

From Multimodal Presentations to Interactive Performances

Elisabeth André
Universität Augsburg
Eichleitnerstr. 30
D-86135 Augsburg
Germany
andre@informatik.uni-augsburg.de

Thomas Rist
DFKI GmbH
Stuhlsatzenhausweg 3
D-66123 Saarbrücken
Germany
rist@dfki.de

1 Abstract

Lifelike characters, or animated 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. In this contribution, we argue in favor of a shift from applications with single presentation agents towards flexible performances given by a team of characters as a new presentation style. 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 may take on an active role as well.

1.1 Keywords

Simulated Dialogues, Presentation Teams, Interactive Performances

2 Introduction

A growing number research projects have started to develop 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 the user 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 3D body models while recorded voices or synthesized speech and sound determine their audible appearance, e.g. see [Lester et al. 99] or [Rickel et al. 99].

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, i.e., 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. So it comes as no surprise that behavior scripting, in one way or another, has been widely used in projects that deal with interface characters. For instance, a straightforward approach is to equip the character with a library of manually authored scripts that determine what the character might do in a certain situation. At runtime, 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 upon an existing general-purpose script-based programming language. For instance, the Microsoft Agent characters [Microsoft 99] can be easily scripted either in Visual Basic or in Java Script allowing the script writer to use the standard control structures of these languages like conditional statements or loop constructs. As an alternative to character specific adjuncts to programming languages, XML-compliant scripting languages may be defined. In any case, the script may be seen as a kind of an 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 in order to avoid characters that behave in a monotonic and too predictable way. Furthermore, the manual scripting of presentation agents can become quite complex and error-prone since synchronization issues have to be considered. In order to avoid extensive scriptwriting 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 been successfully applied to build a number of applications in which information is conveyed either by a single presenter

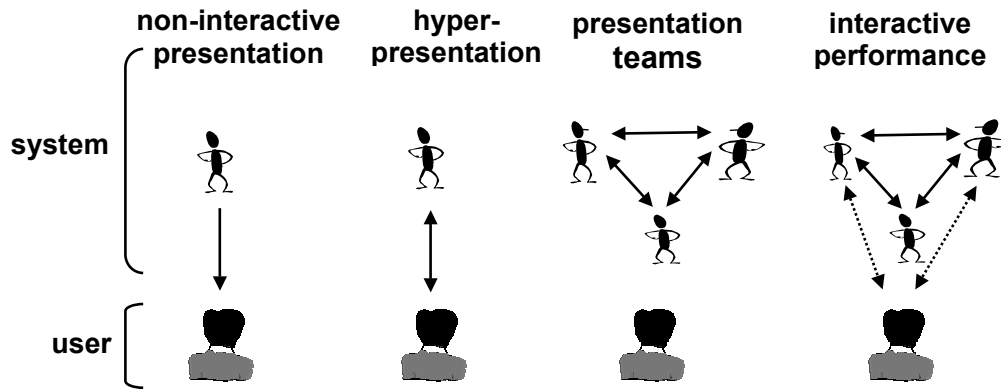


Figure 1: Agent-user relationships in different scenarios

or likewise by 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 whether or not all information to be conveyed by a character is available before a presentation is started. Another aspect is the kind of user interactions that shall be supported during the display of a presentation.

In this contribution, we revisit some of our past and ongoing projects in order 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. Though 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 as no user interaction is foreseen at display time. In contrast, the second scenario may be compared to a hypermedia presentation in which the user may 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 in order to convey information to the observing audience. However, no user intervention is foreseen during a performance. This is in contrast to the fourth scenario where we have the setting of an interactive performance in which the user can take on an active role in the performance. From a technical point of view the fourth scenario is perhaps most challenging as 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 may be used in a corresponding presentation system.

3 Deploying a Single Presenter

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

In our earlier work, we have conducted two projects to develop systems that fall into this category: the PPP (Personalized Plan-Based Presenter) project and the AiA (Adaptive Communication Assistant for Effective Infobahn Access) project [André et al. 99]. In PPP we addressed the automated generation of instructions for the operation of technical devices which were delivered to the user by an animated agent, the so-called PPP-Persona. For instance, to explain how to switch on a device, PPP-Persona showed the user a picture of the device, pointed to the on-off switch while instructing him or 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 the user's request, e.g. 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. In PPP and AiA, the agents' behavior is determined by a script which 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 since it would require to anticipate the needs of all potential users and preparing presentations for them. For instance, in PPP the user could specify a time limit for the presentation. Depending on the setting of this parameter, the generated instructions 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.

