The Affective Virtual Patient: An E-Learning Tool for Social Interaction Training within the Medical Field

Bernhard Jung, Abdul Ahad, Matthias Weber
ISNM International School of New Media
University of Lübeck, Germany
www.isnm.de

Abstract. Training of social interaction skills, particularly for emotionally charged situations is becoming increasingly critical in many professions. In cases such as the medical field, however, conventional real-world or role-playing training is extremely limited or even impossible. An example for this is the treatment of injured children. Therefore, we have developed an e-learning system that uses virtual agent technology to simulate the treatment of and human interaction with a hurt, upset boy and his emotional, easily agitated mother. Inconsiderate behaviors such as asking for insurance cards at the wrong moment, inappropriate joking, or addressing certain questions at the wrong person may contribute to the deterioration of the situation, hinder procedural progress, and lead to complaints about the medical personnel. The PC-based system employs Haptek virtual character technology allowing for highly emotional facial expression and lip-synchronized speech.

1. Motivation

Good command of social interaction skills constitutes a highly critical factor in many professional contexts, particularly in disciplines where there is close contact with people. A typical example is the medical field. Here, ignorance of the social/emotional context of the patient treatment situation can severely complicate diagnosis and therapy processes as the acceptance of medical procedures hinges on doctors' social competence and communication skills. Furthermore, it is estimated that over 70% of malpractice suits against doctors involve a history of crudely flawed communications between doctors and their patients.

To date, however, training of social interaction skills plays only a minor part, if at all, in education of doctors and other medical personnel. Classical textbooks and conventional e-learning tools cannot adequately convey the social and emotional context of procedures performed with real people. Role-playing with real actors is usually too expensive, or in the case of children often not possible at all. Recent advances in 3D computer graphics hardware and animation methods however make the development of novel e-learning tools feasible that build on the inclusion of realistically animated virtual agents. Modern virtual agent technology facilitates e.g. the real-time animation of highly emotional, convincing facial expression and the generation of lip-synchronized speech. Our general idea is to exploit virtual agent technology for the development of a new generation of e-learning tools that support the training of social interaction skills of all kinds of professionals who are frequently exposed to emotionally critical situations. The particular goal of the "Affective Virtual Patient" project is to provide a computer-based training tool that prepares medical students not only to solve a medical problem but also to take into account the social / emotional context of patient treatment situations.

In cooperation with the clinic for pediatric surgery of the Universitätsklinikum Schleswig-Holstein, Campus Lübeck, a prototypical e-learning system for social interaction training called "Affective Virtual Patient" was developed. The scenario models a quite typical situation for pediatric surgery and simulates the interaction with a young, injured, and upset boy and his nervous, easily agitated mother. In comparison to the treatment of adult patients, the scenario places additional demands on the doctors' social interaction skills as clear differences exist between the interaction with adults and children. Also, accompanying parents usually react immediately and in a highly sensitive manner to any interaction with their children. In our prototypical e-learning system, the user (medical student) has to work through a number of individual scenes from admission, over treatment, to check-out of the patient. In each scene, the user must interact with the virtual agents representing the injured boy and his mother. The virtual agents react promptly and change their affective behavior in response to user interactions. For example, asking for insurance cards at the wrong moment, inappropriate joking, or addressing certain questions at the wrong person all cause the agents to become more angry or upset. Ongoing, real-time facial animation effectively conveys the agents' emotional states to the user.

The remainder of this contribution is structured as follows: The next section reviews the state of the art in virtual agent technology; a focus is placed on "embodied conversational agents" and the techniques that are used to build such agents. Section 3 gives a fly-over of our prototype system, the "Affective Virtual Patient". Besides sample screens and user interactions, the overall system architecture is explained. In Section 4, we describe the process of how training scenes for this system can be authored in an XML-based modeling language. Finally we conclude and comment on the future of virtual agents in social interaction training.

