museum without the virtual instructor and using text to describe devices.

3.1.4. Explanation

Explanations can be provided to the learner through audio narration or text. The audio required for the narration can be either pre-recorded (from a human voice) or generated during the presentation from text by a Text-to-Speech (TTS) engine, which has the advantages of faster download times and no need for voice recording sessions. However, the synthesized voice can be perceived by learners as unexpressive, unnatural and impersonal.

When audio narration is chosen, it is spoken by the virtual instructor if present, or simply played during the learning experience.

When text is chosen, in the case of 3D visualization, it is displayed on a semi-transparent and resizable On Screen Display (OSD), as in Figure 6. Text in the OSD is automatically formatted according to display size and the learner can move and resize the OSD as needed. In the case of 2D visualization, text is just written in the Web page.

4. System Architecture and Implementation

The PALE system is composed by five main modules (see Figure 7): the *Data Acquisition* and *Presentation Manager* modules run on client side (a Web Browser), the *Recorder* and *Loader* modules run on server side. When the virtual instructor is used for the presentation, a module implementing the *Virtual Human Architecture* (*VHA*) [22] (see Section 4.3) runs on client side to control and animate the virtual human. Finally, two databases are used, respectively to store learning materials and usage data. In the following, we explain the functions of the different modules.

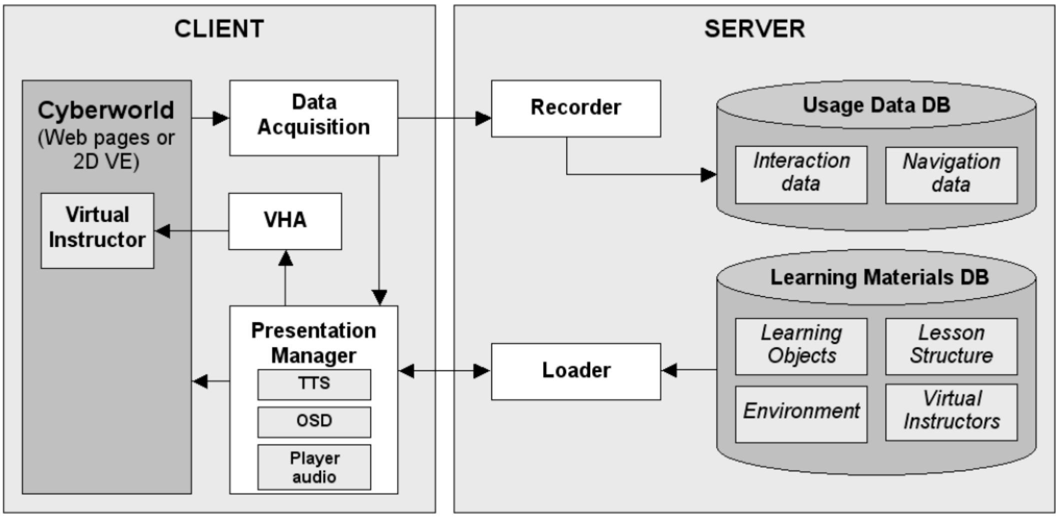

Figure 7. Architecture of PALE.

The Data Acquisition module monitors usage data and sends it both to the Recorder and Presentation Manager modules. In the case of 2D visualization, usage data logs requested pages. In the case of 3D visualization, usage data logs both the navigational behavior of the learner and her interactions with objects. More specifically, at the beginning of a session, Data Acquisition integrates into the VE a proper set of X3D/VRML sensors and then listens for events generated by them. For acquiring navigational information, Data Acquisition uses a `ProximitySensor` node that monitors the current position and orientation of the learner into the VE. For obtaining information on learner's actions, the module senses `TouchSensor` (i.e. touch time), `PlaneSensor` (i.e. object translation), `CylinderSensor` and `SphereSensor` (i.e. object rotation) nodes associated to the objects in the VE. However, since some sensors can generate a lot of events (e.g., the `ProximitySensor` node continuously generates data as the learner moves in the VE), Data Acquisition filters them by sampling at constant time intervals.

The Recorder module stores usage data into the *Usage Data* database, together with proper identifiers for the considered learner and session. This data can be later analyzed, e.g., to (i) evaluate how and to what extent the learner accessed the learning material, (ii) update a model of the learner's knowledge, e.g., for delivering personalized exercises, and (iii) discover problems in the course or in PALE itself.

