

CrossTalk: An Interactive Installation with Animated Presentation Agents

Thomas Rist, Stephan Baldes, Patrick Gebhard, Michael Kipp,
Martin Klesen, Peter Rist, Markus Schmitt

DFKI GmbH, Stuhlsatzenhausweg 3,
D-66123 Saarbrücken, Germany
+49 (0) 681 302 5266
email: crosstalk@dfki.de

ABSTRACT

In this paper, we describe CrossTalk, an interactive installation in which the virtual fair hostess Cyberella presents and explains the idea of simulated dialogues among animated agents to present product information. In particular, Cyberella introduces two further virtual agents, Tina and Ritchie who engage in a car-sales dialogue. Cyberella on the one hand, and Tina and Ritchie on the other hand live on two physically separated screens which are spatially arranged as to form a triangle with the user. The name "CrossTalk" underlines the fact that different animated agents have cross-screen conversations amongst themselves. From the point of view of information presentation CrossTalk explores a meta-theater metaphor that let agents live beyond the actual presentation, as professional actors, enriching the interactive experience of the user with unexpected intermezzi and rehearsal periods. CrossTalk is designed as an interactive installation for public spaces, such as an exhibition, a trade fair, or a kiosk space.

Categories and Subject Descriptors

J.5 [Arts and Humanities]: Performing arts (with virtual actors).

General Terms

Design, Experimentation

Keywords

Virtual conversational characters, interactive installation

1. BACKGROUND

The last decade has seen a general trend in HCI to make human-

computer dialogue more like human-human dialogue. Computers are ever less viewed as tools and ever more as partners or assistants to whom tasks may be delegated. Trying to imitate the skills of human presenters, some R&D projects have begun to deploy animated agents (or characters) in wide range of different application areas including e-Commerce, entertainment, personal assistants, training / electronic learning environments [4, 9]. Based either on cartoon drawings, recorded video images of persons, or 3D body models, such agents provide a promising option for interface development as they allow us to draw on communication and interaction styles with which humans are already familiar.



Figure 1. The presentation agent Cyberella.

1.1 Cyberella

Starting with the development of the so-called PPP Persona presentation agent [14] back in 1994, our group has designed a number of animated conversational characters for a variety of different application tasks, including Cyberella (cf. Figure 1), a female synthetic conversational character working as a receptionist [6]. A user can engage with Cyberella in a typical receptionist conversation, e.g., by asking her about directions how to get to the office of a certain staff member. However, Cyberella

is an affective character and may well react emotionally to a visitor's utterances [2, 6].

1.2 Using Multiple Characters

Most of the current applications assume settings in which the agent addresses the user directly as if it were a face-to-face conversation between human beings [4]. Such a setting seems quite appropriate for a number of applications that draw on a distinguished agent-user relationship. For example, an agent may serve as a personal tutor or as a guide as in the case of Cyberella.

There are also situations in which the emulation of a direct agent-to-user communication - from the perspective of the user - is not necessarily the most effective and most convenient way to present information.

Inspired by the evolution of TV commercials over the past 40 years, our group has discovered role plays with synthetic characters as a promising format for presenting information. A typical TV-commercial of the early days featured a sales person who presented a product by enumerating its positive features – quite similar to what synthetic characters do on web pages today. On TV, however, this format has been almost completely replaced by formats that draw on the concept of short, but entertaining episodes or sketches. Typically, such performances embed product information into a narrative context that involves two or more human actors. One of the reasons that may have contributed to the evolution of commercial formats is certainly the fact that episodes offer a much richer basis compared to the plain enumeration of product features, and thus meets the commercial industry's high demand for originality and unseen spots. We propose a shift from single character settings towards interactive performances given by a team of characters as a new form of presentation. The use of presentation teams bears a number of advantages.

First of all, they enrich the repertoire of possible communication strategies. For example, they allow us to convey certain rhetorical relationships, such as pros and cons, in a more canonical manner. Furthermore, they can serve as a rhetorical device that allows for a reinforcement of beliefs. For instance, they enable us to repeat the same piece of information in a less monotonous and perhaps more convincing manner simply by employing different agents to convey it.

