Integrating a virtual agent into the real world: the virtual anatomy assistant ritchie

Volker Wiendl, Klaus Dorfmüller-Ulhaas, Nicolas Schulz, Elisabeth André

Angaben zur Veröffentlichung / Publication details:

Wiendl, Volker, Klaus Dorfmüller-Ulhaas, Nicolas Schulz, and Elisabeth André. 2007. “Integrating a virtual agent into the real world: the virtual anatomy assistant ritchie.” In *Intelligent Virtual Agents: 7th International Conference, IVA 2007 Paris, France, September 17-19, 2007*, edited by Catherine Pelachaud, Jean-Claude Martin, Elisabeth André, Gérard Chollet, Kostas Karpouzis, and Danielle Pelé, 211–24. Berlin [u.a.]: Springer. https://doi.org/10.1007/978-3-540-74997-4_20.

Integrating a Virtual Agent into the Real World: The Virtual Anatomy Assistant Ritchie

Volker Wiendl, Klaus Dorfmüller-Ulhaas, Nicolas Schulz, and Elisabeth André

University of Augsburg, Lab for Multimedia Concepts and their Applications,
Eichleitnerstr. 30, 86159 Augsburg, Germany

Abstract. Augmented realities, which are partially real and partially virtual, open up new ways for humans to interact with Embodied Conversational Agents (ECAs) since they allow users to meet ECAs in the physical space. Nevertheless, attempts to integrate ECAs as digital overlays in a physical space have been rare. Obvious reasons are the high demands such an integration puts to the animation of ECAs as virtual augmentations of the physical space, their capabilities to perceive not only the virtual, but also the physical world as well as reactive behavior control. In this paper, we describe our technical contributions towards solving these challenges. To illustrate our ideas, we present the virtual anatomy assistant Ritchie that monitors the user's actions in a physical space and dynamically responds to them.

1 Introduction

Augmented realities [1], which are partially real and partially virtual, offer great promise to educational software because they help users to perform real-world tasks in space and may provide useful additional information by enhancing the users' natural surroundings with digital overlays. They allow users to interact with physical and virtual 3D objects in a physical environment and may contribute to a better understanding of spatial concepts than a pure virtual space by conveying spatial cues in a direct manner, see also [2].

Embodied conversational agents (ECAs) bear great potential for educational software because of their motivating and engaging effect, see for example [3]. They may positively influence the learners' interaction experience by rendering presentations more lifelike and appealing and helping learners to reduce fear of failure. Most ECAs developed so far reside on the learner's desktop, have been integrated into web applications or co-habit a 3D virtual environment with learners. Hardly any attempt has been made to integrate them in the learner's real environment.

In this paper, we report on first efforts to combine the benefits of ECAs and augmented realities by employing the metaphor of a digital mirror. The user's physical environment is projected on a screen, and the video image of the real scene is augmented by virtual objects. In addition, a virtual character is rendered into the video image. The character responds to the user's actions in the real world, e.g., by providing instructions or commenting on the user's performance.

The smooth integration of virtual agents into the user's physical environment raises particular challenges to audio-visual rendering, the acquisition and analysis of context data as well as reactive behavior planning. In particular, the following requirements have to be met:

- **Seamless Integration of the Physical and the Real World**
  In order not to destroy the illusion of co-existing virtual and real worlds, we have to make sure that digital overlays are correctly rendered into the video image of the real scene. This includes the accurate handling of occlusions between digital and real objects as well as the generation of realistic shadows for digital objects falling onto real objects and vice versa. In addition, updates of the virtual world have to be performed in real-time in the case of changes in the physical world that affect the virtual world.
- **Context Sensors for the Real and the Digital World**
  Characters that inhabit Augmented Realities need to be aware of the physical as well as the digital context. As a consequence, specific support for the handling of context data is required to fuse the results from the synthetic and real vision processes. Context data may refer to user actions, but also to environmental factors, such as temperature or luminosity.
- **Context-Sensitive Behavior Control**
  The agent needs to be able to dynamically adapt its behavior to changing circumstances with and without explicit user interaction. In order to ensure that the agent spontaneously reacts to external events, it should be possible to interrupt actions, such as animations of the agent, in a natural manner and to provide smooth transitions to follow-up actions.