At the beginning of a session, the Loader module, considering information stored into the Learning Materials database (described in detail in the following section), provides the learner with a set of alternative Cyberworlds (each one characterized by a specific setting of presentation parameters) and the learner can choose the one that best fits her preferences and needs. Considering the choice of the learner, the Loader retrieves required 2D and/or 3D resources from the Learning Material database and sends them to the Presentation Manager. This module, using either pre-defined templates (specifying the Web page layout) in the case of 2D visualization or VE models in the case of 3D visualization, includes the learning materials into the presentation.

During the learning session, the Presentation Manager, considering usage data and information describing the lesson structure, determines the learning material to be presented to the learner and activates presentation animations. Moreover, when the virtual instructor is present, the Presentation Manager is able to send to the VHA the set of animation commands to be executed by the virtual instructor.

In the following, we first describe in detail how the information stored into the Learning Material database is structured, how the Presentation Manager works, and the main VHA functionalities exploited by PALE.

4.1. Learning Material database

Resources and information needed for the presentation are stored into the Learning Materials database. In particular, the database stores information on *Learning Objects*, the digital resources representing the educational content, *Lessons*, specifying how learning content has to be presented to the learner, *Virtual Instructors*, 3D models of H-Anim [23] virtual humans, and the

Environment in which learning takes place. In the following, we describe in more detail each element stored into the Learning Material database.

4.1.1. Learning Objects

The main idea of Learning Objects (hereinafter, LOs) is to break educational content into small instructional components that can be reused a number of times in different learning contexts [24]. Each LO is a structured and standalone resource that encapsulates didactical, educational and/or pedagogical information using a collection of digital resources (typically, multimedia elements). Each LO should be [24]: (i) *self-contained* (can be taken independently), (ii) *reusable*, (can be used in different contexts for different purposes), and (iii) *aggregable* (can be grouped into larger collections of content).

In PALE, each LO is characterized by the following fields (see Figure 8):

- *name*, the identifier of the LO,
- *prerequisites*, the knowledge required for understanding the learning content in terms of a set of other learning objects,
- a set of digital *resources*, describing the learning content together with presentation-related metadata.

Resources can be classified in two different categories, narrative *descriptions* and visual *representations* of the learning content, each one definable using alternative data formats: narrative descriptions can be stored using audio and textual data, while visual representations can be defined using images, movies and 3D models.

For example, the LO depicted in Figure 8 contains information on the `Card Reader` device and is characterized both by descriptive information in the form of textual and audio data, and by visual representations of the device in the form of an image and a 3D model. Moreover, the considered LO can be accessed by the learner only if the LO named `Memory Devices` has been previously presented.

In PALE, each LO resource is defined by (i) *url* of the resource, and (ii) *metadata* specifying additional information associated to the resource. Metadata plays a fundamental role in PALE, since (i) it allows the system both to synchronize a narrative description with animations on a LO representation, and (ii) it enables the virtual instructor to interact with a LO representation.