Using multiple characters is also a good means to convey social aspects, such as interpersonal relationships between emotional characters, e.g., see [13, 15].

Last but not least, the single members of a presentation team can serve as indices which help the user to organize the conveyed information. For instance, we may convey meta-information, such as the origin of information, or present information from different points of view, e.g. from the point of view of a businessman or the point of view of a traveler.

Looking at past and current projects conducted at DFKI we observe an ongoing evolution of character-based presentation systems. This evolution starts from systems like PPP in which a single character presents information content in the style of a TV-presenter. It continued with systems in which role plays with several characters are used to convey information, and it will bring about interactive systems that invite the user to barge-in ongoing conversations among multiple characters.

1.3 The “Inhabited Market Place”

The Inhabited Market Place (IMP, [1]) is an example of a system that employs presentation teams to convey information about products like cars. As the name indicates, IMP is a virtual place, e.g., a showroom, where seller agents provide product information to potential buyer agents in form of a typical multi-party sales dialogue. The user who observes the simulated dialogue will learn about the features of a car.

From the point of view of the system, the presentation goal is to provide the user with facts about a certain car. However, the presentation is neither just a mere enumeration of the plain facts about the car, nor does it assume a fixed course of the dialogue between the involved agents. Rather, IMP supports the concept of adaptivity. It allows the user to specify prior to a presentation (a) the agents' role, (b) their attitude towards the product (in our case a car), (c) their initial status, (d) their personality profile and (d) their interests. Taking into account these settings, a variety of different sales dialogues can be generated for one and the same product.



Figure 2. The agents Tina and Ritchie engaging in a car sales dialogue in the IMP.

1.4 CrossTalk

Seeking for potential exhibits that could be demonstrated at the CeBIT 2002 computer fair, both the Cyberella system as well as the Inhabited Market Place were favored candidates. During the discussions who of the staff members would be willing to serve at the stand the question arose, why not delegate this task to Cyberella and have her present the Inhabited Market Place on behalf of us. The combination of the two systems resulted in the interactive CrossTalk installation.

A major conceptual contribution of CrossTalk is the introduction of a meta-theater metaphor, taking forward the “computers-as-theatre” paradigm that has been originally presented by Brenda Laurel [12] and since applied by others too, e.g., [7, 11]. Our motivation for introducing the meta-theater metaphor is a practical one. We consider it as a good means to raise the visitor's attention and to enhance his/her interactive experience with animated presentation agents.

2. FUNCTIONAL DESCRIPTION OF CROSSTALK

As sketched in Figure 3, CrossTalk has three main components that are spatially arranged in the form of a triangle:

- the screen of the Cyberella system;
- the screen of the Inhabited Market Place;
- a centre control stand with a touch screen mounted on top, and a camera that notices approaching and leaving visitors.

At the CeBIT booth two 17" LCD monitors were used for the display of the characters. Figure 4 shows a snapshot of the installation taken at the CeBIT stand. While other projection techniques, e.g., data beamers, might be used as well, the display screens should be positioned as shown in the sketch of Figure 3 in order to create the impression of a cross-screen conversation among Cyberella and the couple Tina and Ritchie.

2.1 Basic Role Castings in CrossTalk

In CrossTalk Cyberella's primary task is that of a fair hostess. She welcomes visitors who approach the stand and offer them a demonstration of the Inhabited Market Place (IMP) which she can show on the opposing screen.

In addition, Cyberella plays the role of a mediator between the human stand visitor and the IMP application. For instance, IMP allows a user to set the mood of the agents involved in a product presentation. However, in CrossTalk the visitor does not interact directly with the IMP system. Rather Cyberella will ask the user what settings to choose and then passes the user's input on to the IMP.