In the next section, we first provide an overview on existing work aiming at integrating virtual characters in the user's physical environment. We then describe the setup of the Virtual Anatomy Assistant Ritchie which teaches anatomy of the human body using a real skeleton. In the subsequent three sections, we detail our technical contributions to meet the challenges identified above. We first focus on the registration problem between the virtual and real world as well as graphical issues that have to be handled in order to smoothly embed ECAs in augmented reality scenarios. After that, we describe the employed context sensors, the ACOSAS Core Engine handling context data for dynamic behavior behavior control as well as the ACOSAS Player that implements the connection to the graphical rendering and text-to-speech synthesis. Finally, we report on the results of a user study conducted at CeBIT 2007, the world's largest computer fair.

2 Related Work

Various attempts have been made to integrate virtual characters into the user's physical world - either by projecting virtual displays on real objects or vice versa by integrating video images of real objects into a virtual scene.

Cassell and colleagues [4] present a system which allows children to play with natural figurines inhabiting a physical castle in collaboration with a virtual character. The character is projected on the screen and a digital image of physical castle on the screen provides the user with the impression that the castle continues into the screen. The authors make use of RFID technology in order track which figurines in the castle the child moves.

Kruppa and colleagues [5] developed a character that is capable of freely moving along the walls of a real room by making use of a steerable projector mounted at the ceiling of the room and a spatial audio system. The character is aware of the user's position and orientation in the room and thus may provide situated advice, but unlike Ritchie it is not able to track the user's pointing gestures in space. Furthermore, the application does not make use of any additional digital overlays apart from the character.

Barakonyi and Schmalstieg [6] present an Augmented Reality framework that integrates work on autonomous agents with work on ubiquitous computing. The framework supports the creation of a large variety of agent-based applications, such as mobile agents embodied by virtual and physical objects as well as agents that are able to migrate from one Augmented Reality application to another. Unlike Ritchie, their agents do, however, not integrate sophisticated conversational skills that include dynamically generated gestures, mimics and speech.

In our earlier work, we made use of a video see-through Head-Mounted Display (HMD) to combine video recordings from the real scene with digital overlays including the virtual character Ritchie [7]. In this Augmented Reality application, Ritchie jointly explores with the user a table-top application that combines virtual buildings of the city center of Augsburg with a real city map being laid out on a real table. The benefit of the approach followed in this paper in comparison to our earlier work lies in the fact that the user may move freely around without having to wear obtrusive equipment or being wired. Furthermore, several users may participate in the installation at the same time taking on the role of an observer or actor.

Most similar to the work described in this paper are applications with agents making use of the mirror paradigm. A very early example is the ALIVE system [8] that incorporates body gestures influencing the behavior of a virtual dog called Silas. In contrast to our setup, virtual objects appear only in front of the video image since occlusion problems between real and virtual objects are not handled. In our application setup, we integrate a real skeleton that gives tactile feedback while the user is positioning virtual organs. Herewith, virtual objects are occluded by the real skeleton and virtual organs cast shadows on real objects in order to enhance a mixed reality impression. A more recent application example that is based on the mirror paradigm includes the Invisible Person [9] which is only visible in the mirror, but not in the real world. In one application, the Invisible Person is acting as a game partner in the TicTacToe game where both the user as well as the Invisible Person may select pads of a game board by moving to a certain position and conducting specific postures. Neither Silas nor the Invisible Person make use of speech and communicative gestures to interact with the human user, however. Cavazza and colleagues [10] employed the mirror paradigm for digital story telling to enable users to participate in a story both as an actor and a spectator. In their application, a video image of the user is integrated in a virtual world while in our case a virtual character is integrated as a digital overlay in the real world. The application by Cavazza and colleagues therefore rather falls in the area of augmented virtuality as opposed to ours which is an augmented reality application. In addition, our approach allows not only for the interaction with digital, but also for the interaction with physical objects.

3 The Virtual Anatomy Assistant

As an example of an educational virtual agent application, we have designed and implemented Ritchie, a Virtual Anatomy Assistant, that helps the user to locate organs of the human body. As shown in Fig. 1 the user stands in front of a backprojection screen where a real-time camera image of the user's physical environment is shown. The attendance of a user is detected by a recognition sensor built on top of the whole installation. Next to the user is a real skeleton. The user's task is to attach virtual organs rendered into the video image to the real skeleton using a tangible pointing device. In Fig. 2 one can see the typical interaction procedure. The user first selects one of the organs from a menu and then moves it to the skeleton position where he or she wishes to place it. In addition, users may explore the virtual organs as part of their physical environment. For example, by rotating the pointing device, they may look at the organs from different sides. The position and rotation of the pointing device is tracked by an optical tracking system called IRTraX and offered by inoptech[1]. It also tracks the position of the skeleton to enable the system to update the transformations of the organs already being placed.