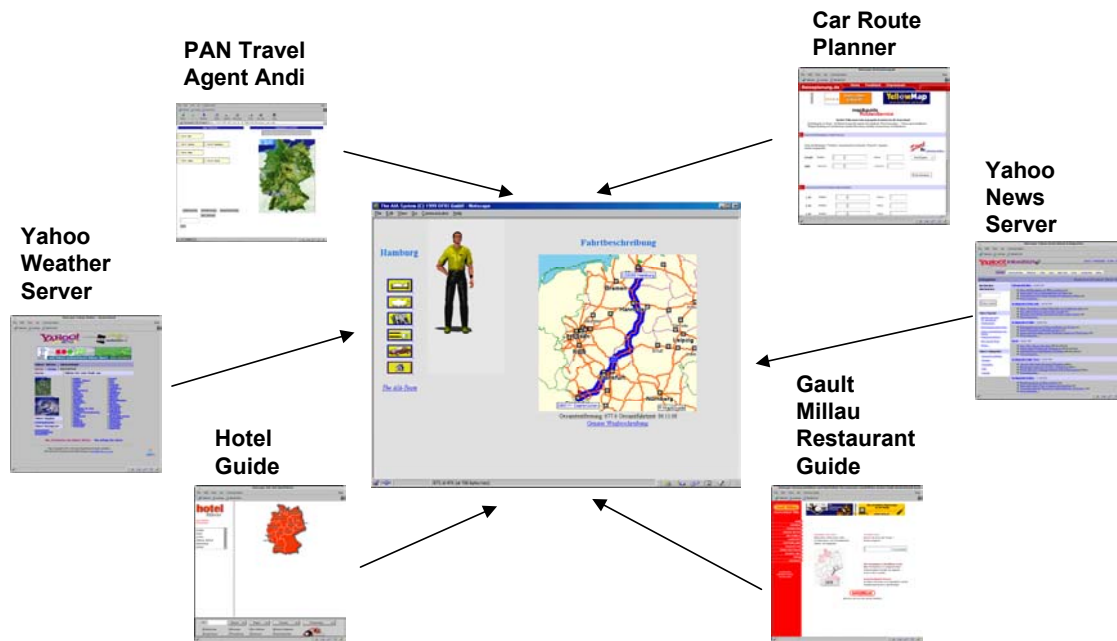


Figure 2: The AiA Travel Agent

In the case of the AiA system, manual scripting is even more unpracticable since 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 and formalize action sequences for composing multimedia material and designing scripts for presenting this material to the user as operators of a planning system [André et al. 99]. The effect of a planning operator refers to a complex communicative goal (e.g. 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 in order to achieve this goal (e.g. to show an illustration of a certain object and to 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 will be 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 et al. 99]).

4 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 may serve as a personal guide or assistant in information spaces like 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. 99] 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.

Based on this observation, we started to investigate a new style of presentation which conveys information in a less direct manner. We employ 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 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. Finally, the single members of a presentation team can serve as indices which help the user to organize the conveyed information.



Figure 3: The Inhabited Market Place

For instance, we may convey meta-information, such as the origin of information, or to 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. This presentation style is currently explored within the Inhabited Market Place [André et al. 00].

As suggested by the name, the Inhabited Market Place is a virtual place in which seller agents provide product information to potential buyer agents. Figure 3 shows a dialogue between several seller and buyer agents. In addition to the agents provided by the MSAgentRing, we created our own agents, such as the character shown in the middle, see <http://www.msagentring.org>

One of the salesmen is trying to convince the buyers of the potential benefits of a certain car. 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 which may be chosen by the user as well. To illustrate this, we list two dialogue fragments that have been generated for different initial parameter settings (see Figure 4).

