AI Magazine Volume 22 Number 4 (2001) (© AAAI)

Controlling the Behavior of Animated Presentation Agents in the Interface

Scripting versus Instructing

Elisabeth André and Thomas Rist

■ Lifelike characters, or animated agents, provide a promising option for interface development because they allow us to draw on communication and interaction styles with which humans are already familiar. In this contribution, we revisit some of our past and ongoing projects to motivate an evolution of character-based presentation systems. This evolution starts from systems in which a character presents information content in the style of a TV presenter. It moves on with the introduction of presentation teams that convey information to the user by performing role plays. To explore new forms of active user involvement during a presentation, the next step can lead to systems that convey information in the style of interactive performances. From a technical point of view, this evaluation is mirrored in different approaches to determine the behavior of the employed characters. By means of concrete applications, we argue that a central planning component for automated agent scripting is not always a good choice, especially not in the case of interactive performances where the user might take on an active role as well.

A growing number of research projects in academia and industry have started to develop lifelike characters or agents as a metaphor for highly personalized human-machine communication. Work in this area is motivated by a number of supporting arguments, including the fact that such characters allow for communication styles common in human-human dialogue and thus can release users from the burden to learn and familiarize themselves with less native interaction techniques. Furthermore, well-designed characters show great potential for making interfacing with a computer system more enjoyable. One aspect when designing a character is to find a suitable visual and audible appearance. In fact, there is now a broad spectrum of characters that rely on either cartoon drawings, recorded (and possibly modified) video images of persons, or geometric three-dimensional (3D) body models for their visual realization with recorded voices or synthesized speech and sound to determine their audible appearance.

Audio-visual attractiveness, however, is not everything. Rather, the success of an interface character in terms of user acceptance and interface efficiency very much depends on the character's communication skills and its overall behavior. On a very low level of abstraction, the behavior of an agent can be regarded as the execution of a *script,* that is, a temporally ordered sequence of actions including body gestures, facial expressions, verbal utterances, locomotion, and (quasi-) physical interactions with other entities of the character's immediate environment. It comes as no surprise then that behavior scripting, in one way or another, has been widely used in projects that deal with interface characters. For example, a straightforward approach is to equip the character with a library of manually authored scripts that deter-

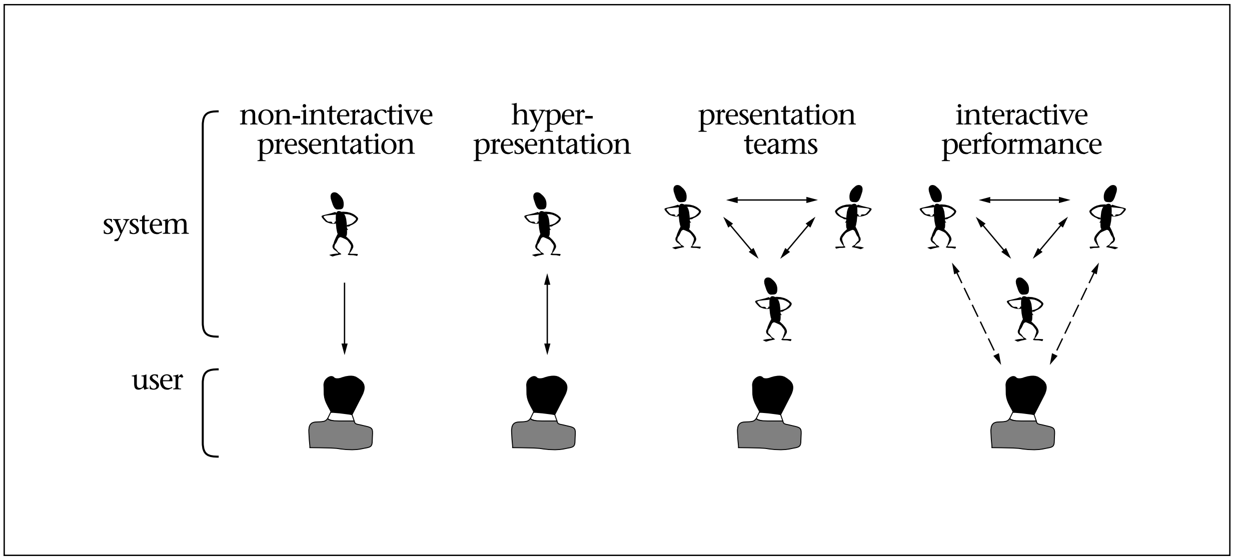

Figure 1. Agent-User Relationships in Different Presentation Scenarios.

mine what the character might do in a certain situation. At run time, the remaining task is to choose from the library a suitable script that meets the constraints of the current situation and at the same time helps to accomplish a given task. What is specified in a character script is also a matter of the level of abstraction and the expressiveness of the scripting language. In some cases, the scripting language is built on an existing general-purpose script-based programming language. For example, the Microsoft (1999) AGENT characters can easily be scripted either in VISUAL BASIC or in JAVA script, allowing the script writer to use the standard control structures of these languages, such as conditional statements or loop constructs. As an alternative to character-specific adjuncts to programming languages, XML-compliant scripting languages, such as TVML (Hayashi, Gakumazawa, and Yamanouchi 1999), can be defined. In any case, the script can be seen as a kind of application programming interface (API) that allows users to specify the agent's behavior at a certain level of abstraction.

Unfortunately, the problem with manually authored scripts and script libraries is that the author has to anticipate scripts for all possible situations and tasks and that the scripts must allow for sufficient variations to avoid characters that behave in a monotonic and very predictable way. Furthermore, the manual scripting of presentation agents can become quite complex and error prone because synchronization issues have to be considered. To avoid extensive script writing but, nevertheless, to enable a rich and flexible character behavior, one can use a generative mechanism that composes scripts according to a set of composition rules. Our contribution to this area of research was the development of a plan-based approach to automate the process of writing scripts that are forwarded to the characters for execution. This approach has successfully been applied to build a number of applications in which information is conveyed either by a single presenter or a team of presentation agents. While exploring further application fields and new presentation styles, we identified, however, some principle limitations of scripting presentations with characters. One decisive factor is the question of whether all information to be conveyed by a character is available before a presentation is started. Another aspect is the kind of user interactions that will be supported during the display of a presentation.

In this contribution, we revisit some of our past and ongoing projects to motivate an evolution of character-based presentation systems, as illustrated in figure 1.

The first of the four depicted scenarios represents presentation systems that deploy a single character to present information. Although automated presentation planning will be used to work out the complete script, from the perspective of the user, a generated presentation appears quite similar to the display of a video clip because no user interaction is foreseen at display time. In contrast, the second scenario can be compared to a hypermedia presentation in which the user can actively influence the course of a presentation at certain decision points. Moving on to the third scenario actually means a shift from face-to-face presenter-user setting to a user-as-observer setting. That is, two or more characters give a performance on the screen to convey information to the observing audience. However, no user intervention is foreseen during a performance. Finally, there is the setting of an interactive performance in which the user can take on an active role, as in the fourth scenario. From a technical point of view, the fourth scenario is perhaps most challenging because one has to resolve on an operational level the conflict between predestination and freedom of interaction.