To avoid a possible dominance of the virtual character we have rendered Ritchie smaller than the user on a real socket next to the skeleton. Additionally, the socket helps to avoid occlusion problems between the user and the character.

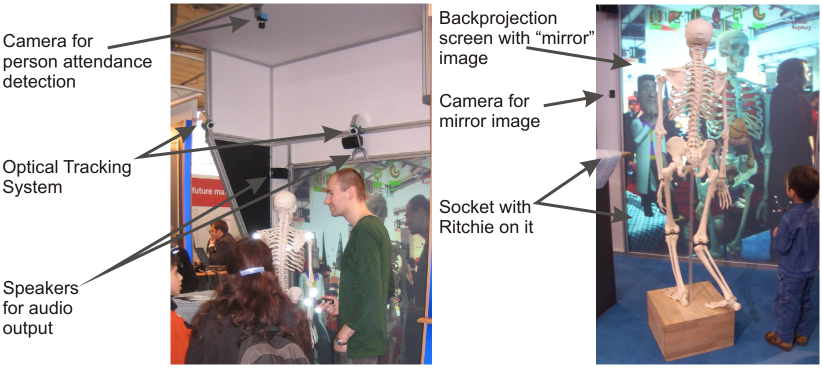
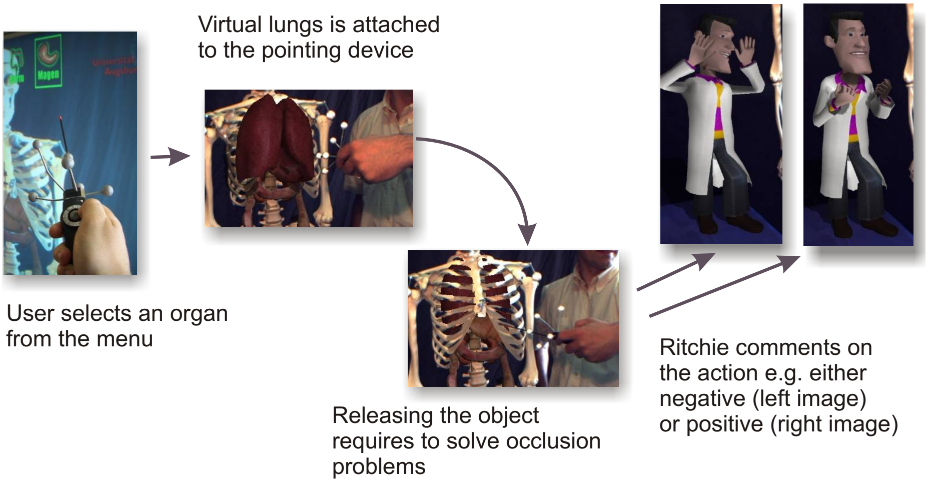

Camera for person attendance detection
Optical Tracking System
Speakers for audio output
Backprojection screen with "mirror" image
Camera for mirror image
Socket with Ritchie on it

Fig. 1. The setup of the Virtual Anatomy Assistant application

[1] http://www.inoptech.com



Virtual lungs is attached to the pointing device

User selects an organ from the menu

Releasing the object requires to solve occlusion problems

Ritchie comments on the action e.g. either negative (left image) or positive (right image)

Fig. 2. A typical interaction between user, skeleton, and the virtual character Ritchie

4 Seamless Integration of Physical and Real Worlds

Augmented Reality should appear to the user as if the virtual and real objects coexisted in the same space. In order to achieve this, a basic process called registration is needed that requires the use of tracking systems. The process of registration is divided into two subtasks. First, a calibration step is needed to estimate static offset and tracking parameters. Calibration can be done offline. Second, the manipulation of objects and typically the user's viewing position have to be tracked in real-time. In the proposed digital mirror setup, the latter is unimportant due to the fact that the user views the real scene from the position of the video camera capturing the real world.

Our digital mirror setup uses an infrared-based stereoscopic tracking system to track the skeleton and a pointing device similar to a mouse with six degrees of freedom. Unfortunately, tracking coordinates are calculated in the coordinates of the stereoscopic tracking system, typically. However, to coincide the virtual objects manipulated by the pointing device with the video image of the projection screen, we need to know the spatial transformation between the video camera and the stereoscopic tracking system. We solved this problem as follows:

1. Calibrate the video camera with the GMLCamera calibration tool.[2]
2. Integrate a system (such as ARToolkitPlus [11]) that recognizes an artificial marker and estimates its pose from the perspective of the video camera.
3. External camera calibration of the stereoscopic tracking system that estimates the relative transformation parameters between infrared cameras.