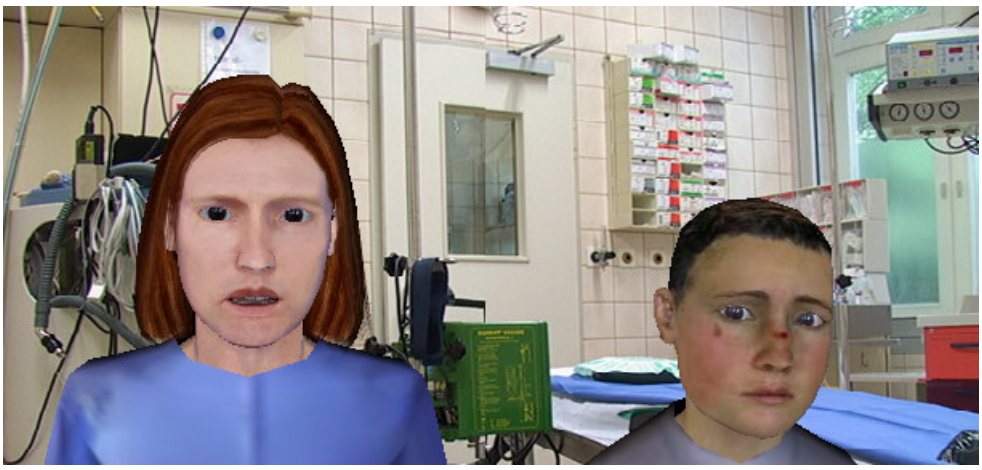

Figure 1: The setting: An upset mother with her injured child enters the clinic

2. Virtual Agents – Enabling Technologies and Related Work

Communication between humans happens on many channels. As e.g.. Mehrabian (1968) points out facial expression, body language, and gestures play a very important role in human communication. The research fields of "Intelligent Virtual Agents" (e.g. Rist et al., 2003; Gratch et al., 2002; Prendinger & Ishizuka, 2003) or "Embodied Conversational Agents" (Cassell et al., 2000) examine interactive 3D computer graphics-based simulated humans that can be animated in real time and possess capabilities for verbal and nonverbal communication. A goal is that these agents should act in a believable way and possess human-like, social interaction capabilities (e.g. Bickmore & Cassell, 2004).

Virtual Agents for Social Interaction Training are an extension to these embodied conversational agents. These agents are used in e-learning applications that aim at improving the social interaction skills of their users. Social interaction is "the process by which we act and react to those around us" (Giddens et al. 2003) and involves both verbal and non-verbal communication. Non-verbal communication is accomplished via several communication channels like body language, gesture, and facial expression. Facial expression can be highly indicative of someone's emotional state and thus is of particular importance in social interaction (Ekman 1993, 1999). Ekman (1993) has identified a set of basic emotions that can be communicated via facial expression; his system of facial features is commonly used as basis for many computer animations of facial display. Computer models of the virtual agents' underlying emotional states often represent emotions as a points or regions in a multi-dimensional space, building e.g. on the work of Wundt (1911); the OCC model conceptualizes emotions as valenced reactions e.g. to external events and other agents' actions and is also commonly used in virtual agent implementations (Ortony et al., 1998). See (Trappl et al., 2003) for an overview of modeling of emotion and affect in virtual agents.

Another important part for virtual agent systems, especially in training applications, concerns the story-telling (e.g. Ponder et al, 2003). The story has to consider many eventualities, e.g. which interaction possibilities exist for the user in a given situation and what transitions between the individual scenes can occur. Further, if affective feedback is important then the course of the story will influence how the virtual agents "feel" after certain actions by the user, e.g. do they feel understood or not and therefore become more friendly / satisfied or angry.

Enabling technologies used to achieve the goal of believable, communicative, and socially aware virtual agents thus include at least:

- real-time face and body animation,
- speech-recognition or at least textual or multi-choice input,
- speech-generation with so called text-to-speech (TTS) systems or playback capability of prerecorded speech, combined with lip-synchronous animation of speech,
- an emotion simulation that determines the virtual agent's affective state in the current situation, e.g. how angry / afraid / happy the agent is; and
- a story telling component.