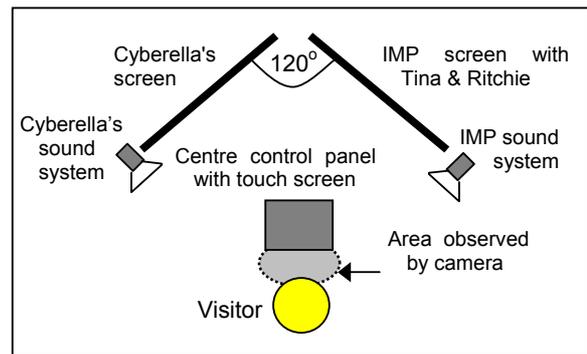


Figure 3. Main components and spatial layout of the CrossTalk installation.

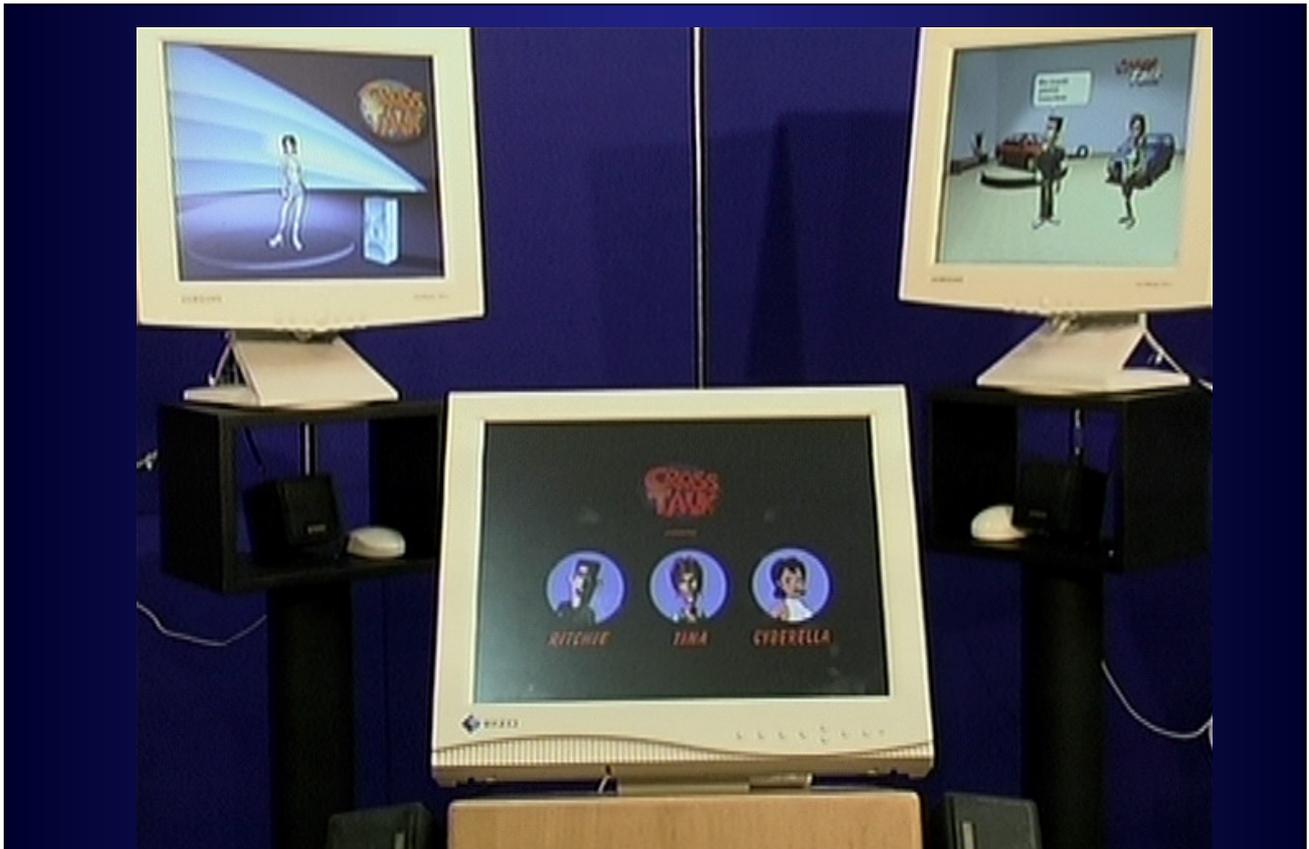


Figure 4. Snapshot of the CrossTalk installation at the CeBIT 2002 fair. The two screens in the back belong to the Cyberella system (left) and the IMP System (right). A visitor can interact with Cyberella via the touch screen in the front.

In contrast, Tina and Ritchie live in the IMP and interchangeably take on the role of a car seller or a potential car buyer. Once triggered by Cyberella to perform a sales dialogue, Tina and Ritchie work through the features of a certain car by means of a question-answering dialogue. Clearly notable variations in the course of such dialogues are due to the specific settings of the agent's current mood. As mentioned above, such settings can be provided by the stand visitor by the way of Cyberella.

2.2 Presentation versus "Off-Duty" Mode

When manning a stand at a tradeshow it is quite natural for the staff members to switch back and forth between presentation activities and more private conversations with colleagues, e.g., when no visitor is present or particularly interested in getting a demo. In CrossTalk, we emulate such a switch between an active presentation mode on the one hand, and an off-duty mode on the other hand.

Tina and Ritchie will switch to presentation mode, too. They are now ready to perform another car sales dialogue. That is, the buyer agent will ask particular questions about a car on display.

When a visitor approaches the centre control stand of the installation (cf. Figures 3 and 4), she/he gets recognised by a camera system mounted below the touch screen of the centre control panel. The camera permanently observes an area of the panel's front side. A "visitor recognised" event will then trigger Cyberella and have her switch to presentation mode. That is, she will greet the interested visitor and offer a presentation of the IMP system.

In response, the seller agent will highlight and explain features of the car that relate to the buyer's questions. Depending on their mood settings, both the buyer and seller will comment on the mentioned car attributes either in a positive or negative way.