[2] `http ://research.graphicon.ru/calibration/gml-c++-camera-calibration-toolbox.html`

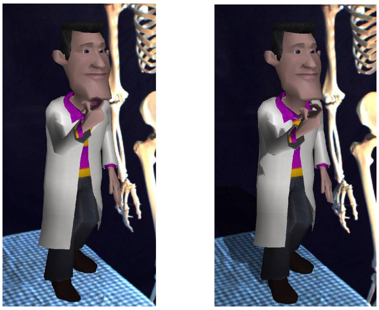


Fig. 3. The integration of virtual shadows on real objects improves the impression of a seamless integration

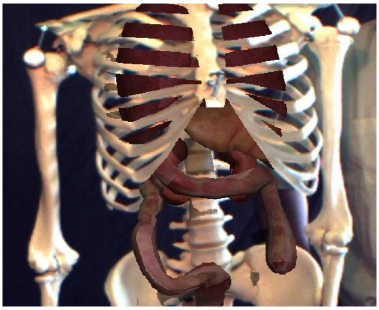


Fig. 4. For a realistic impression of augmented reality virtual objects have to be occluded by real ones

4. Mount an artificial marker perpendicular to the calibration angle of the stereoscopic tracking system [12] such that the relative offset between the two coordinate systems can be easily measured using a ruler.
5. Move this calibration rig into the interaction volume and ensure that there is no line-of sight problem. Then, store pose parameters of the marker and calibration angle. Finally, calculate the world transformation parameters that represent the transformation between objects in the video camera space and the coordinate system of the calibration angle.

As a result of this offline calibration, objects tracked by the infrared-based tracking system are given in the world coordinate system and can be precisely projected onto the video image by using the inverse world transformation matrix.

Apart from rendering the virtual objects and the video image, the graphical renderer has to account for techniques to convey the illusion of an augmented reality. The video image of the real environment is rendered as a full screen background image using a virtual camera model incorporating the parameters of the physical video camera, e.g. focal length, and the principal point. As a consequence, virtual objects are registered properly, however, they appear always in front of the real objects resulting in the shortcoming that the viewer does not have the impression of the virtual objects being integrated into the scene. To overcome this problem and reduce the gap between the real and virtual world, we rely on a technique which is also known under the term “black object rendering”. To this end, we create three dimensional models of some real objects (e.g. the skeleton) that are visible in the video image. The virtual copies are placed in the scene in a way that their screen projection exactly overlaps with the corresponding real objects in the video image. The models are just rendered to the depth buffer and not the color buffer so that they get their color from the video texture but occlude the other virtual objects which are intersecting them. Collisions between the character and the user have been avoided by placing the character on a socket (see Section 3).

To further improve the illusion of a consistent scene, we have the virtual objects cast shadows on the real world objects. To this end, we do not draw the video texture as a simple background full screen quad, but project the video image onto the occlusion objects, modulating the video texture with the shadow intensity of the occlusion models. The resulting shadows give important clues regarding the distance between the virtual objects and the real objects and result in the impression of a better integration.

5 Context Sensors

Situated agents need to take the user's physical context into account. Recent projects equip animated agents with a set of sensors to detect and track people in front of the screen. Examples include kiosk agents, such Mack [13] or Max [14]. For the anatomy assistant, we have included four different sensors that are connected through the Virtual Reality Peripheral Network (VRPN)[3] library:

1. Mouse Sensor: handles button clicks of the pointing device
2. Keyboard Sensor: for administrative purposes
3. Person Detection: indicates the attendance of a user
4. 3D Optical Tracking System: measures the 6DOF of a pointing device and the skeleton.

While the keyboard sensor is only used for administrative purposes (e.g. to reset the application), the person detection and the pointing device has direct influence on the character's behavior. The person attendance detection is realized by a camera that is mounted above the skeleton and was already used in the COHIBIT system described in [15]. If it indicates the presence of a user, it forwards the changed state to ACOSAS that triggers a transition from the "Start Idle Phase" to the "Visitor Arrived" state (see Section 6).

The context data of the tracking system and the mouse button sensor is primarily used for selecting and placing a virtual organ, but secondarily a change of an organ's state also has an influence on the character's comments. The fusion of the real context data with the virtual state of the world is being done by our ACOSAS system (see Section 6).