For most of these tasks, many commercial and sometimes free open source tools are available on the market. In our system, facial-animation and lip-synchronous facial animation is achieved with Haptek technology (www.haptek.com). Haptek's PeoplePutty is used for modeling of the virtual agents' appearance. The open source library Cal3d can be used for full body animation (cal3d.sourceforge.net). For some tasks, such as simulation of emotion dynamics and story telling, customized solutions were implemented. Section 3 describes the software architecture of our prototype system in more detail.

Related virtual agent systems concerned with patient simulations and training applications include amongst others the following: The "Virtual Patient" (Kiss et al. 2004) system and the virtual agent system described by (Kizakevich et al., 2003) are used for training of medical procedures but are not primarily concerned with aspects of doctor to patient dialogue. The "pedagogical" agent STEVE is a virtual reality based interactive training system for maintenance tasks (Rickel & Johnson, 1999). Its successor system MRE uses virtual agents for training of decision-making processes in military contexts (Hill et al., 2003). Also, a few other virtual agent systems have been developed for purposes of social interaction training: (Hubal et al., 2003) have developed a system for training the social

interaction skills of police officers. An earlier system (Hubal et. al 2000) is closest to our application area and aims at improving the interaction skills of doctors when interviewing patients. Our system combines, however, aspects for training of both medical procedures and social interaction skills (the focus is on the latter). It is also a Windows based application that builds on many standard components, supports extensibility to alternative scenarios, and is easily installed on any modern PC.

3. Affective Virtual Patient: System Walk-Through & Architecture

The "Affective Virtual Patient" is an e-learning system for training medical students and doctors in social interaction with their patients. Patients are represented via socially aware, conversational virtual agents. The agents can display very expressive emotions and produce speech in a lip-synchronous fashion. The scenario involves the simulated treatment of an injured, upset boy who enters pediatrician's clinic with his nervous, easily agitated mother (see Figure 2). The system is however highly configurable and can easily be extended to different application scenarios (see Section 4).