In contrast, when a visitor leaves the area close to the centre control stand, a "visitor gone" event is sent to Cyberella. In turn, Cyberella will inform Tina and Ritchie that they can interrupt their car-sales performance and switch to "off-duty" mode.

In off-duty mode the characters still keep acting and talking. This time, however, they either engage in a small talk or start exercising car sales performances. The motivation for including this mode is twofold. Firstly, talking characters are more likely to attract new visitors to come close to the installation, i.e., close enough to get noticed by Cyberella (the camera). Secondly, visitors may step a bit back (i.e., they are no longer seen by the camera) but still watch what the characters do. Switching from presentation mode to off-duty mode often raises the interest of the leaving visitor again since they are now curious what else the characters can talk about.

3. SCRIPTING OF CONVERSATIONS

As indicated in the introduction, the CrossTalk installation has emerged from the combination of the two stand-alone systems Cyberella and IMP.

The Cyberella system comprises a natural language dialogue component that analyses textual input from the user, performs a domain-specific semantic interpretation of input sentences, and generates appropriate responses which are either verbal, gestural

or by facial expression. Technical details about the Cyberella system can be found in [2, 6].

In the case of IMP, the dialogue contributions of all agents participating in a car-sales talk are determined by a centralised action planner. The task of this component can be compared with the task of a script writer who acts out all parts and dialogue contributions for the actors in a theatre play.

Of course, in order to obtain a believable result, the script writer, as well as our automated planning component, have to consider the knowledge and personalities of all characters and must be able to anticipate a reasonable unfolding of the scene. Since the car sales domain is a relatively closed domain, a broad variation of car sales dialogues can be automatically generated by means of a relatively small number of dialogue patterns. The current IMP system comprises approximately 30 planning operators to represent typical dialogue moves in this domain. For a more detailed description of the underlying planning approach we refer to [1, 3].