6 Context-Sensitive Behavior Control

In order to enable a human author to easily specify context-sensitive stories with ECAs, we designed and implemented the ACOSAS (**A** **CO**ntext **S**ensing **A**uthoring **S**ystem) framework. Similar to the approach presented by Gebhard and colleagues [16], ACOSAS models the flow of a reactive story by means of cascaded finite state machines. However, unlike their system, ACOSAS [17] provides specific support for the integration of context data. The fact that ACOSAS is able to access context information from the user's physical environment and to fuse real with virtual context makes it in particular suitable for the realization of the Virtual Anatomy Assistant.

[3] `http://www.cs.unc.edu/Research/vrpn/index.html`

6.1 The ACOSAS Core Engine and the ACOSAS Player

For improved reusability we have split ACOSAS in a core module handling the story and context sensors, and a player application for the connection to the different output modules, such as rendering and text-to-speech. Figure 5 shows the interaction between the ACOSAS Core Engine, the context sensors as well as the ACOSAS player.

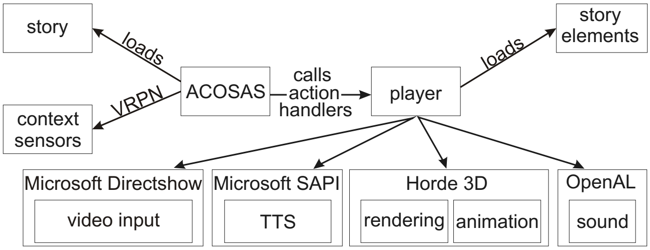


Fig. 5. Module Layout of the ACOSAS Core Engine and the ACOSAS Player: The different output modules are linked by the player application. The ACOSAS Core Engine can control their functionalities by calling registered action handlers.

As input, the ACOSAS Core Engine expects a story file. The header of this file contains links to context sensors as well as a world description file containing the story elements loaded by the player. The body of the story file essentially consists of a specification of the finite state machine representing the story.

The connection between the ACOSAS Core Engine and the ACOSAS Player is established via so-called action handlers that are triggered by the ACOSAS Core Engine during the traversal of the finite state machine. The ACOSAS Player loads story elements, such as the character, the organs and specific sounds, and provides action handlers to manipulate these elements. The main consideration behind the introduction of action handlers was to abstract from the technical details of the underlying output modules.

6.2 Modeling the Story with ACOSAS

The nodes of the finite state machines correspond to scenes while the edges represent transitions between scenes. Figure 6 shows the story nodes created for the Virtual Anatomy Assistant. There are two major story states. The first one is “Start Idle Phase” which refers to a situation when no user is present. In this state, the story loops through different sub nodes in which the character monologizes or performs specific animations to attract visitors. The second one is “Visitor Arrived” which refers to a situation when a user approaches the booth. To ensure that Ritchie is permanently alive, we also included a “Game Idle Phase” in which Ritchie encourages the user to start or continue with the placement task. The utterances used are chosen by a script that takes into account the state of the interaction device. While the user is placing an organ the character would not disturb him or her. The states "User placed organ" and "All organs placed" specify Ritchie's behaviors when commenting on the user's actual performance. By using these fixed states with variable scripted actions we achieve great flexibility. Even if the user responds in an unexpected manner, for example, by starting a conversation with people closeby, the character would not get lost since it always remains in valid predefined states. According to the state machine created for the Virtual Anatomy Assistant, it would then try to get the user's attention back. For the Virtual Anatomy Assistant, the user's perceivable actions are restricted to approaching and leaving the installation and the placement of objects. It is important to note, however, that the ACOSAS system enables us to model character responses to arbitrary user actions including not only vision-based, but also audio-based context sensors. For the european project CALLAS[4], for example, we have developed an emotion recognition sensor based on acoustic features.

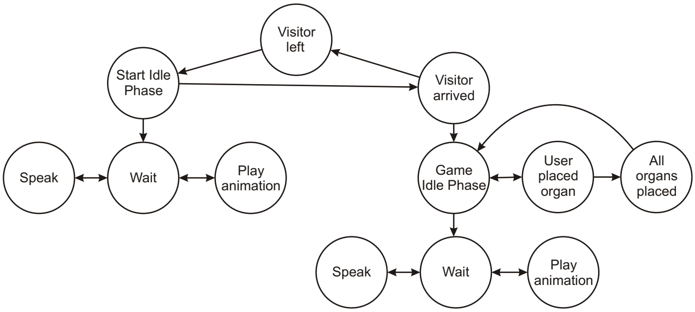


Fig. 6. The Virtual Anatomy Assistant as realized within ACOSAS. In each node various actions for controlling Ritchie are triggered. Some of them have random parameters to offer a greater variety in the character's behavior.