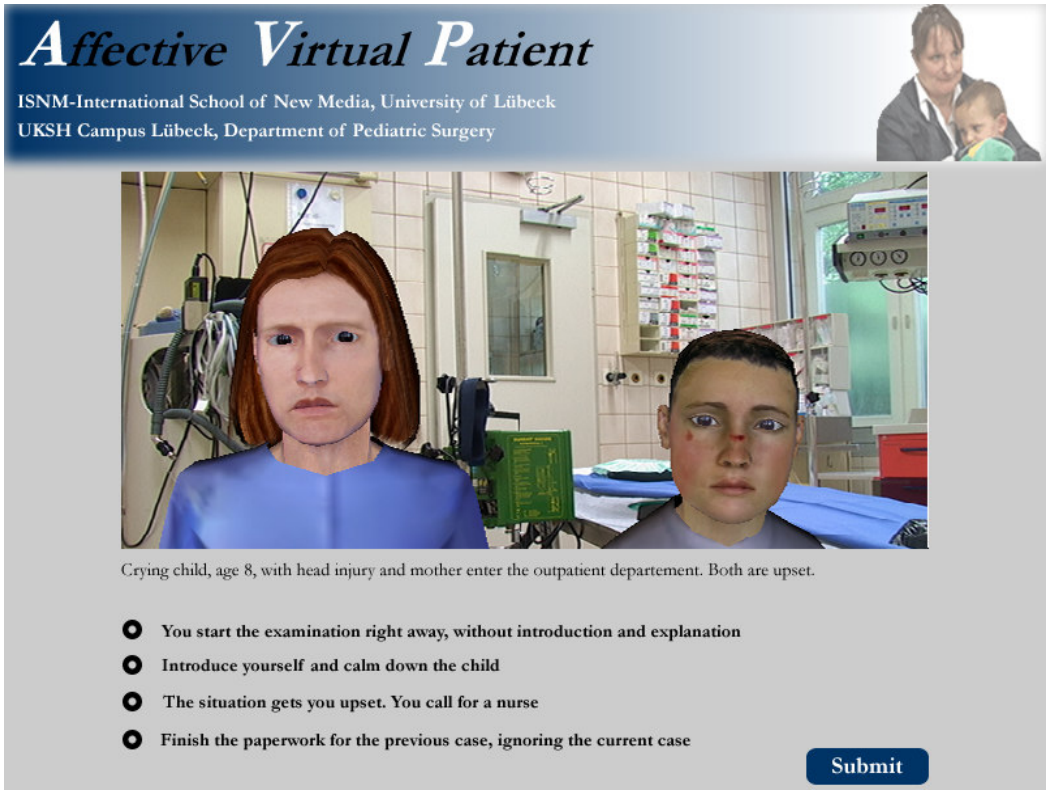

Figure 2: Initial scene of the Affective Virtual Patient system

The present application consists of 6 scenes. Each scene starts with a short description of the current situation and task and provides the user with 3 to 5 interaction choices. For example, Figure 2 shows the initial scene where the mother straightaway starts complaining about the irresponsible behavior and unavailability of doctors to set the tone for an intense learning experience. If an interaction choice is selected, the virtual agents respond through the playback of prerecorded speech segments. If the correct choice was taken, the story is advanced to the next state. Otherwise, the story will remain in the same scene but with the virtual agents recognizably more upset. Animation of the virtual agents is continuous, i.e. also

occurs in the absence of user interactions. Further, short video clips with real-world hospital footage are shown between the individual scenes. Upon completion of the whole story, an evaluation score is presented that reflects the number of mistakes made by the user.

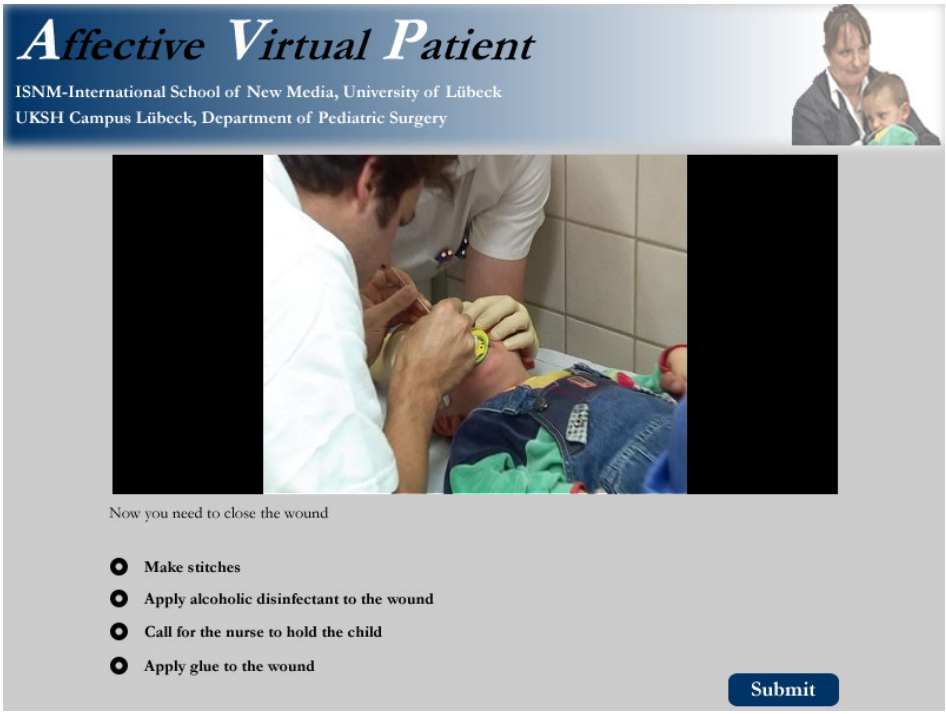

Figure 3: Video footage from the hospital is displayed as transition between scenes

The system architecture of the Affective Virtual Patient is composed of three main components, which are:

- XML processing to read in the story book from an external file
- User Interface and application logic
- Media components (audio, video, character animation)

The core software module that contains the application logic and generates the user interface is written in Visual Basic.Net. It is a Windows based desktop application that uses .Net Framework technologies and provides the support to integrate components for both Windows media player SDK and Haptek ActiveX. The Windows media player SDK (www.microsoft.com) gives the opportunity of playing media elements like videos and audio while the Haptek ActiveX component (www.haptek.com) controls character appearance, movement, and facial animation.