In the following sections, we pick a number of concrete application examples to describe in more detail both the characteristic features of the presentation scenarios as well as the machinery that can be used in a corresponding presentation system.

Related Work

A number of research projects have discovered lifelike agents as a new means of computer-based presentation. A popular application is the simulation of human TV speakers. Examples include Noma and Badler's (1997) virtual weather reporter as well as Thalmann and Kalra's (1995) TV presenter. The BYRNE system (Binsted and Luke 1999) makes use of an embodied commentator that is represented by a talking head to comment on matches of the RoboCup simulation league (Kitano et al. 1997). Furthermore, animated agents have been used as presenters in various tutoring systems. Prominent examples include HERMAN-THE-BUG, the pedagogical agent of DESIGN-A-PLANT (Lester et al. 1999), or STEVE, a virtual trainer for 3D environments (Rickel and Johnson 1999). To facilitate the production of presentation sequences, most systems rely on a basic repertoire of actions that allow for a high-level specification of the agents' behavior. For example, the input for Noma and Badler's weather reporter consists of text to be uttered by the presenter and commands, such as pointing and rejecting, which refer to the presenter's body language. DESIGN-A-PLANT and STEVE go a step further and automatically compose presentation scripts from design primitives according to a repertoire of construction rules. In DESIGN-A-PLANT, such primitives include complex multimedia objects, such as audio or video clips, that have been manually designed by a multidisciplinary team of graphic artists, animators, musicians, and voice specialists. On the one hand, this approach allows the authoring of high-quality presentations because the human author has much control over the material to be presented. On the other hand, enormous effort by the human author is required to produce the basic repertoire of a course. Instead of creating animation sequences for a course offline and putting them dynamically together, the 3D character STEVE is directly controlled by commands, such as look at, walk to, or grasp an object. In this case, the character interacts with virtual objects in the same way as a human does in a real environment with direct access to the objects.

A new presentation style has been introduced by the use of multiple characters that convey information by means of simulated dialogues. In some cases, these dialogues are manually scripted as in the AGNETA and FRIDA system (Höök et al. 2000) that incorporates narratives into a web environment by placing two characters on the user's desktop. These characters watch the user during the browsing process and make comments on the visited web pages. In other cases, the content of the dialogue script for the single characters is based on a real conversation between humans. For example, Yabe, Takahashi, and Shibayama (2000) animate discussions from news groups by casting 3D characters as the authors of the single-discussion contributions. Systems that aim at a simulation of conversations between humans usually automate at least parts of the generation process. Cassell and colleagues (Cassell et al. 1994) automatically create and animate dialogues between a bank teller and a bank employee with appropriate synchronized speech, intonation, facial expressions, and hand gestures. Walker, Cahn, and Whittaker (1997) concentrate on the linguistic capabilities of computer characters (for example, a waiter and a customer) and examine how social factors influence the semantic content, the syntactic form, and the acoustic realization of conversations. The generation of their dialogues is essentially influenced by the power the listener has on the speaker and the social distance between them. MR. BENGO (Nitta et al. 1997) is a system for the resolution of disputes that uses three agents: (1) a judge, (2) a prosecutor, and (3) an attorney who is controlled by the user. The prosecutor and the attorney discuss the interpretation of legal rules. Finally, the judge decides on the winner.

Last but not least, our work was inspired by research on interactive drama that aims at integrating a user in a scenario—either as an audience member or an active participant. To allow for user interaction, systems usually incorporate decision points in a narrative-style script (Mateas 1997) or model their characters as autonomous agents that select and instantiate actions under consideration of dramatic constraints, such as the plot of a story or the characters' role and personality (Hayes-Roth and van Gent 1997). Even though the focus of

A number of research projects have discovered lifelike agents as a new means of computer-based presentation. A popular application is the simulation of human TV speakers.

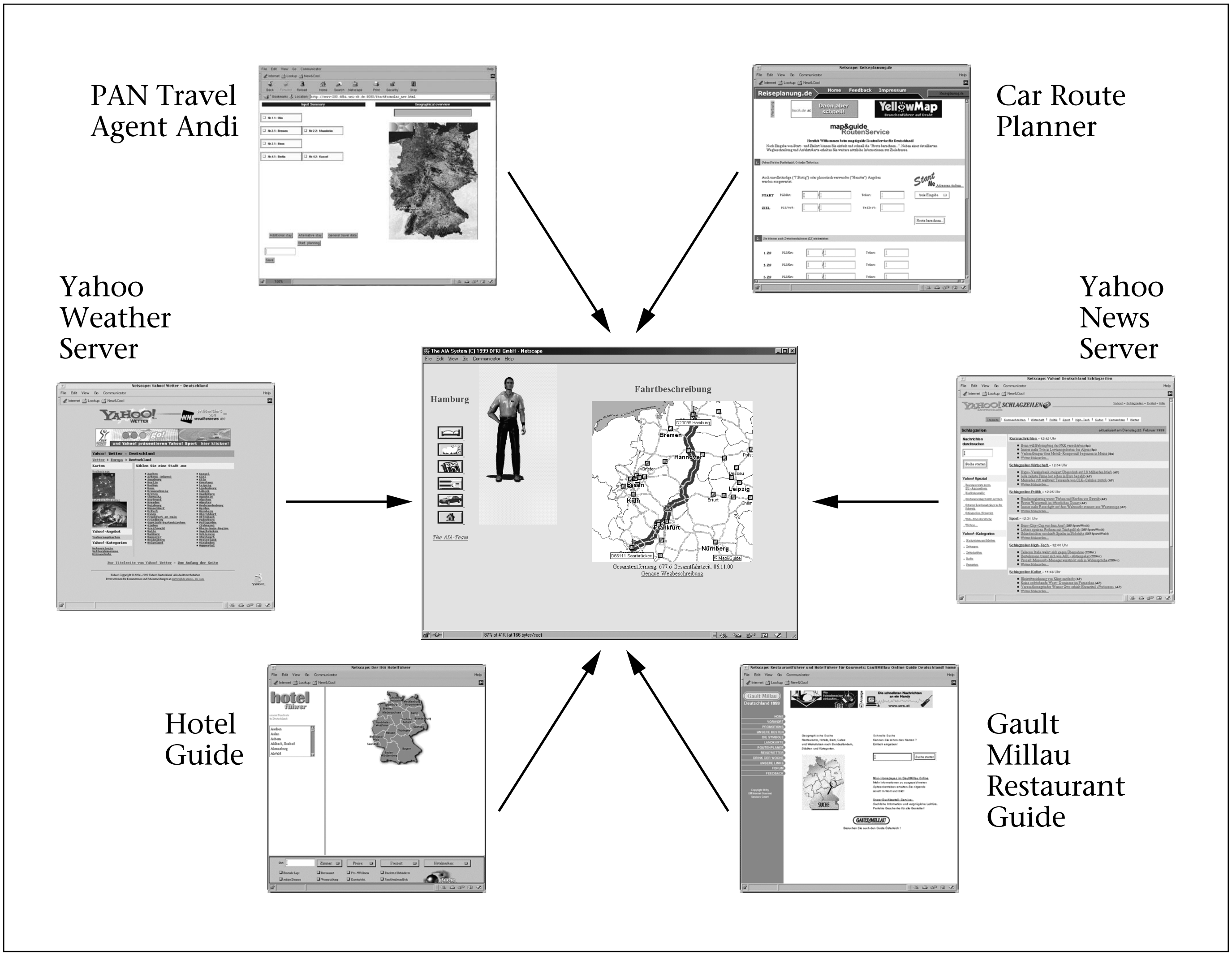

Figure 2. The AIA Travel Agent.