A transition between story nodes is taken if all corresponding conditions are satisfied and all blocking actions (see below) have been finished. Conditions may refer to parameters that are received via context sensors. For example, the transition between the "Start Idle Phase" and the "Visitor arrived" state depends on data received by the person detector. Other possible sources for the parameter values are LUA scripts or just hard-coded constant values. We opted for the choice of LUA [18] as a scripting language since it is employed by a number of commercial games (such as World of Warcraft), it's very fast and also easy to use. Another advantage is the fact that values or actions provided by LUA functions can be changed on the fly without restarting the application.

For each story node, we may specify actions that will be triggered if the corresponding node is entered, left or interrupted. In the latter case, there are actions that were activated when entering the node, but not yet finished when

[4] http://www.callas-newmedia.eu/

leaving it. In the Virtual Anatomy Assistant, such a situation may occur when Ritchie is waiting for visitors. If the person attendance detector signals that a user has arrived, the current animation of the character has to be interrupted and the "Visitor arrived" node is entered. To interrupt a story node, we may, for example, specify an action that ends the character's current animation by fading the animation to a neutral state. Next to the predefined actions implemented in the player, the designer can also create more complex ones by using the scripting language that can access various LUA binding functions (e.g. accessing context values) provided by the ACOSAS core module and the player.

6.3 Behavior Output

As mentioned in Section 6.1 the ACOSAS Player has to implement the handlers for actions triggered by the ACOSAS Core Engine and the connection to the output libraries, such as graphics, sound and text-to-speech.

For the graphical duties, we are using the Horde3D rendering engine[5] developed at our lab. Among other things it supports morph targets that can be easily used for the synchronisation of synthesized speech based on a text-to-speech library, such as the Microsoft Speech API, and the virtual agent's lip movements.

A morph target defines the final positions of some vertices. Beginning from a base pose, the vertex positions are linearly interpolated. By setting the interpolation factors it is possible to control the strength of a facial expression and to mix several targets. Our implementation uses a morph target system to fade between the previous and current viseme of the text to be spoken.

Another challenge is the animation of the character's body. Our objective was to create convincing movements of the character and smooth transitions between separate animation clips. The movements of the character were realized using a skeletal animation system. The animations were created with a 3D modeling package and applied directly to the model.

To enable smooth transitions from one animation to another, we make use of a technique called animation blending. With our system, it is possible to put several animation clips on discrete channels. A weight for each channel determines how much the associated animation contributes to the final result relative to the other channels. To perform a transition between two channels, we do blending by fading one channel towards zero and the other towards one.

Another important requirement is the possibility to play several animations at the same time. This is called animation mixing. In our case, we have a rather long subtle idle animation where Ritchie performs slight movements of his head, blinks with its eyes and breathes. This animation is played by using an endless action and has to be combined with some randomized and casual gestures. To achieve this effect, we use additive animations. We calculate the difference in translation and rotation of the gesture animation relative to the character's standard pose. This difference is added to the base animation which is in our case the idle clip.

[5] http://www.nextgen-engine.net/

Therewith we have a simple way to combine arbitrary independent gestures with very satisfying visual quality and a minimum amount of effort.

7 Evaluation

The Virtual Anatomy Assistant was evaluated at CeBIT 2007, the world's largest annual trade show for information and telecommunication technology.

7.1 Method

Over a period of 8 days, hundreds of people interacted with the Virtual Anatomy Assistant out of which 71 (56 male and 15 female) participated in the evaluation and filled in the questionnaire. All subjects were native speakers of German. 19 participants were under 18 years old, 52 over 18 years old.

All participants were given an individual introduction to the system. In particular, they were shown how to select menu options using the pointing device and how to place organs in the skeleton. After that, they had the chance to try out the system themselves. On average, the experiment took about 10 minutes per participant. After the experiment, each participant filled in an anonymous post-questionnaire.

The post-questionnaire used 14 attitude statements with a 5-ary rating scale (disagree, somewhat disagree, neutral, somewhat agree, agree) to evaluate how the participants perceived the interaction with the system (9 questions) and the behavior of the agent (5 questions).