Figure 4 shows the system architecture. In this architecture, the central .Net module implements both the application logic and the user interface. It receives the XML data elements, which represent all facets of the story (for details on authoring these data elements see Section 4). An appropriate data structure to store these elements in the application is provided by the central .Net module. Once stored, it dynamically generates the Windows form application from the XML descriptions and plays the corresponding media elements. As the complete story is defined in an external XML file, the core system can be used with different languages, visualize as many scenes as required, and incorporate a flexible number of dialogue options.

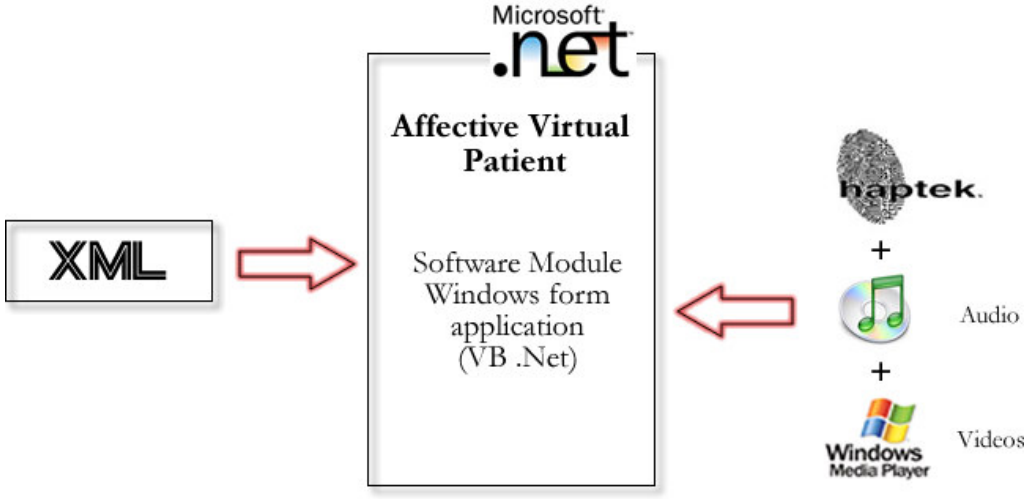

Figure 4: System Architecture of Affective Virtual Patient

4. Authoring

A goal of the "Affective Virtual Patient" was to develop a generic platform that facilitates the easy extension to different application areas of virtual agents for social interaction training. The authoring process of such applications involves the production of various media files (3D characters, video, audio) as well as the definition of a "story" composed of the various scenes and interaction choices for the user; an XML-based format was designed to model such stories (cf. the left and right components of the system architecture in Figure 4).

Affective Virtual Patient (AVP) makes use of XML technology in order to create a dynamic and easily extensible system. An external XML document contains a storyboard composed of dialogues, scene descriptions, emotion values, media / visual aid references and the navigator, which advances the application state from one scene to the next.