interactive drama is usually not on the communication of information, dramatic elements offer great promise in presentation scenarios as well (see also Laurel [1991]).

Deploying a Single Presentation Agent

In many cases, the success of a presentation depends not only on the quality of the multimedia material but also on how it is presented to the user. Inspired by human speakers, we decided to use an animated character that shows, explains, and verbally comments on textual and graphic output on a window-based interface.

Sample Applications: PPP-PERSONA and Its Siblings

In our earlier work, we conducted two projects to develop systems that fall into this category: (1) the PPP (personalized plan-based presenter) Project (André and Rist 1996) and the AIA (adaptive communication assistant for effective infobahn access) Project (André, Rist, and Müller 1999). In PPP, we addressed the automated generation of instructions for the operation of technical devices that were delivered to the user by an animated agent, the so-called PPP-PERSONA. For example, to explain how to switch on a device, PPP-PERSONA showed the user a picture of the device, pointed to the on-off switch, and instructed him/her verbally how to manipulate the switch. In the AIA Project, we developed a series of personalized information assistants that aimed at facilitating user access to the web. Besides the presentation of web contents, the AIA agents provide orientation assistance in a dynamically expanding navigation space. Figure 2 shows one of our applications, which is a personalized travel agent. Based on

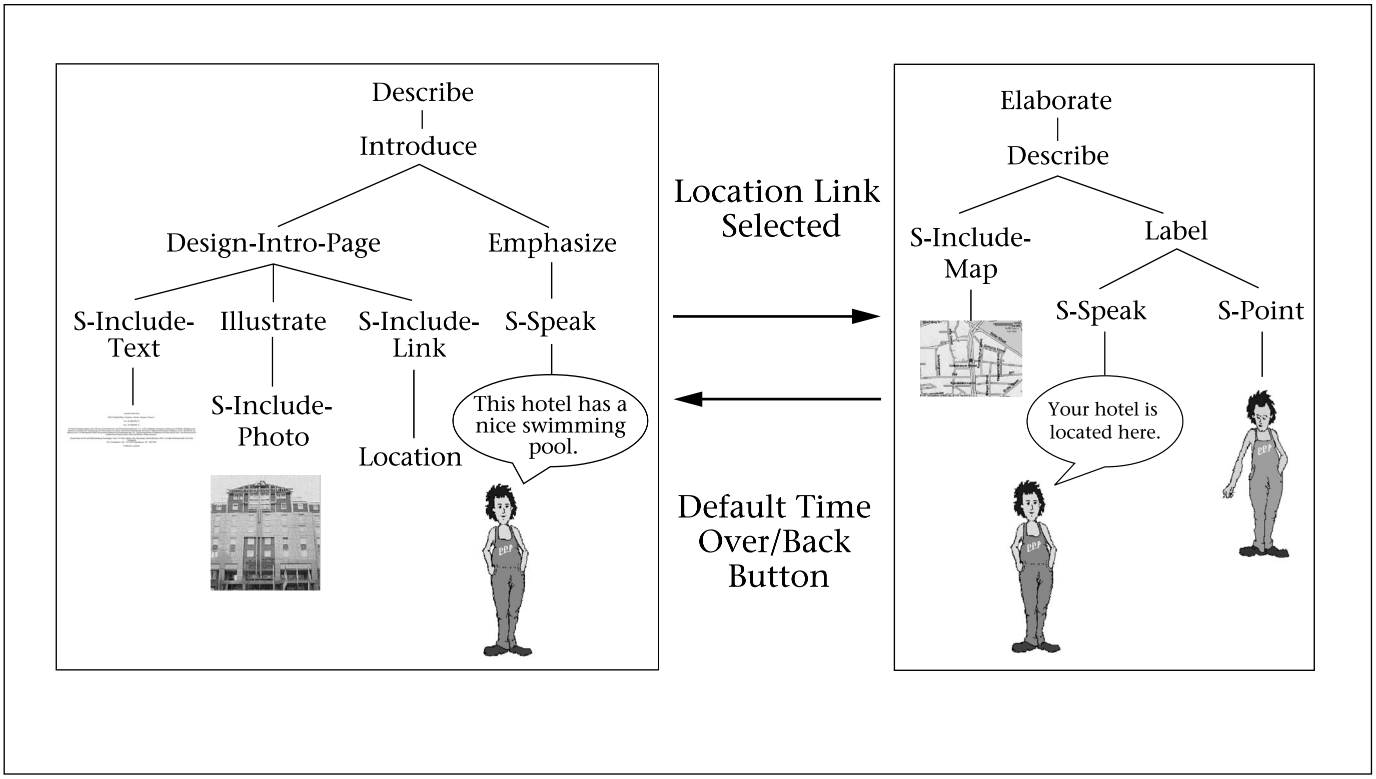

Figure 3. Presentation Structures.

the user's request, for example, to provide travel information for a trip from Saarbrücken to Hamburg, AIA retrieves relevant information from the web, reorganizes it, encodes it in different media (such as text, graphics, and animation), and presents it to the user as a multimedia web presentation.

Scripting the Behavior of Single-Presentation Agents

In PPP and AIA, the agents' behavior is determined by a script that specifies the presentation acts to be carried out as well as their spatial and temporal coordination. Creating scripts manually is, however, not feasible for many applications because it would require them to anticipate the needs of all potential users in preparing presentations for them. For example, in PPP, the user could specify a time limit for the presentation. Depending on the setting of this parameter, the generated instructions would have varied significantly with respect to the provided degree of detail. Manual scripting would have required a large library of different presentation scripts, taking into account all potential settings of the time limit. In the case of the AIA system, manual scripting is even more unpracticable because the information to be presented dynamically changes, and there is simply not enough time to manually create and update presentations. Based on these observations, we decided to automate the script-generation process.