7.2 Results for the Interface

To show that the mean value of a rating was significantly above or below the neutral value of 3, we applied t-tests for one sample. Overall, the interface was perceived as positive. The participants thought it was illustrative to place organs in the real skeleton with a mean value of 4.48 (t(70)=13.723, p<0.001), they found it not too difficult to concentrate both on the real skeleton and the presentation on the screen with a mean value of 2.61 (t(70)=-2.293, p<0.03), and they did not find the interface confusing with a mean value of 1.92 (t(70)=-8.469, p<0.001). The participants found it easy to select menu options with a mean value of 3.92 (t(70)=7.633, p<0.001). They found it less easy (mean 3.32) to place the organs at the wanted position (t(70)=2.529, p<0.02).

We did not find any empirical evidence that the females experienced the interaction with the system differently than the males. We noticed, however, that different age groups (i.e. participants under 18 years old and participants over 18 years old) responded differently to the mirror metaphor. In particular, we compared the responses of participants under 18 years old with those of participants over 18 years old. The average age of the first age group was 13.63. The youngest participant in this group was 10 and the oldest 17. For the second age group, we did not record the exact age because we assumed that people were reluctant to give that information.

Table 1. Mean Ratings for Age-Specific Perception of the Interface

Mean Scores	Children	Adults
Interface was illustrative	4.84	4.35
Reminded me of a mirror	3.95	2.87
Found it helpful to see myself on the screen	3.79	2.65
Overlooked that I was visible on the screen	1.74	2.73

Applying a two-tailed t-test to the two age groups showed that there were significant differences between the two age groups for 4 out of 9 statements. Overall, the younger age group developed a better understanding of the mirror metaphor and gave it a more positive rating (see Table 1). In particular, the younger participants found it more illustrative to place organs in the real skeleton (t(69)=2.086, p<0.05) than the older participants, they felt rather reminded of a mirror (t(69)=2.929, p<0.006) than the older participants, they found it more helpful to see themselves on the screen (t(69)=3.227, p<0.003) than the older participants , and they were less likely to overlook that they were visible on the screen than the older participants (t(69)=-2.391, p<0.03).

7.3 Results for the Character

Overall, the results of the experiment indicate that the participants perceived the character as intended. It was rated above mean in regard to its entertaining value with a mean value of 3.92 (t(70) = 6.284, p<0.001). Furthermore, the participants had the impression that the character was aware of them with a mean value of 4.23 (t(70) = 10.936, p<0.001). Even though the character was permanently talking, it was not regarded as distracting with a mean value of 1.83 (t(70) = -10.141, p<0.001). Nevertheless, the participants did not consider Ritchie as superiorly helpful. As a reason, we indicate that the character commented on the actions by the participants, but it did not provide any hints regarding where to place the objects. Its main purpose was to encourage the participants to start with their task and to entertain them by witty comments. We also thought Ritchie's special kind of humor would make the participants feel less embarrassed about making mistakes. However, this could not be confirmed by our experiments.

Interestingly, we did not find any significant gender or age differences regarding the perception of the character. The results regarding gender are in line with the results obtained by Kipp and colleagues [15] for the Autostadt scenario. For adults, we observed, however, a medium correlation between the participants' ratings of the task and their ratings of the character which could not be attested for the younger age group. The easiest the placement task was for the adults, the more they appreciated Ritchie's entertaining value (Pearson Product-Moment Correlation: r=0.501; p<0.001). Obviously, the adults had more sense for Ritchie's humor when figuring out less difficulties with the task. For young people under 18, there was no significant correlation between the two variables.

We ascribe this effect to the heterogeneous composition of this age group. Especially, younger children might have problems with Ritchie's sarcastic nature which probably had a stronger influence on their ratings than the difficulty of task.

8 Conclusion

In this paper, we reported on our efforts to integrate an embodied conversational agent into the user's real world. The technological contributions of this paper include (1) the combination of sophisticated tracking systems with state-of-the art rendering techniques to achieve a seamless integration of the physical and the real world and (2) a module for context-aware behavior control that enables spontaneous character responses to a user's actions in the physical world which may even include interruptions of a character's ongoing actions. To illustrate our ideas, we presented the Virtual Anatomy Assistant Ritchie which is based on these technologies. An evaluation of Ritchie confirmed that it contributed to a positive perception of the interaction experience by its entertaining value without being distracting. The participants had the impression that the character was aware of their presence and noticed what they were doing. Furthermore, the users found it illustrative to place virtual objects in a real skeleton.

Future work will concentrate on the conduction of a controlled experiment in order to shed light on the benefits of augmented realities for spatial learning tasks. Furthermore, we will equip the virtual character with a component for spatial reasoning so that it will be able to provide hints to the users regarding the placement of objects.

Acknowledgments. This work was partly funded by the European Commission, project IST-2005-2.5.7 Callas.