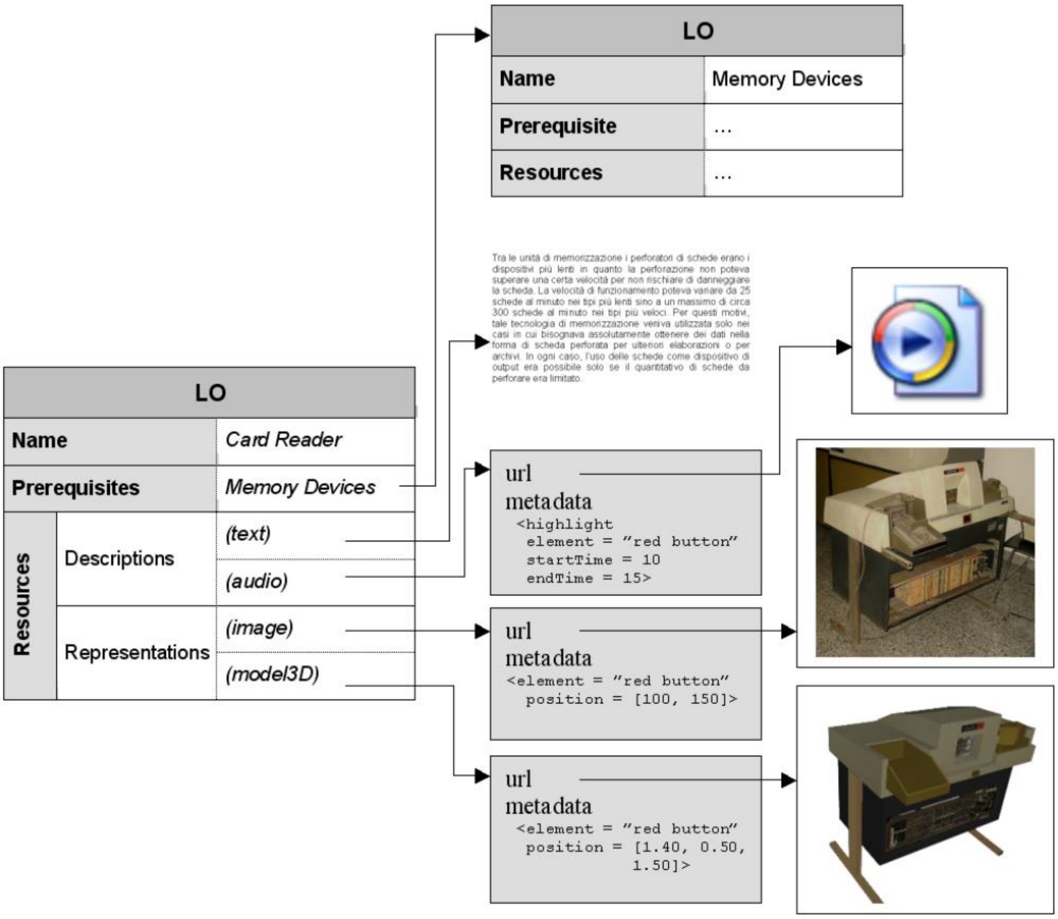

Figure 8. PALE Learning Object providing information on the card reader.

Information specified into metadata depends on the type of resource; in the following, we describe metadata associated to visual representations and narrative descriptions.

Representation Metadata

The main purpose of representation metadata is to store the information required for generating and running animations. Representation metadata can specify two different types of information:

- *positioning information*, required for dynamically generating animations, that identifies specific elements on the LO representation by using:
  - 2D coordinates in the case of images;
  - 2D coordinates and timing information in the case of movies;
  - 3D coordinates in the case of 3D models.
- *animation data*, that stores pre-computed animations. It consists of:

- position, orientation, scalar, color and/or coordinate X3D/VRML interpolators together with timing information in the case of 3D models;
- couples `(startTime, endTime)` that identify portions of movies.

In the following, we provide a simple example considering content metadata of the main printer in the Computer Science Museum. In the case of 3D visualization with virtual instructor, positioning information can specify the location of buttons on the printer. This data can be used to allow the virtual instructor to interact with the printer, for example to push the button that turns the device on (see Figure 9).

Figure 9. The Virtual instructor turning on the main printer.

In the case of 2D visualization, the printer is represented by an image and the turn on button is identified by means of 2D coordinates on the image itself. This information can be used for example to highlight on the image the turn on button. If there is a movie showing a set of procedures to operate the printer, the movie clip showing the turn on procedure can be extracted from the entire movie by using proper `startTime` and `endTime` values. Finally, in the case of 3D visualization and when the virtual instructor is absent, specific pre-computed animations are defined by X3D/VRML interpolators and timing information. For example, the animation demonstrating the opening procedure can be defined by specifying both the rotation to be applied to the upper panel of the printer and the duration of the animation.

Description Metadata

Description metadata specifies a list of *system commands* to be executed at particular instants when presenting the LO. There are two kind of commands, to respectively *highlight* a specific part on the selected LO representation or *play* an animation (stored in the representation metadata or built from them).

Each command produces different effects according to the values of the presentation and interaction parameters. For example, the highlight command can use the virtual instructor gaze and pointing gestures for guiding the learner's attention toward the element of the representation that needs to be highlighted (see Figure 10).

Figure 10. Highlighting the red button of the card reader using 3D visualization and virtual instructor.

If the virtual instructor is absent, the same highlight command draws the learner's attention by drawing arrows or colored circles (see Figure 11) over the LO representation.