Compared to the original versions of Cyberella and IMP, the perhaps most interesting feature of CrossTalk manifests itself in the smooth interleaving of pre-scripted subdialogues with fully automatically generated car-sales dialogues.

3.1 Pre-scripted Sub-Dialogues

While a broad variation of car sales dialogues can be automatically generated by means of a relatively small number of dialogue patterns, an approach for the automated generation of small talk dialogues (which would be interesting enough for a visitor to listen to) appears much more challenging. We therefore decided to rely on a pre-authored repertoire of scripted small-talk scenes from which the system would randomly chose when in off-duty mode.

A total of 180 different scenes were composed by one of the authors with experience in theater acting and directing. Some scenes cover themes related to every-day belongings, such as what to do in the evening, how to get home, or where to get cigarettes. Other scenes refer to the world of the theater or movies. So the agents may reflect on their stagecraft, or what to do professionally after the CeBIT convention. An excerpt of a pre-authored episode is shown in Figure 5.

```
...
Tina: I'm so happy when this is over.
Cyb: What are you going to do next?
Tina: Got a job offer from Neckermann.
Cyb: Let me guess, something to do with online catalogues?
Ritchie: <laughter>
Tina: Yeah, something like that.
Ritchie: You're kidding, aren't you?
...
```

Figure 5. Excerpt of a pre-authored dialogue.

The specification of such pre-authored dialogues can also include special tags. Some of these tags allow an author to explicitly specify an agent's non-verbal behaviour, such as the

agent's gaze, gesturing, and body postures. The set of gestures for the CrossTalk agents come from a repertoire based on empirical studies by Kipp [10]. For all three agents the set of gestures and postures has been modelled by a professional animator and turned into libraries of animation clips. Figure 6 shows an excerpt of a pre-scripted episode including tags to specify the agent's non-verbal behaviour.

Scene: OFF-Chat stage-direction

Ritchie: [TINA AS_LookLeft] Ok, if you are interested leave me your number.
 [V_LookToCy]
Tina: Well, <Pau=300> ok.
 [RITCHIE V_LookToActor]
 Sounds ... great. [AS_Glasses]
 I'll think about it.
Cyberella: [GS_Chide] My agent will contact you.
Ritchie: Yeah. Sure. [GS_DoubtShrug] All right.

Figure 6. Excerpt of a pre-authored dialogue including tags that specify non-verbal behaviour.

A further set of tags concern the control of the interactive touch screen. For instance, an author can specify which graphics and input menus should be displayed on the touch screen at a certain point in time. This way, screens that solicit user input can be synchronized with accompanying utterances and gestures of Cyberella.

For the convenience of the writer of pre-scripted episodes, a dialogue compiler has been developed that takes as input a written dialogue script, just as shown in Figure 6, and converts it to an internal representation that can be understood by the dialogue planner. That is, the dialogue segment will be represented by means of a planning operator. When selected in the planning phase, the operator adds the corresponding dialogue segment to the overall script.

3.2 Interweaving Pre-scripted Dialogues with Automated Scripting

In off-duty mode, the agents usually perform a random sequence of pre-scripted scenes (small talk). A special kind of off-duty conversations, however, is achieved by combining pre-scripted episodes with automatically generated car sales talks. The idea is that, when in off-duty mode, the agents may "rehearse" car sales dialogues. In such a "rehearsal", Tina and Ritchie start performing an arbitrary part of the car sales dialogue which is at some point interrupted by a *rehearsal crisis*: for instance, the actors start arguing about pronunciation or the correct version of a text passage, or one of them "forgot" what to say next. After a crisis has been resolved the rehearsal continues in a regular way.

Technically, a rehearsal consists of a chunk of generated car sales dialogue where a pre-scripted scene, the rehearsal crisis, is inserted. An excerpt of a sample rehearsal with a "crisis" is shown in Figure 7. In the scenario Tina acts as a virtual customer who wants to get information about a certain car from Ritchie acting as a virtual car dealer.