We rely on our earlier work on presentation design (André and Rist 1995) and formalize action sequences for composing multimedia material and designing scripts for presenting this material to the user as operators of a planning system. The effect of a planning operator refers to a complex communicative goal (for example, to describe a technical device in PPP or a hotel with vacancies in AIA), whereas the expressions of the body of the operator indicate which acts have to be executed to achieve this goal (for example, to show an illustration of a certain object and describe it). In addition, the plan operators allow us to specify spatial and temporal layout constraints for the presentation segments corresponding to the single acts. The input of the presentation planner is a complex presentation goal. To accomplish this goal, the planner looks for operators whose headers subsume it. If such an operator is found, all expressions in the body of the operator are set up as new subgoals. The planning process terminates if all subgoals have been expanded to elementary production-retrieval or presentation tasks (for details, see André, Rist, and Müller [1999]).

Handling User Interaction

In AIA and PPP, we primarily use the animated characters for presenting information. User feedback is only supported in a rather limited way—through a hypermedia-style interface. To allow for the dynamic expansion of the navigation space, we do not script the complete presentation in advance. Instead, we expand certain parts of a presentation only on demand. The basic idea is to associate with each hyperlink a presentation goal for which the planner is started again if the user selects it at presentation run time. This method has the advantage that presentations can be adapted to the user's previous navigation behavior and the information that has been conveyed thus far.

Figure 3 shows an example. We assume that the planner is called with the goal *describe hotel* and comes up with a presentation consisting of a spoken utterance by the PERSONA, a photo of the hotel, and a text paragraph that includes a hyperlink referring to the location of the hotel. The discourse structure of this presentation is shown on the left. If the user selects the location link at run time, the planner is started again with the goal *elaborate location,* which leads to the discourse structure on the right.

Lessons Learned from AIA and PPP

In the AiA and the PPP projects, we take a human screen writer as a role model who creates a script for a speaker prior to a presentation. The approach seems appropriate for the generation of presentations that (1) follow a narrative-style structure either linear or with a finite number of branching points only and (2) allow a clear temporal separation of script-writing time and presentation-display time. In case both conditions are satisfied, highly coherent presentations can be generated, for example, by using a plan-based approach.

Even though PPP and AIA allow for the dynamic expansion of scripts at run time, the user has, however, only limited possibilities for interaction. The major disadvantage of a scripting approach lies in the fact that all user interactions have to be anticipated to some extent. For example, the system has to decide at scripting time where to incorporate choices in terms of hyperlinks.

Presentation Teams

Frequently, systems that use presentation agents rely on settings in which the agent addresses the user directly as if it were a face-to-face conversation between human beings. Such a setting seems quite appropriate for a number of applications that draw on a distinguished agent-user relationship. For example, an agent can serve as a personal guide or assistant in information spaces such as the World Wide Web (as the PPP-PERSONA and its siblings). However, there are also situations in which the emulation of a direct agent-to-user communication is not necessarily the most effective way to present information. Empirical evidence suggests that at least in some situations, indirect interaction can have a positive effect on the user's performance. For example, Craig and colleagues (Craig et al. 1999) found that in tutoring sessions, users who overheard dialogues between virtual tutors and tutees subsequently asked significantly more questions and also memorized the information significantly better.

Frequently, systems that use presentation agents rely on settings in which the agent addresses the user directly as if it were a face-to-face conversation between human beings. Such a setting seems quite appropriate for a number of applications that draw on a distinguished agent-user relationship.

Based on this observation, we started to investigate a new style of presentation that conveys information in a less direct manner. We use a team of characters that do not interact directly with the user but with other characters as if they were performing a play to be observed by the user. The use of presentation teams bears a number of advantages. First, 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 example, they enable us to repeat the same piece of information in a less monotonous and perhaps more convincing manner simply by using different agents to convey it. Finally, the single members of a presentation team can serve as indexes that help the user to organize the conveyed information. For example, we can convey metainformation, such as the origin of information, or present information from different points of view, for example, from the point of view of a businessperson or a traveler. This presentation style is currently explored within two projects: The INHABITED MARKETPLACE Project (André et al. 2000a) and the MAGIC LOUNGE Project (Rist et al. 2000).

Sample Applications: The INHABITED MARKETPLACE and the MAGIC MONITOR

The objective of the *inhabited marketplace* (IMP) is to investigate sketches, given by a team of lifelike characters, as a new form of sales presentation (André et al. 2000a). As suggested by the name, the IMP is a virtual place in which seller agents provide product information to potential buyer agents. Figure 4 shows a dialogue between Merlin playing the role of a car seller and Genie and Robby as buyers. For the graphic realization of this scenario, we use the Microsoft AGENT package; see Microsoft (1999).

Figure 4. INHABITED MARKETPLACE.

Genie has uttered some concerns about the high running costs that Merlin tries to play down. From the point of view of the system, the presentation goal is to provide the observer—who is assumed to be the real customer—with facts about a certain car. However, the presentation is not just a mere enumeration of the plain facts about the car. Rather, the facts are presented along with an evaluation that takes into account the observer's interest profile that can be specified prior to a presentation. In addition, the presentation reflects the characters' personality features that can be chosen by the user as well. In its current version, however, the inhabited marketplace does not allow user interaction during run time. The scenario was inspired by Jameson and colleagues (Jameson et al. 1995), who developed a dialogue system that models noncooperative dialogues between a car seller and a buyer. However, although the objective of Jameson and colleagues is the generation of dialogue contributions that meet the goals of the system that can either take on the role of the seller or the buyer, our focus is on the development of animated agents that convey information by giving performances.

The *MAGIC MONITOR* is a tool for illustrating message exchange within the MAGIC LOUNGE virtual meeting space. Cartoon-style characters are used to represent different conversation partners (which can be humans or virtual conversational agents). That is, in the MAGIC MONITOR, the single characters essentially serve to encode the source of information. The tool allows for the playback of recorded message exchanges according to different structuring criteria, such as timelines or dialogue threads. Figure 5 displays a snapshot of the graphic user interface. The MAGIC MONITOR uses a dedicated commentator character, which is used to represent the MAGIC LOUNGE system itself. In figure 5, a facilitator agent is located in the lower right

Figure 5. Magic Monitor.

area of the screen. The five characters in the central area represent different users. One of them is uttering the text of a message that was sent by the user it represents.

Generating Scripts for Presentation Teams

In the IMP, the system takes on a similar role as in the AIA and PPP systems, with the difference that it now has to script the behavior of a group of actors that engage in a dialogue. Again, we follow a communication-theoretic view and consider the automated generation of such scripts a planning task. Nevertheless, a number of extensions became necessary to account for the new communicative situation. First, information is no longer presented by a single agent that stands for the presentation system but, instead, is distributed over the members of a presentation team whose activities have to be coordinated. Second, information is not conveyed by executing presentation acts that address the user but by presenting a dialogue between several characters to be observed by him/her. Third, to generate effective performances with believable dialogues, we cannot simply copy an existing character. Rather, characters have to be realized as distinguishable individuals with their own areas of expertise, interest profiles, personalities, emotions, and audiovisual appearance.