Figure 11. Highlighting the red button of the card reader using 2D visualization and no virtual instructor.

The syntax used for specifying system commands depends on the type of resource used for the narrative description. If the description is in the form of an audio file, each system command is specified by a triple consisting of the type of command (*highlight* or *play*), a reference to a representation metadata (e.g., identifying a part of the representation) and a set of timing information.

For example, considering the LO `card reader` (represented in Figure 8), metadata associated to audio description is composed by the single command `<highlight element="red button" startTime = 10 endTime = 15>`. This command specifies that the red button of the card reader (whose location is specified into representation metadata) has to be highlighted from the tenth to the fifteenth second of audio reproduction.

If a narrative description is given in textual format, system commands are written into the text by means of tags. In this case, each command is specified by a pair consisting of the type of command and a reference to a representation metadata (in this case, timing information is not required). For example, considering the LO about the card reader, suppose that the textual description "... *the card reader can be turned on by pressing the red button for five seconds* ..." needs to be associated with an highlight command of the `red button` on the selected `card reader` representation. This would be achieved by inserting the highlighting command into the original textual description as follows "... *the card reader can be turned on* `<highlight element = "red button">` *by pressing the red button for five* *seconds* `</highlight>` ...". Also in this case, the final effect depends on the value of the *explanation* parameter. When text is used for explanations, the text inside the command tags is underlined and the learner can click on it to activate the corresponding command. If audio is used for explanations, a synthesized voice is generated from the textual description, and each command is executed when the text inside its tags is being read. In this way, animations and visual effects are automatically synchronized with the explanation.

4.1.2. Lessons

For each lesson, the learning materials database specifies:

- the order in which LOs have to or should be presented to the learner;
- the environments (3D VE and Web pages) to be used for the lesson and where LO representations have to be positioned in the environments themselves;
- the virtual human that should be used as virtual instructor.

4.1.3. Virtual Instructor

The Learning Materials database contains the set of virtual human models that can be used for embodying the virtual instructor and associated animations. To guarantee the interchangeability of different virtual instructors, PALE requires them to be compliant with the H-Anim standard [23], i.e. the international standard for representing humanoids in network-enabled 3D graphics and multimedia environments. Associated animations are stored as sequence of postures (each one characterized by a specific configuration of the virtual human joints) and timing information. The standard joint structure defined by H-Anim allows one to run the same animation on different virtual human models.

4.1.4. Environment

The environment is specified as a template for a 3D VE or a Web page.

In the case of a 3D VE, the template contains the geometric model of the environment (e.g., walls, windows and doors) and information on its topology, which is used to automatically compute collision-free paths for the virtual instructor and for viewpoint movement with constrained navigation. Topological information contains:

- a *2D map* of the VE specifying which areas can be traveled and which ones contain objects preventing navigation (e.g., walls). If the VE is composed by multiple floors (e.g., a building), one needs to provide a 2D map for each of them;
- *connectivity information*; in the case of multi-floor VEs, information on how different floors are connected (e.g., where stairs are) is also required to compute a path connecting arbitrary start and end positions.

In the case of a 2D Web page, the template defines the Web page layout, including where the topics menu, virtual human, and LO representations have to appear.

4.2. Presentation Manager module

The Presentation Manager module, using information on the lesson structure and learner's actions, is able to retrieve at execution-time the learning materials to be presented and integrate associated resources into the presentation. LO representations (e.g., an image, movie or 3D model) are simply integrated into the environment template, while LO descriptions are transformed for presentation according to the value of *explanation* parameter and then shown or reproduced.

For example, if a narrative description is in textual format but audio narration is chosen, the Presentation Manager module uses a TTS to transform the text into audio format and reproduce it. Moreover, if the virtual instructor is present during audio reproduction, the Presentation Manager computes and sends to the VHA the correct visemes (i.e. mouth configurations corresponding to a particular sound in spoken language). This data is used by the VHA to synchronize the virtual instructor lip movements with spoken text.

If explanations have to be given using text in 3D visualization, the Presentation Manager formats textual descriptions according to the OSD size.