References

1. Azuma, R.T.: A survey of augmented reality. Presence: Teleoperators and Virtual Environments 6, 355–385 (1997)
2. Shelton, B.E., Hedley, N.R.: Exploring a cognitive basis for learning spatial relationships with augmented reality. Technology, Instruction, Cognition and Learning 1, 323–357, 154 (2002)
3. van Mulken, S., André, E., Müller, J.: The persona effect: How substantial is it? In: Proceedings of the Thirteenth Conference of the British Computer Society Human Computer Interaction Specialist Group - People and Computers XIII, Sheffield, pp. 53–66 (1998)
4. Cassell, J., Ananny, M., Basu, A., Bickmore, T., Chong, P., Mellis, D., Ryokai, K., Smith, J., Vilhjálmsson, H., Yan, H.: Shared reality: physical collaboration with a virtual peer. In: CHI '00 extended abstracts on Human factors in computing systems, pp. 259–260. ACM Press, New York, USA (2000)
5. Kruppa, M., Spassova, M., Schmitz, M.: The virtual room inhabitant - intuitive interaction with intelligent environments. In: Zhang, S., Jarvis, R. (eds.) AI 2005. LNCS (LNAI), vol. 3809, Springer, Heidelberg (2005)

6. Barakonyi, I., Schmalstieg, D.: Ubiquitous animated agents for augmented reality. In: Proceedings of IEEE and ACM International Symposium on Mixed and Augmented Reality (ISMAR'06), IEEE Computer Society Press, Los Alamitos (2006)
7. Dorfmüller-Ulhaas, K., André, E.: The synthetic character ritchie: First steps towards a virtual companion for mixed reality. In: Proceedings of IEEE International Symposium on Mixed and Augmented Reality (ISMAR'05), IEEE Computer Society Press, Los Alamitos (2005)
8. Maes, P., Darrell, T., Blumberg, B., Pentland, A.: The alive system: full-body interaction with autonomous agents. Computer Animation, 11 (1995)
9. Psik, T., Matković, K., Sainitzer, R., Petta, P., Szalavari, Z.: The invisible person: advanced interaction using an embedded interface. In: EGVE '03: Proceedings of the workshop on Virtual environments 2003, pp. 29–37. ACM Press, New York, USA (2003)
10. Cavazza, M., Charles, F., Mead, S.J., Martin, O., Marichal, X., Nandi, A.: Multimodal acting in mixed reality interactive storytelling. IEEE-Multimedia 11, 30–39 (2004)
11. Wagner, D., Schmalstieg, D.: ARToolKitPlus for pose tracking on mobile devices. In: Proceedings of 12th Computer Vision Winter Workshop (CVWW'07) (2007)
12. Dorfmüller-Ulhaas, K.: Optical Tracking - From User Motion To 3D Interaction. PhD thesis, Institut 186 für Computergraphik und Algorithmen, Vienna (2002)
13. Nakano, Y.I., Reinstein, G., Stocky, T., Cassell, J.: Towards a model of face-to-face grounding. In: Dignum, F.P.M. (ed.) ACL 2003. LNCS (LNAI), vol. 2922, pp. 553–561. Springer, Heidelberg (2004)
14. Kopp, S., Jung, B., Leßmann, N., Wachsmuth, I.: Max - a multimodal assistant in virtual reality construction. Künstliche Intelligenz 17, 11–17 (2003)
15. Kipp, M., Kipp, K.H., Ndiaye, A., Gebhard, P.: Evaluating the tangible interface and virtual characters in the interactive cohibit exhibit. In: Gratch, J., Young, M., Aylett, R., Ballin, D., Olivier, P. (eds.) IVA 2006. LNCS (LNAI), vol. 4133, pp. 434–444. Springer, Heidelberg (2006)
16. Gebhard, P., Kipp, M., Klesen, M., Rist, T.: Authoring scenes for adaptive, interactive performances. In: AAMAS '03: Proceedings of the second international joint conference on Autonomous agents and multiagent systems, pp. 725–732. ACM Press, New York, USA (2003)
17. Erdmann, D., Dorfmüller-Ulhaas, K., André, E.: Integrating VR-authoring and context sensing: Towards the creation of context-aware stories. In: Göbel, S., Malkewitz, R., Iurgel, I. (eds.) TIDSE 2006. LNCS, vol. 4326, pp. 151–162. Springer, Heidelberg (2006)
18. Ierusalimschy, R., de Figueiredo, L.H., Filho, W.C.: LUA — an extensible extension language. Software Practice and Experience 26, 635–652 (1996)