...
Tina: How fast can it drive ?
Ritchie: This is a fast car. Maximum speed is 180 KM/H
Tina: *Ahm. What else could I ask ?*
Ritchie: *Everything. Ask me whether you may invite me for dinner.*
Tina: <laughter>
Cyb: *Hey guys, come on, that's not in the script!*
Tina: Ok.
Tina: Does it has leather seats ?
Ritchie: Of course. It's a very luxurious car.
 ...

Figure 7. Excerpt of an automatically generated rehearsal dialogue with an inserted crisis part (in italics) that has been pre-authored.

4. CONCLUSIONS

Originally CrossTalk was developed for the CeBIT convention with the objective to attract CeBIT visitors and motivate them to enter the DFKI booth. Up to now, no formal evaluation of the installation has been performed – neither during the CeBIT nor at other occasions¹ where the system has been demonstrated to a public audience. However, we have observed several interesting reactions from people who have seen the system:

- Most visitors found the installation entertaining, some of them spending more than 15 minutes to watch the characters. Especially the interweavement of the two modes obviously increased the entertaining value of the system.
- Visitors observing the characters in both off-duty and interaction mode reported that watching the characters doing small-talk was more interesting instead of listening to car sales dialogues. This was not really a surprise, since the off-duty scenes contain jokes and personal comments.
- Since the small-talk conversations and the rehearsal crisis dialogues were pre-authored, we were concerned that some of our colleagues who had to share the CeBIT booth for several days with the talking characters could get tired of the installation. However, this fear was unwarranted. Since there was a large number of pre-authored scripts which were randomly selected, there was a relatively small repletion rate per day which was almost not noticed by the stand personnel.
- The cross-screen conversation between Cyberella and the actors Tina and Ritchie achieved a high level of believability. Consequently many users assumed that they could give verbal responses, when prompted by Cyberella. This situation even occurred in cases, when visitors were explicitly told that feedback could be given via the touch screen only.

¹ Among these occasions was an open-house event in the context of the Girls Day initiative of the German Ministry for Education and Research for girls between 10-18 years.

We conclude that the installation suffices its purposes but at the same time see many opportunities to extend and improve CrossTalk further. As pointed out in the previous section, many both visitors wished to talk to Cyberella which would require the addition of a speech interface. As suggested by Cassell and Bickmore [5]. Cyberella may also switch back and forth between a task-oriented presentation mode and a small-talk mode when engaging in a conversation with a visitor.

5. TECHNICALITIES

As mentioned above, CrossTalk emerged from a combination of the preexisting Cyberella system and the IMP system. To combine these two systems and also to include a touch screen interface, an information router based on TCP/IP sockets builds the backbone for the CrossTalk components. For the sake of performance, we recommend the run the three main components of CrossTalk's (i.e., Cyberella, IMP, and the user interface to Cyberella) on different PCs. However, CrossTalk does not require any special hardware rather than ordinary PC's and three separate screens. As sketched in Figure 3, a spatial separation of the display screens for Cyberella on the one hand, and Tina and Ritchie on the other hand is essential to convey the impression of a cross-screen conversation.

Most software components of CrossTalk are implemented in Java. The implementation of the dialogue planner is based on the Java-based JAM Agents architecture framework [9]. The outcome of the planning process are commands to be executed by the agents.

All our animated characters have been designed by one of the co-authors who works as a visual artist and animator. While the single frames of the Cyberella character are drawn by hand and animated with Macromedia's Director, Tina and Ritchie are 3D characters that have been modeled and animated with the tool 3D-Studio Max. Within CrossTalk we use the Microsoft Agent Toolkit² for the runtime animation of the characters and the L&H TTS3000 text-to-speech engine for speech synthesis.