To meet these requirements, we extended our repertoire of communicative acts by dialogue acts, such as "responding to a question" or "making a turn," and defined plan operators that code a decomposition of a complex communicative goal into dialogue acts for the single agents. Dialogue acts include not only the propositional contents of an utterance but also its communicative function, such as taking turns or responding to a question. This is in line with Cassell et al. (2000), who regard conversational behaviors as fulfilling propositional and interactional conversational functions.

An example of a dialogue operator is listed in figure 6. It represents a scenario where two agents discuss a feature of an object. It only applies if the feature has a negative impact on any value dimension and if this relationship can easily be inferred. According to the operator, any disagreeable buyer produces a negative comment referring to this dimension (NegativeResp). The negative response is followed by a response from the seller (RespNegativeResp).

The character's profile is treated as an additional filter during the selection instantiation and the rendering of dialogue strategies. For example, we can define specific dialogue strategies for characters of a certain personality and formulate constraints that restrict their applicability.

The plan operators implemented to date typically result in dialogue sequences similar to that presented in figure 7. One of the buyers asks a question to which the seller provides an answer. The buyer then has the option of providing an evaluation that can be commented on by the seller again.

In the MAGIC MONITOR, the system plays a twofold role. First, it takes on the role of a screen writer that determines how a conversation within the MAGIC LOUNGE should be played back by the single characters. Second, it is actively involved in the presentation as a facilitator. To determine the presentation behavior of both the facilitator and the user representatives, we rely on the same planning approach as in the IMP. Parameterizable presentation operators determine how recorded message exchange will be presented in the MAGIC MONITOR. For example, our basic strategy for message playback according to the timeline of occurrence comprises the following substeps: (1) have the facilitator announce the next message by referring to aspects, such as sender, date, time, topic and speech act; (2) have the character representing the sender of the message move to the center podium; (3) if any, have the specific message recipients move closer to the podium and play a listen animation; (4) have the sender read the message while it plays an animation that enforces the speech act associated with the message; and (5) have all characters move back to their assigned home positions.

Variations of the basic scheme have been defined to cope with messages, such as log in, log out, and registration, and to take into account various aspects such as consecutive messages from the same user or the need for a more compact presentation.

```
NAME: "DiscussValue1"
HEADER: (DiscussValue ?attribute)
CONSTRAINTS:
(*and*    (Polarity ?attribute ?dimension "neg")
          (Difficulty ?attribute ?dimension "low")
          (Buyer ?buyer)
          (Disagreeable ?buyer)
          (Seller ?seller))
INFERIORS:
(A1 (NegativeResp ?buyer ?dimension))
(A2 (RespNegResp ?seller ?attribute ?dimension))
```

Figure 6. Example of a Dialogue Operator.

Robby: How much gas does it consume?

Merlin: It consumes 10 litres per 100 km.

Peedy: That's bad for the environment!

Merlin: Bad for the environment? It has a catalytic converter, and it is made of recyclable materials.

Figure 7. Generated Dialogue Fragment of the Sales Scenario.

Lessons Learned from Presentation Teams

In the IMP and the MAGIC MONITOR, several different characters are allocated to convey information units to the users. The distribution and organization of contributions followed human-to-human conversations. In the case of the MAGIC MONITOR, this task was relatively easy to achieve because the system can exploit the structure of the recorded chat-style conversations. In the case of the IMP, there is no natural dialogue structure as such coded in the domain data (that is, the car database). To make the presentation coherent and, thus, the whole scenario more plausible, we superimposed a plotlike structure as a frame within which the characters make dialogue turns. For example, there is a greeting scene followed by a scene in which the car attributes are discussed and, finally, a conclusion scene after the customers made up their minds.

In both applications, the behavior of all agents is scripted by a single script-writing component prior to the performance. Such an approach bears the advantage that it enables the generation of coherent dialogues. It requires, however, that all the knowledge to be communicated to the audience is known in advance. Consequently, it is less suitable in sit-

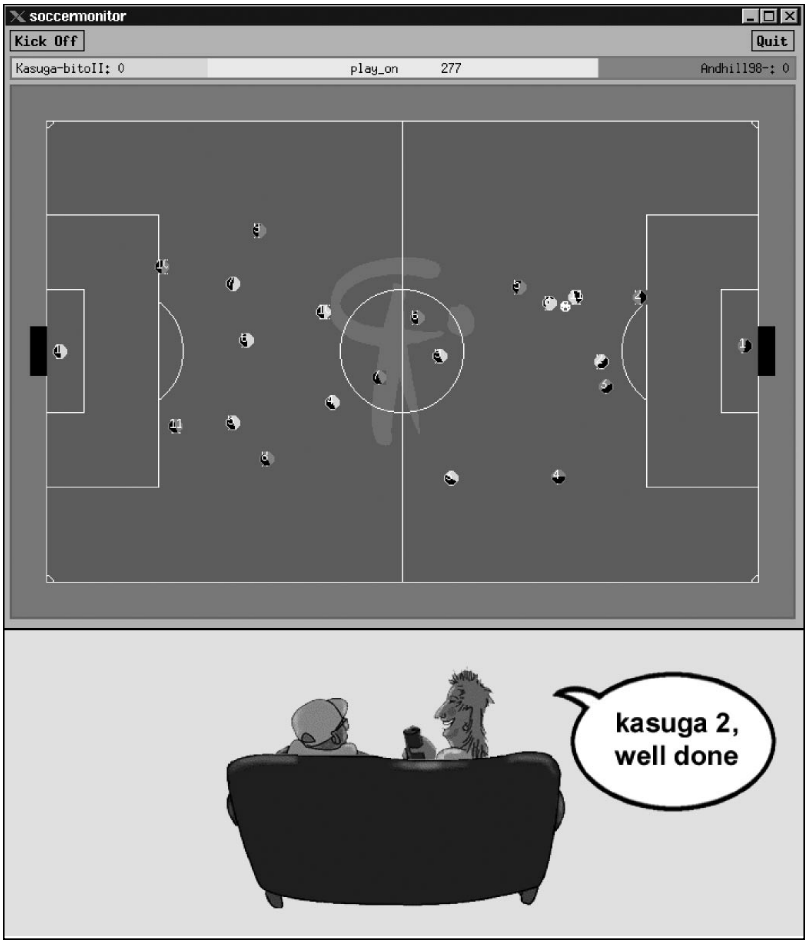

Figure 8. Commentator Team Gerd and Matze.

uations where the agents have to immediately respond to events at presentation run time, such as new incoming information to be presented or user interactions.

From a knowledge engineering point of view, there is also the argument that the definition of planning operators becomes a more complicated task as the number of involved agents increases because of the combinatorics of the set of possible interagent relationships that have to be taken into account.

Instructing Improvisational Presentation Agents

There are a number of application fields for presentation agents where scripts for the agents cannot be worked out in advance because all or part of the information to be presented becomes available only at presentation run time. Any kind of reportage or commentary of live data falls into this category.