4.3. Virtual Human Architecture

The VHA is a software architecture that allows one to develop interactive H-Anim [23] virtual humans by separating the programming of geometrical, kinematic, physical and behavioral aspects [22]. Although the VHA supports a wide range of different functionalities, in PALE it is mainly used for generating virtual instructor animations at execution-time and applying them to the virtual human model. Starting from any animation command (e.g., `<play animation = ”walkTo” element = ”printer”>`), the VHA computes corresponding animations by considering anthropometric measures of the H-Anim model (e.g., lengths of arms and legs), environment and object information (e.g., where an object is positioned into the VE).

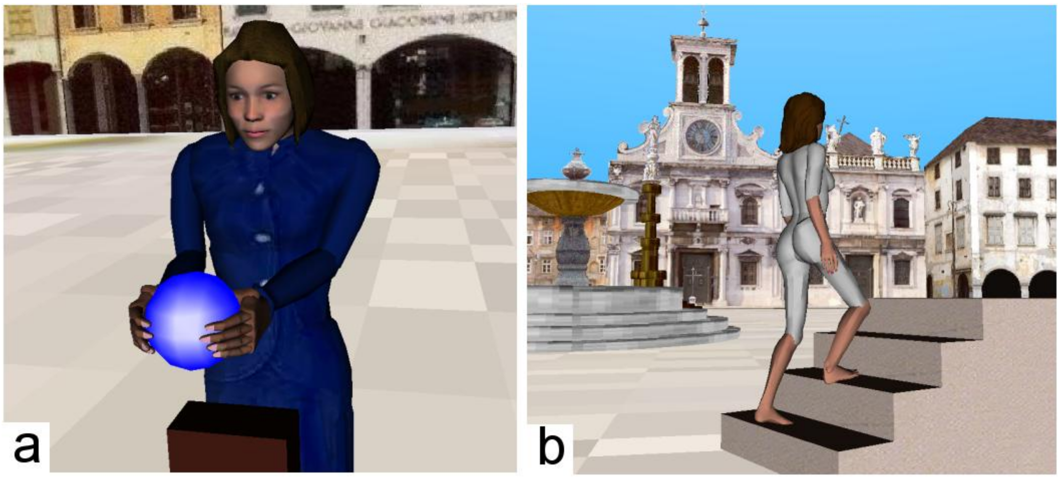

Figure 12. Examples of parameterized animations supported by VHA: a) grasping and b) walking on uneven surfaces.

The VHA supports different *animation models*, each one describing how a specific animation has to be generated starting from a set of given parameters. Animation models use pre-computed data, inverse kinematics, physically-based simulation, path planning and collision detection algorithms to generate the motion. Different animation models are supported, including a walk model, a grasp model (Figure 12 shows examples of animations produced by the walk and grasp models), a model controlling the virtual human facial expressions and a model managing gaze and head movements.

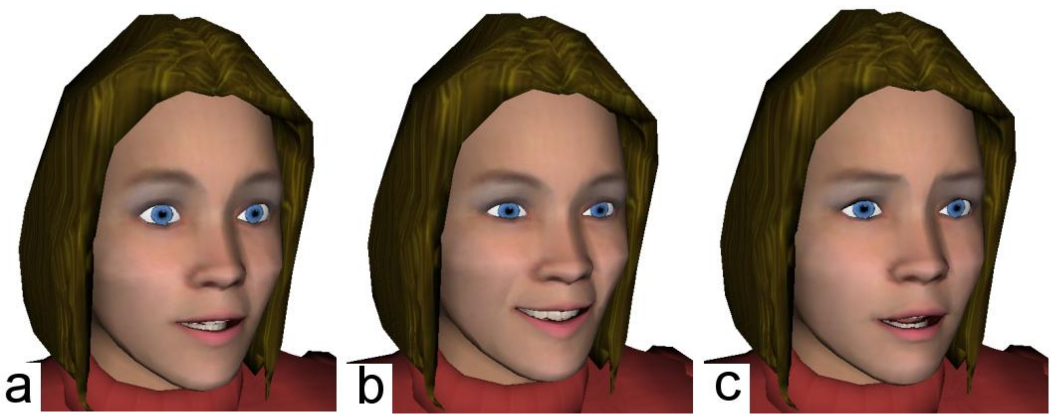

Figure 13. Examples of expressions supported by VHA: a) surprise, b) joy and c) sadness.