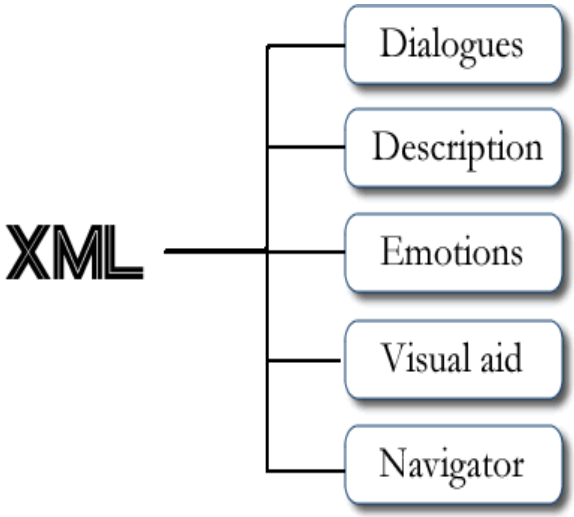

Figure 5: XML container for AVP

The dialogue structure or "story" of the Affective Virtual Patient is defined by means of an XML document. Such a document contains exactly one story, denoted by the XML element <*story*>. Documents are composed of multiple scenes. Each scene is defined within the <*scene*> element that also contains an attribute "*id*" which denotes the scene number. Descriptions for scenes are created with the <*description*> element. The <*option*> element

finally does the main trick; it has an attribute "*type*" which denotes whether this option is right or wrong. For each option, a successor scene and influences on emotions can be described. This can be done via two other elements contained in <*option*> elements:

- ***moveto:*** Acts like a navigator by defining the next scene that results from a particular user interaction.
- ***emotion:*** Holds the virtual agent's name and the effect of a particular user interaction on the agent's mood (i.e. does the agent become happier or more upset/frustrated when the user chooses an particular option?).

```
<story>
<scene id="1">

    <description>Crying child, age 8, with head injury and mother enter the outpatient departement.</description>

     <option type="wrong">You start the examination right away, without introduction and explanation
           <moveto next="1">
             <emotion agent="mother" change="-2"/>
             <emotion agent="child" change="-2"/>
           </moveto>
     </option>

     <option type="correct">Introduce yourself, calm down the child
           <moveto next="2">
             <emotion agent="mother" change="+2"/>
             <emotion agent="child" change="+2"/>
             <videoplay name="moviel.avi"/>
           </moveto>
     </option>
          .
          .
          .
</scene>
  .
  .
</story>
```

Figure 6: XML-based story description

Visual aids like videos and audios provide an additional amount of multimedia to our e-learning system. Playing short video clips in between the scenes enhances the learning experience and helps to visualize the real-world situation better. For each option, the XML file may contains a <*videoplay*> element that defines which video to play when the option is selected.

To author a new application, besides defining the story, it is necessary to produce appropriate characters, audio files, and videos. Character modeling is accomplished using Haptek's PeoplePutty (www.haptek.com), a tool that provides the capability of creating interactive and highly emotional 3D characters. PeoplePutty can also be used to generate lip-synchronous audio files. In cooperation with the Universitätsklinikum Schleswig-Holstein, Campus Lübeck two actors were asked to perform the role of mother and doctor in our predefined story. While acting the predefined story with all it's possibilities their voices were recorded in a .wav format file. These recordings were passed to a tool supplied with People Putty, which converts these .wav files into an .ogg format that is enriched with information for lip-synchronous character animation and speech playback.

5. Conclusion

We have described the "Affective Virtual Patient", an e-learning system that makes use of virtual agent technology to better prepare medical students for emotionally charged patient treatment situations. The virtual agents are continuously animated and, in particular, are able to convey their current affective state via highly emotional facial expressions. I.e. by using virtual agent technology, the social / emotional context of the doctor-patient interaction – which is hard to represent in classical textbooks or conventional e-learning systems - is naturally reflected in the learning situation. While a rigorous empirical evaluation of the system's effectiveness is part of future work, initial feedback suggests that the system clearly raises awareness of the importance of social/emotional context in doctor-patient dialogue and thus serves to motivate students to improve their social interaction skills.

Many potential application areas for social interaction training using virtual agents can be imagined, from training of doctors, police, military, and call-center workers to the general promotion of intercultural understanding. Of particular interest are all kinds of emotionally charged situations where one's social interaction skills make the difference between an escalating and de-escalating course of affairs. By building on standard software components and support for XML-based authoring, the Affective Virtual Patient system was designed as a general platform that can easily be extended to such alternative scenarios of virtual agents in social interaction training.

Acknowledgments

The authors want to thank Prof. Lucas Wessel and Frank Albrecht, both clinic of pediatric surgery, Universitätsklinikum Schleswig-Holstein, Campus Lübeck for their help in drafting the Affective Virtual Patient's scenario, for their support with the video production, and the audio recording. Thanks also to Panayiotis Petropoulos for his video recordings from the hospital.