Sample Application: Gerd and Matze Commenting on RoboCup Soccer games

As an example, we discuss ROCCO II (André et al. 2000b), an automated live report system that generates commentaries for the simulator league of Robot World Cup Soccer. Figure 8 shows a screen shot of the system that was taken during a typical session. In the upper window, two (software agent) teams play a soccer game while they are commented on by two soccer fans: Gerd and Matze sitting on a sofa. As in the IMP, the user has the option of experimenting with different character profiles. The user can characterize Gerd and Matze by their personality and their sympathy for a certain team.

Generating Live Commentary

Unlike the agents in the car sales scenario, the RoboCup commentators have to comment on a rapidly changing environment. Because events on the soccer field evolve while time progresses, no global organization of the presentation is possible. Instead, the commentators have to respond immediately to incoming data. Furthermore, they have to meet severe time constraints imposed by the flow of the game. In some cases, it might even be necessary for the commentators to interrupt themselves. For example, if an important event (for example, a goal kick) occurs, utterances should be interrupted to communicate the new event as soon as possible. In such a situation, it is not possible to prescript utterances. Instead scripting has to be done at run time, for example, either by a centralized script-writing component or the single agents themselves.

We decided to use a self-scripting approach and realize Gerd and Matze as (semi-) autonomous agents. To implement this scenario, we assign each agent its own reactive planner and code the agents' dialogue strategies as operators of the single planners. Dialogue contributions then result from autonomous characters trying to achieve their individual goals. The goal of the single commentators is to inform the viewer about ongoing events in the scene. We assume that both commentators share all knowledge about the events on the soccer field, which is provided by ROCCO's incremental event-recognition component (André et al. 2000b). Assuming a discrete time model, at each increment of a time counter, the recognition component selects relevant events, formulates corresponding presentation tasks, and writes them into a buffer. In addition, the buffer contains presentation tasks that refer to the presentation of background information. If an event has been communicated, or in case, the topicality of an event falls below a threshold, the corresponding presentation task is removed from the buffer.

Each agent has a set of strategies at its disposal for accomplishing presentation tasks or interacting with the other agent. As in the sales scenario, the character's profile is used to parameterize the presentation planning process. In addition, the system comprises various strategies for coordinating the dialogue behavior of the single agents. In the current version of ROCCO II, each commentator concentrates on the activities of a certain team. That is, there is an implicit agreement between the characters concerning the distribution of dialogue contributions. Another possibility we are currently exploring is to use a probabilistic approach in which extravert characters get a word in with a higher probability than introvert characters. Responses to dialogue contributions of the other commentator are possible provided the speed of the game allows for it. Furthermore, the commentators can provide background information on the game and the involved teams.

Lessons Learned from Gerd and Matze

Although in the IMP, a plot structure was used for the generation of a plausible discourse, Gerd and Matze rely on a character-centered generation approach. Instead of specifying the agents' behavior to the last detail, we just provide them with instructions that they can work out themselves at presentation run time. Instructions include information on the character's role and personality profile that constrain the character's goals and behaviors but still provide enough space for improvisation. Such an approach seems appropriate for scenarios that require immediate responses to external events. It is, however, more difficult to ensure the coherence of a dialogue because no global organization of the information is possible. Think of two people giving a talk together without clarifying in advance who is going to explain what.

Toward Interactive Performances

In the systems presented to this point, user interaction was rather limited or not possible at all. In our search for new presentation styles, we are currently exploring possibilities for involving the user in presentation scenarios. One basic idea is provide the user with the option of taking an active role in the performance if he/she wants to do so. If not, however, the characters will give a performance on their own—maybe encouraging the user to give feedback from time to time. Such a scenario bears a lot of similarities to improvisational theatre (compare Johnstone [1989]). First, there is no predefined script. Instead, the dialogue between the user and the characters evolves while time progresses. Furthermore, the scenario is open ended. Neither the characters nor the users are able to tell what exactly might happen next.

Sample Application: The INHABITED MARKETPLACE 2

As a first example of a system that presents information in the style of an interactive performance, we are currently developing a second version of the IMP. The presentation task and scenario are similar to the original version (compare section Sample Applications: The INHABITED MARKETPLACE and the MAGIC MONITOR). Following the principle idea described earlier, our goal is now to allow the user to step into the role of an accompanying customer who can pose questions and support, reinforce, or reject arguments made by the seller or her/his cobuyers.

Structuring Interactive Performances

Because the dialogue behavior of the agents does not follow a script, there is the danger that the dialogues get rather incoherent. As a consequence, the user might lose interest and even leave the IMP. To provide more structure to the dialogues without constraining the characters' improvisational behaviors too much, we use a slightly modified version of a dramaturgy framework that has been proposed by our collaborators from IDAU Denmark for the PUPPET Project (see Klesen et al. [2001]). A major element of this framework, which goes back to Greimer and Courtes's (1982) ACTANT model, is that of an underlying conflict that is established by introducing a protagonist, which persecutes a certain goal, and a second character, the antagonist, that tries to accomplish a counter goal (figure 9). Both the protagonist and the antagonist can be supported by one or more helpers. Once started, a certain "dramatic story" would unfold over time just by having the involved actors play their assigned roles.

In the case of the IMP, we decided to model a buyer and a customer with conflicting interests. Although the seller tries to present a car in a positive light, the customer persecutes the opposite goal, namely, to point out the weaknesses of the car. In addition, we foresee a helper character that is to support the virtual customer and is played by the user. As a helper agent, the user can interact in the following ways: He/she can support the virtual customer

In the systems presented to this point, user interaction was rather limited or not possible at all. In our search for new presentation styles, we are currently exploring possibilities for involving the user in presentation scenarios.

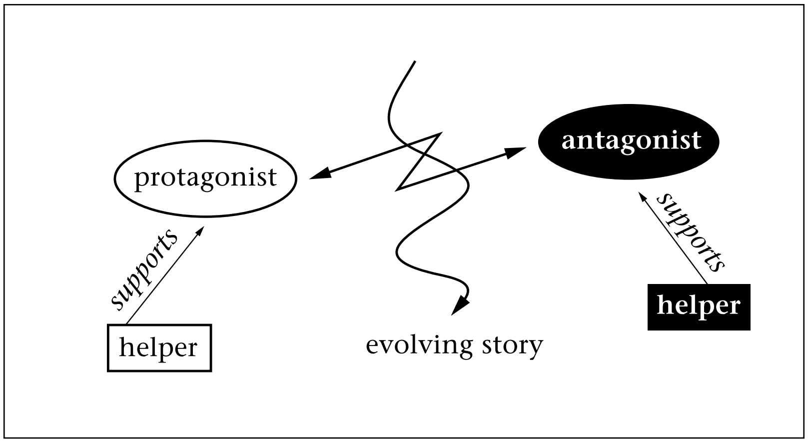

Figure 9. ACTANT *Model.*

directly by confirming its statements or expressing approval. In response to a seller's statement, he/she can utter disbelief or mention understanding problems.