The walk model allows the virtual human to walk on both even and uneven terrains in multi-floor environments. The walk model combines the automatic generation of human walking motion (based on the model proposed by [25]) with trajectory information (derived by a path planning algorithm). In particular, the walk model first uses the trajectory information and the data representing the walking surface topology to derive foot and pelvis trajectories. Then, it computes proper joint rotation values for the virtual human legs by using an inverse kinematics system. Finally, the walk model employs generic pre-computed data to animate the upper part of the virtual human body.

The grasp model employs the Multi-sensor approach [26] to simplify the collision-detection algorithm. This technique hangs spherical sensors to the articulated figure and activates each sensor for any collision with other objects or sensors. Then, the model computes joint rotation values suitable to place the virtual human hands in the required positions by using an inverse kinematics system.

For facial animation, given the set of (pre-computed) data describing possible virtual human expressions (including visemes), the VHA is able to smoothly interpolate different expressions, controlling also the intensity of each expression. For example, when audio is used for explanations, the TTS (belonging to the Presentation module) continuously sends to the VHA information on the current viseme and the system generates a smooth interpolation between the previous and the current expression.

Figure 14. Three different instants of a virtual instructor explanation; VHA controls simultaneously virtual human posture, lip and gaze movements.

In its current implementation, the VHA supports 22 different visemes, while, according to the classification proposed by Ekman [27], it uses six basic facial expressions (i.e. joy, sadness, anger, surprise, fear and disgust, see Figure 13) for representing virtual human emotions.

It is also important to note that the VHA is able to control concurrently different parts of the virtual human; for example, it is able to animate the virtual human body while controlling both its facial expression and gaze (see Figure 14).

5. Conclusions

In this paper, we have presented PALE, a system that is able to support different presentation and interaction techniques for educational Web-based Cyberworlds. PALE is not intended as a full-featured educational system, since it focuses on content presentation and does not cover other features of learning systems, such as collaboration tools, student and domain modeling. With PALE, the learner can experience 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, PALE can be configured according to particular instructor's strategies or considering the availability of learning material (e.g., availability of movies or 3D models). Finally, the possibility of collecting navigational information as well as information on learner's actions allows one to use it in combination with methods for evaluating the learner's performance or, alternatively, the effectiveness of a learning content and presentation techniques. For example, data analysis applications (e.g., spreadsheets and statistical tools) can be employed for studying actions data, while navigational information can be analyzed by using tools such as VU-Flow [28][29][30] in the case of 3D visualization, while navigational information referring to 2D visualization can be studied by using various visual tools, such as VISVIP [31] and WebQuilt [32].

The primary limitation of PALE is the complexity and effort required for defining LO metadata. In particular, while information on the topology of the VE 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 determined by making calculations or by trial-and-error. 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) and to preview presentations would help greatly in solving this kind of problems.

While preliminary tests on the two educational Cyberworlds presented in the paper have shown promising results, a formal evaluation and comparison of the available interaction and presentation techniques is needed.

Finally, while using the VHA frees the courseware designer from dealing with low-level aspects in programming the virtual instructor behavior, it has the effect of constraining the virtual instructor possible pedagogical and social interaction. For example, the current implementation of the VHA is neither able to control more than one virtual instructor at the same time (as it would be needed to effectively demonstrating collaborative tasks) nor to understand (spoken or written) sentences expressed in natural language by the learner.

References

[1] Kayssi A, El-Haji A, Assir M.E, Sayyid R (1999) Web-based tutoring and testing in a computer networks course. Computer Applications in Engineering Education, 7(1):1-7

[2] Lin B.S, Hsieh C.T (2001) Web-based teaching and learner control: A research review. Computers & Education, 37:377-386

[3] Findlater L, McGrenere J (2004) A comparison of static, adaptive, and adaptable menus. Proceedings of CHI, ACM Press, New York, pp 89-96

[4] Weld D.S, Anderson C, Domingos P, Etzioni O, Krzysztof G, Lau T, et al. (2003) Automatically personalizing user interfaces. Proceedings of IJCAI, pp 1613-1619

[5] X3D (2004) ISO/IEC FDIS 19775. http://www.web3d.org/x3d/specifications/index.html

[6] VRML (1997) The Virtual Reality Modeling Language. http://www.web3d.org/x3d/specifications/vrml/ISO_IEC_20772-All/index.html

[7] Chittaro L, Ranon R (2006) Web3D Technologies in Learning, Education and Training: Motivations, Issues, Opportunities. Computers & Education, in press.