6. ACKNOWLEDGMENTS

The work presented in this paper is a joint effort with contributions from the EU funded IST projects NECA, SAFIRA, and Magicster, and from the project MIAU funded by the German Ministry for Education and Research. Thanks also to our colleague Renato Orsini for the preparation of a CrossTalk video clip which is accessible from <http://www.dfki.de/crosstalk>

7. REFERENCES

- [1] André, E., Rist, T., van Mulken, S., Klesen, M., and Baldes, S.: The Automated Design of Believable Dialogues for Animated Presentation Teams. In: Cassell, J., Sullivan, J., Prevost, S., Churchill, E. (eds.): *Embodied Conversational Agents*, Cambridge, MA: MIT Press, 2000, pp 220-255.
- [2] André, E., Klesen, M., Gebhard, P., Allen, S., and Rist, T.: Exploiting Models of Personality and Emotions to Control the Behavior of Animated Interface Agents. In: Rickel, J. (eds.), *Proceedings of the Workshop on "Achieving Human-Like Behavior in Interactive Animated Agents"* in conjunction with the 4th, Int. Conf. on Autonomous Agents, Barcelona, Spain, 2000, pp 3-7.
- [3] André, E., Rist, T.: Controlling the Behaviour of Animated Presentation Agents in the Interface: Scripting vs. Instructing. In: *AI Magazine*, Vol 22, No. 4, 2001, pp53-66.
- [4] Cassell, J., Sullivan, J., Prevost, S., and Churchill, E. (eds.) *Embodied Conversational Agents*. Cambridge, MA: MIT Press, 2000.
- [5] Cassell, J., Bickmore, T.: Negotiated Collusion: Modelling Social Language and Its Interpersonal Effects in Intelligent Agents. *User Modelling and Adaptive Interfaces*. (Forthcoming)
- [6] Gebhard, P.: Enhancing Embodied Intelligent Agents with Affective User Modelling. In: Vassileva, J. and Gmytrasiewicz, P. (eds.), UM2001, In: *Proceedings of the Eighth International Conference* (Doctoral Consortium summary) Berlin: Springer, 2001.
- [7] Hayes-Roth, B., van Gent, R.: Story-Making with Improvisational Puppets. In: *Proceedings of Autonomous Agents '97*, 1997, pp 92-112.
- [8] Huber M.: JAM: A BDI-theoretic mobile agent architecture. In: *Proceedings of the Third Conference on Autonomous Agents*, New York: 2001, pp 236-243.
- [9] Johnson, W.L., Rickel, J.W. and Lester J.C.: Animated Pedagogical Agents: Face-to-Face Interaction in Interactive Learning Environments. *International Journal of Artificial Intelligence in Education* 11, 2000, pp 47-78.
- [10] Kipp, M.: From Human Gesture to Synthetic Action. In: *Proceedings of the Workshop on "Multimodal Communication and Context in Embodied Agents"* Montreal, 2001, pp 9-14.
- [11] Klesen, M., Szatkowski, J., and Lehmann, N.: A Dramatized Actant Model for Interactive Improvisational Plays. In: A. de Antonio, R. Aylett, and D. Ballin (eds.) *Proceedings of the Third International Workshop on Intelligent Virtual Agents*. Lecture Notes in Artificial Intelligence 2190. Heidelberg: Springer Verlag, 2001.
- [12] Laurel, B.: *Computers as Theatre*. Reading Mass.: Addison-Wesley. 1993.
- [13] Prendinger, H. and Ishizuka, M.: Social Role Awareness in Animated Agents. In: *Proceedings of the Fifth Conference on Autonomous Agents*, New York: ACM Press. 2001, pp 270-377.
- [14] Rist, T., André, E.: Adding Animated Presentation Agents to the Interface In: *Proceedings of the International Conference on Intelligent User Interfaces IUI '97*, 1997, pp 79-86.
- [15] Rist, T., Schmitt, M.: Avatar Arena: An Attempt to Apply Socio-Physiological Concepts of Cognitive Consistency in Avatar-Avatar Negotiation Scenarios. In: *Proceedings of AISB'02 Symposium on Animated Expressive Characters for Social Interactions*, London, 2002, pp 79-84.

² <http://www.microsoft.com/msagent>