The user's utterances influence the course of the dialogue in different ways. First, they can cause changes in the virtual agents' emotional state, which, in turn, will have an impact on the agents' dialogue behavior. Second, the user's utterances can result in new dialogue goals for the virtual agents, which they try to satisfy by applying matching dialogue operators. For example, if the user questions a claim made by the seller, the seller looks for dialogue operators that provide evidence for the claim. In case the user utters understanding problems related to a seller's statement, the seller searches for alternatives to convey the statement or for operators that elaborate on the statement. The implementation of the agents' dialogue behavior is based on the JAM agent architecture; see Huber (1999).

Essentially, there are two reasons for assigning the user the role of a helper. First, it allows us to restrict the set of possible interactions in a reasonable manner that facilitates the analysis of natural language utterances. Currently, we use spin, a java-based pattern-matching tool developed within the SmarkKom Project, both for the analysis and generation of natural language utterances; see Wahlster, Reithinger, and Blocher (2001). Second, as a helper agent, the user can respond to utterances of the seller or the buyer, but he/she is not expected to take the initiative in a dialogue. In the extreme case, he/she might even decide just to listen. The virtual agents continue with their discussion and provide new impulses to avoid the conversation from breaking down. At each point in time, the user has the option of joining the discussion again. The novelty of the approach lies in the fact that it allows the user to dynamically switch between active and passive viewing styles.

Conclusions

The objective of this article was to discuss various possibilities for constraining and structuring multimedia presentations that use lifelike characters. We started with a centralized planning approach that was used to automatically generate scripts for a single presenter or a team of presenters. Such a scripting approach facilitates the generation of well-structured and coherent presentations. However, it requires a clear separation of scripting and display time, which is only possible if all the information to be presented is known beforehand. However, there are also situations in which the underlying information dynamically changes at presentation display time. Examples include the presentation of live data as well as presentations that allow for an active participation of the user. For these applications, we propose a character-centered approach in which the scripting is done by the involved characters at presentation display time. The general idea here is not to specify the agents' behavior to the last detail but give them instructions instead that they can refine and work out at presentation run time. Table 1 provides an overview of the evolution of systems from script-based approaches to interactive performances.

Our future work will concentrate on the definition and evaluation of improvisational constraints for interactive performances, such as IMP II, to provide the user with an interesting and enjoyable experience. For example, we will investigate whether it makes sense to have the user appear as the seller's helper agent or allow him/her to switch roles during a performance. In our earlier applications, all involved characters were controlled by a computer-based behavior-selection mechanism. Consequently, we could start from the assumption that all involved characters followed the presentation constraints—either given by the script or the character's role and personality profile. If we involve human agents in such a scenario, such a behavior can no longer be guaranteed.

In our future work, we will address the following questions: (1) How can we make sure that the user understands his/her role and behaves accordingly? (2) What should be done if presentation constraints are violated—intentionally or unintentionally?

Acknowledgments

Part of this research was supported by the German Ministry for Education, Science, Research,

	Metaphor	Scripting Time	Script Producer	Structuring Principle	Degree of Predeterminism	Technical Realization
Noninteractive Presentation	Scripted talk	Prior to presentation, offline	Separate system component	Script centered	Fixed script, no interaction	Centralized planning component
Hyper-Presentation	Scripted talk, interactive narration	Switching between scripting and displaying	Separate system component	Script centered	Predefined choice points, expandable script	Centralized planning component
Noninteractive Presentation Teams	Script-based theater	Prior to presentation, offline	Separate system component	Plot centered	Fixed script, no interaction	Centralized planning component
Reactive Presentation Teams	Improvisational theater	During presentation, online	Involved characters	Character centered	Open ended	Distributed reactive planners
Interactive Performance	Improvisational theater	During presentation, online	Involved characters and user	Character centered, dramatic elements	Open ended	Distributed reactive planners

Table 1. From Script-Based Presentations to Interactive Performances.

and Technology (BMBF) under contracts 01 IW 806 and 01 IW 001 and by the European Community under contracts EP-25458 and EP-29335. We would like to thank Steve Allen, Stephan Baldes, Sabine Bein, Ralph Engel, Patrick Gebhard, Michael Kipp, Thomas Kleinbauer, Martin Klesen, Jochen Müller, Susanne van Mulken, Stefan Neurohr, Peter Rist, Ralph Schäfer, and Wolfgang Wahlster for their valuable contributions to the research described in this article.

References

André, E., and Rist, T. 1996. Coping with Temporal Constraints in Multimedia Presentation Planning. In Proceedings of the Eighth National Conference on Artificial Intelligence, 142–147. Menlo Park, Calif.: American Association for Artificial Intelligence.

André, E., and Rist, T. 1995. Generating Coherent Presentations Employing Textual and Visual Material. *Artificial Intelligence Review* (Special Issue on the Integration of Natural Language and Vision Processing) 9(2–3): 147–165.

André, E.; Herzog, G.; and Rist, T. 1997. Generating Multimedia Presentations for RoboCup Soccer Games. In *RoboCup '97: Robot Soccer World Cup I,* ed. H. Kitano, 200–215. New York: Springer-Verlag.

André, E.; Rist, T.; and Müller, J. 1999. Employing AI Methods to Control the Behavior of Animated Interface Agents. *Applied Artificial Intelligence* 13(4–5): 415–448.

André, E.; Rist, T.; van Mulken, S.; Klesen, M.; and Baldes, S. 2000a. The Automated Design of Believable Dialogues for Animated Presentation Teams. In *Embodied Conversational Agents,* eds. J. Cassell, J. Sullivan, S. Prevost, and E. Churchill, 220–255. Cambridge, Mass.: MIT Press.

André, E.; Binsted, K.; Tanaka-Ishii, K.; Luke, S.; Herzog, G.; and Rist, T. 2000b. Three RoboCup Simulation League Commentator Systems. *AI Magazine* 21(1): 57–65.

Badler, N. I.; Phillips, C. B.; and Webber, B. L. 1993. *Simulating Humans: Computer Graphics, Animation, and Control*. New York: Oxford University Press.

Binsted, K., and Luke, S. 1999. Character Design for Soccer Commentary. In *RoboCup-98: Robot Soccer World Cup II,* eds. M. Asada and H. Kitano, 22–33. New York: Springer-Verlag.

Cassell, J.; Bickmore, T.; Campbell, L.; Vilhjalmsson, H.; and Yan, H. 2000. The Human Conversation as a System Framework: Designing Embodied Conversational Agents. In *Embodied Conversational Agents,* eds. J. Cassell, J. Sullivan, S. Prevost, and E. Churchill, 29–63. Cambridge, Mass.: MIT Press.