[8] Lester J, Converse S, Kahler S, Barlow T, Stone B, Bhogal R (1997) The persona effect: affective impact of animated pedagogical agents. In CHI '97: Proceedings of CHI, ACM Press, New York, pp 359-366

[9] Shneiderman B, Maes P (1997) Direct Manipulation vs. Interface Agents. Interactions 4(6): 42-61

[10] van Mulken S, Andrè E, Muller J (1998) The persona effect: how substantial is it? Proceedings of CHI, ACM Press, New York, pp 56-66

[11] Lester J.C, Stone B.A (1997) Increasing believability in animated pedagogical agents. Proceedings of Autonomous Agent, ACM Press, New York, pp 16-21

[12] Cassell J, Bickmore T, Billinghurst M, Campbell L, Chang K, Vilhjàlmsson H, Yan H (1999) Embodiment in Conversational Interfaces: REA. Proceedings of CHI, ACM Press, New York, pp 520-527

[13] Sproull L, Subramani M, Kiesler S, Walker J.H, Water K (1996) When the interface is a face. Human-Computer Interaction, 11:97-124

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

[15] Bandura A (1977) Social Learning Theory. General Learning Press.

[16] Norman D (1997) How People Might Interact With Agents. In: Bradshaw J (ed.), Software Agents, MIT Press, pp 49-55

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

[18] Atkinson R.K (2002) Optimizing Learning from examples using animated pedagogical agents. Educational Psychology, 94(2): 416-427

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

[20] Moreno R, Mayer R.E (2004) Personalized Messages That Promote Science Learning in Virtual Environments. Educational Psychology, 96(1): 165-173

[21] Takeuchi A, Nagao K (1995) Situated facial displays: Towards social interaction. Proceedings of CHI, ACM Press, New York, pp 450-455

[22] Ieronutti L, Chittaro L (2005) A Virtual Human Architecture that Integrates Kinematic, Physical and Behavioral Aspects to Control H-Anim Characters. Proceedings of Web3D, ACM Press, New York, pp 75-83

[23] H-Anim (2005) ISO/IEC FCD 19774. http://www.h-anim.org/Specifications/H-Anim200x/ ISO_IEC_FCD_25774/

[24] Wiley D (2000) Connecting learning objects to instructional design theory: A definition, a metaphor, and a taxonomy. In: Wiley D.A (ed.), The Instructional Use of Learning Objects, online version: http://reusability.org/read/chapters/wiley.doc

[25] Chung S, Hahn J.K (1999) Animation of Human Walking in Virtual Environments. Proceedings of Computer Animation, IEEE Computer Society Press, pp 4-15

[26] Huang Z, Boulic R, Magnenat-Thalmann N, Thalmann D (1995) A Multi-Sensor Approach for Grasping and 3D Interaction. Proceedings of CGI '95, Academic Press, pp 235-253

[27] Ekman P (1982) Emotion in the human face. Cambridge University Press

[28] Chittaro L, Ieronutti L (2004) A Visual Tool for Tracing Behaviors of Users in Virtual Environments. Proceedings of AVI 04, ACM Press, New York, pp 40-47

[29] Ieronutti L, Ranon R, Chittaro L (2005) High-Level Visualization of Users' Navigation in Virtual Environments. Proceedings of INTERACT 05, Springer-Verlag, Berlin, pp 873-885

[30] Chittaro L, Ranon R, Ieronutti L (2006) Visual Analysis of Navigation in Virtual Environments. IEEE Transactions on Visualization and Computer Graphics, Special issue on Visual Analytics, in press.

[31] Cugini J, Scholtz J (2002) VISVIP: 3D visualization of paths through web sites. Proceedings of Database and Expert Systems Applications, IEEE Computer Society Press, pp 259-263

[32] Hong J, Landay J (2001) Webquilt: A Framework for Capturing and Visualizing the Web Experience. Proceedings of World Wide Web, ACM Press, New York, pp 717-724

[33] Ieronutti L, Ranon R, Chittaro L (2004) Automatic Derivation of Electronic Maps from X3D/VRML Worlds. Proceedings of Web3D, ACM Press, New York, pp 61-70