References

T. Bickmore and J. Cassell. Social Dialogue with Embodied Conversational Agents. In J. van Kuppevelt, L. Dybkjaer and N. Bernsen (eds.) *Natural Intelligent and Effective Interaction with Multimodal Dialogue Systems.* New York: Kluwer Academic, 2004.

J. Cassell, J. Sullivan, S. Prevost, and E. Churchill. *Embodied Conversational Agents*, Cambridge: MIT Press. 2000.

P. Ekman. Facial Expression and Emotion. *American Psychologist*, 48(4), pp. 384-392, 1993.

P. Ekman. Basic Emotions. In T. Dalgleish and T. Power (eds.) *The Handbook of Cognition and Emotion*, pp. 45-60. Sussex, U.K.: John Wiley & Sons, Ltd. 1999.

J. Gratch, J. Rickel, E. André, J. Cassell, E. Petajan and N. Badler. Creating Interactive Virtual Humans: Some Assembly Required. *IEEE Intelligent Systems, 17 (4)*, pp. 54-63. 2002.

A. Giddens, M. Duneier and R. P. Appelbaum. Introduction to Sociology. 4th edition, 2003.

R. W. Hill, J. Gratch, S. Marsella, J. Rickel, W. Swartout, D. Traum. Virtual Humans in the Mission Rehearsal Exercise System. *KI – Künstliche Intelligenz, 4/03*, pp. 5-10, 2003.

R. C. Hubal, P. N. Kizakevich, C. I. Guinn, K. D. Merino and S. L. West. The Virtual Standardized Patient – Simulated Patient-Practitioner Dialog for Patient Interview

Training. In J.D. Westwood, H.M. Hoffman, G.T. Mogle, D. Stredney and R.A. Robb (eds.) *Medicine Meets Virtual Reality 2000: Envisioning Healing: Interactive Technology and the Patient-Practitioner Dialogue*, IOS Press, Amsterdam, 2000.

R. C. Hubal, G. A. Frank and C. I. Guinn. Lessons learned in modeling schizophrenic and depressed responsive virtual humans for training. *Proceedings of the 8th International Conference on Intelligent User Interfaces*, pp 85-92. ACM Press 2003.

B. Kiss, B. Benedek, G. Szijártó, G. Csukly, L. Simon and B. Takács. Virtual Patient: a Photo-real Virtual Human for VR-based Therapy. *Medicine Meets Virtual Reality 2004*. Long Beach, CA USA, 2004.

P. N. Kizakevich, L. Lux, S. Duncan, C. Guinn. Virtual Simulated Patients for Bioterrorism Preparedness Training. *Medicine Meets Virtual Reality 2003*, pp. 165-167, IOS Press, 2003.

A. Mehrabian. Communication without words. *Psychology Today*, 2(4), 1968.

A. Ortony, G.L. Clore, and A. Collins. *The Cognitive Structure of Emotions*, Cambridge Univ. Press, Cambridge, UK, 1988.

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. *9th Eurographics Workshop on Virtual Environments IPT/EGVE 2003*, pp. 97-106. Zurich, Switzerland, 2003.

H. Prendinger & M. Ishizuka (Eds.). *Life-Like Characters – Tools, Affective Functions*, a*nd Applications*. Springer. 2003.

J. Rickel and W. L. Johnson. Animated agents for procedural training in virtual reality: Perception, cognition, and motor control. *Applied Artifcial Intelligence*, 13:343-382, 1999.

T. Rist, R. Aylett, D. Ballin and J. Rickel (eds). *Intelligent Virtual Agents. 4th International Workshop, IVA 2003*. Springer, LNAI 2792, 2003.

R. Trappl, S. Payr, P. Petta. *Emotions in Humans and Artifacts*. MIT Press, 2004.

W. Wundt. *Grundzüge der physiologischen Psychologie*. Leipzig: Engelmann, 1911.