Cassell, J.; Pelachaud, C.; Badler, N. I.; Steedman, M.; Achorn, B.; Becket, T.; Douville, B.; Prevost, S.; and Stone, M. 1994. Animated Conversation: Rule-Based Generation of Facial Expression, Gesture, and Spoken Intonation for Multiple Conversational Agents. *Computer Graphics* (SIGGRAPH '94 Proceedings) 28(4): 413–20.

Craig, S. D.; Gholson, B.; Garzon, M. H.; Hu, X.; Marks, W.; Wiemer-Hastings, P.; and Lu, Z. 1999. AUTOTUTOR and Otto Tudor. Paper presented at the Workshop on Animated and Personified Pedagogical Agents, Organized in Conjuction with the Ninth World Conference on Artificial Intelligence in Education, 19–23 July, LeMans, France.

Greimers, A., and Courtes, J. 1982. *Semiotics and Language: An Analytical Dictionary.*

Bloomington, Ind.: Indiana University Press.

Hayashi, M.; Gakumazawa, Y.; and Yamanouchi, Y. 1999. Automatic Generation of Talk Show from Dialog Using TVML. In Proceedings of Digital Convergence for Creative Divergence (ICCC'99), ed. K. Ono, 325–332. Washington, D.C.: ICCC.

Hayes-Roth, B., and van Gent, R. 1997. Story-Making with Improvisational Puppets. In Proceedings of Autonomous Agents, 92–112. New York: Association of Computing Machinery.

Höök, K.; Persson, P.; and Sjölinder, M. 2000. Evaluating Users' Experience of a Character-Enhanced Information Space. *Artificial Intelligence Communications* 13(3): 195–212.

Huber, M. 1999. JAM: A BDI-Theoretic Mobile Agent Architecture. In Proceedings of the Third Conference on Autonomous Agents, 236–243. New York: Association of Computing Machinery.

Jameson, A.; Schäfer, R.; Simons, J.; and Weis, T. 1995. Adaptive Provision of Evaluation-Oriented Information: Tasks and Techniques. In Proceedings of the Fourteenth International Joint Conference on Artificial Intelligence, ed. C. S. Mellish, 1886–1893. Menlo Park, Calif.: International Joint Conferences on Artificial Intelligence.

Johnstone, K. 1989. *IMPRO: Improvisation and the Theatre*. New York: Routledge and Kegan Paul.

Kitano, H.; Asada, M.; Kuniyoshi, Y.; Noda, I.; Osawa, E.; and Matsubara, H. 1997. RoboCup: A Challenging Problem for AI. *AI Magazine* 18(1): 73–85.

Klesen, M.; Szatkowski, J.; and Lehmann, N. 2001. A Dramatized Actant Model for Interactive Improvisational Plays. In *Proceedings of the Third International Workshop on Intelligent Virtual Agents,* eds. A. deAntonio, R. Aylett, and D. Ballin, 181–194. Lecture Notes in Artificial Intelligence 2190. Heidelberg: Springer-Verlag.

Laurel, B. 1991. *Computers as Theatre*. Reading, Mass.: Addison-Wesley.

Lester, J. C.; Voerman, J. L.; Towns, S. G.; and Callaway, C. B. 1999. Deictic Believability: Coordinated Gesture, Locomotion, and Speech in Lifelike Pedagogical Agents. *Applied Artificial Intelligence* 13(4–5): 383–414.

Mateas, M. 1997. An Oz-Centric Review of Interactive Drama and Believable Agents. Technical Report, CMU-CS-97-156, School of Computer Science, Carnegie Mellon University.

Microsoft. 1999. Microsoft Agent: Software Development Kit. Redmond, Wash.: Microsoft Press. Also available at http://microsoft.public.msagent..

Nitta, K.; Hasegawa, O.; Akiba, T.; Kamishima, T.; Kurita, T.; Hayamizu, S.; Itoh, K.; Ishizuka, M.; Dohi, H.; and Okamura, M. 1997. An Experimental Multimodal Disputation System. Paper presented at the Workshop on Intelligent Multimodal Systems, Organized in Conjunction with the Fifteenth International Joint Conference on Artificial Intelligence, 23–29 August, Nagoya, Japan.

Noma, T., and Badler, N. 1997. A Virtual Human Presenter. Paper presented at the Workshop on Animated Interface Agents: Making Them Intelligent, Organized in Conjunction with the Fifteenth International Joint Conference on Artificial Intelligence, 23–29 August, Nagoya, Japan.

Rickel, J., and Johnson, W. L. 1999. Animated Agents for Procedural Training in Virtual Reality: Perception, Cognition, and Motor Control. *Applied Artificial Intelligence* 13(4–5): 343–382.

Rist, T.; Brandmeier, P.; Herzog, G.; and André, E. 2000. Getting the Mobile Users In: Three Systems That Support Collaboration in an Environment with Heterogeneous Communication Devices. In Proceedings of Advanced Visual Interfaces 2000, eds. V. D. Gesù, S. Levialdi, and L. Tarantino, 250–254. New York: Association of Computing Machinery.

Thalmann, N. M., and Kalra, P. 1995. The Simulation of a Virtual TV Presenter. In *Computer Graphics and Applications,* 9–21. Singapore: World Scientific.

Wahlster, W.; Reithinger, N.; and Blocher, A. 2001. SmartKom: Multimodal Communication with a Life-Like Character. Paper presented at Eurospeech 2001, Seventh European Conference on Speech Communication and Technology, 3–7 September, Aalborg, Denmark.

Walker, M.; Cahn, J.; and Whittaker, S. J. 1997. Improving Linguistic Style: Social and Affective Bases for Agent Personality. In Proceedings of Autonomous Agents'97, 96–105. New York: Association of Computing Machinery.

Yabe, J.; Takahashi, S.; and Shibayama, E. 2000. Automatic Animation of Discussions in Usenet. In Proceedings of Advanced Visual Interfaces 2000, eds. V. D. Gesù, S. Levialdi, and L. Tarantino, 84–91. New York: Association of Computing Machinery.

Elisabeth André is a full professor of computer science at the University of Augsburg, Germany. Prior to that, she worked as a principal researcher at DFKI GmbH. She is the chair of the Association of Computational Linguistics Special Interest Group on Multimedia Language Processing (SIGMEDIA). Furthermore, she is on the editorial boards of *Artificial Intelligence Communications* (AICOM), *Cognitive Processing* (International Quarterly of Cognitive Science), and *Universal Access to the Information Society* (UAIS) and is the area editor for intelligent user interfaces of the *Electronic Transactions of Artificial Intelligence* (ETAI). Her e-mail address is andre@dfki.de.

Thomas Rist is currently a senior researcher at the German Research Center for Artificial Intelligence (DFKI). His areas of experience include knowledge-based graphics, intelligent multimedia systems, lifelike interface agents, and AI techniques. He is currently a member of the European i3 (Intelligent Information Interfaces) coordination group and is in charge of several managing tasks. Within the ERCIM Computer Graphics Network Task II Group, he worked on an initiative for the development of a standard reference model for intelligent presentation systems. His e-mail address is rist@dfki.de.