The two dialogues partially discuss the same car attributes, but from different points of view. In both cases, one of the buyers criticizes the high gas consumption of the car. But in the first case, it is thinking of the environment while, in the second case, it is concerned about the high costs. With regards to recent attempts in the area of collaborative browsing, the use of multiple presenters would also allow for performances that account to a certain extent for the different interest profiles of a diverse audience.

As in the AiA project, 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 of all, information is no longer presented by a single agent which stands for the presentation system, but instead 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 a dialogue between several characters to be observed by him or 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 audio-visual appearance.

Peedy: <i>How much gas does it consume?</i>	Peedy: <i>How much gas does it consume?</i>
Robby: <i>It consumes 8 liters per 100 km.</i>	Robby: <i>It consumes 8 liters per 100 km.</i>
Peedy: <i>Isn't that bad for the environment?</i>	Peedy: <i>I'm worrying about the running costs.</i>
Robby: <i>Bad for the environment? It has a catalytic converter. It is made of recyclable material.</i>	Robby: <i>Forget the running costs. Think of the prestige.</i>

Figure 4: Dialogue Samples

To account for this, 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 contents of an utterance, but also its communicative function, such as taking turns or responding to a question, see also [Cassell et al. 00]. The character’s profile is considered by treating it as an additional filter during the selection instantiation and rendering of dialogue strategies.

The Inhabited Market Place allows for flexible role assignment of the agents. The user, however, has only the option of joining the presentation as a passive observer.

4 Interactive Performances

Currently, we are investigating possibilities to involve the user in presentation scenarios as well. The basic idea is provide the user with the option of taking an active role in the performance if she or he wishes 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. 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.

As a first example of a system that presents information in the style of an interactive performance we are currently developing a further version of the Inhabited Market Place. The presentation task and scenario are similar to the original version. Following the principle idea described above, our goal is now to allow the user to step into the role of an accompanying buyer or a seller who can pose questions, support, reinforce or reject arguments made by the other agents. In Figure 3, the user who is shown from behind has just made a comment on the car.

Since the agents have to dynamically respond to user interactions, it is no longer possible to pre-script utterances. For such scenarios, we propose a character-centered approach. Instead of specifying the agents’ behavior to the

last detail, we just provide a character with a description of its role and profile according to which it has to behave at presentation runtime. Technically, we have realized this approach by a system of distributed planners.

5 Conclusions

The objective of this paper was to discuss various possibilities for constraining and structuring multimedia presentations that employ 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. But, it requires a clear separation of scripting and display time. This is only possible if all the information to be presented is a priori known. However, there are also situations in which the underlying information dynamically changes at presentation display time. For these applications, we propose a character-centered approach in which the scripting is done by the involved characters at presentation display time.

6 Acknowledgements

The work described here has been partially funded by the BMBF project MIAU and the EU projects MagiCster and NECA.

7 References

- André, E., T. Rist, and J. Müller. 1999. Employing AI methods to control the behavior of animated interface agents. *Applied Artificial Intelligence* 13:415–448.
- André, E., Rist, T., van Mulken, S., Klesen, M. and Baldes, S. 2000. The Automated Design of Believable Dialogues for Animated Presentation Teams. In: Cassell et al. (eds.): *Embodied Conversational Agents*, 220-255, Cambridge, MA: MIT Press.
- Cassell, J., T. Bickmore, L. Campbell, H. Vilhjalmsson, and H. Yan. 2000. The human conversation as a system framework: Designing embodied conversational agents. In: Cassell et al. (eds.): *Embodied Conversational Agents*, 29-63, Cambridge, MA: MIT Press.
- Craig, S. D, B. Gholson, M. H. Garzon, X. Hu, W. Marks, P. Wiemer-Hastings, and Z. Lu. 1999. Auto Tutor and Otto Tudor. In *AIED-Workshop on Animated and Personified Pedagogical Agents*, 25–30. Le Mans, France.
- Microsoft. 1999. *Microsoft Agent: Software Development Kit*. Redmond, Wash.: Microsoft Press. <<http://microsoft.public.msagent>>
- Lester, J. C., J.L. Voerman, S.G. Towns, and C.B. Callaway. 1999. Deictic Believability: Coordinated Gesture, Locomotion, and Speech in Lifelike Pedagogical Agents. *Applied Artificial Intelligence* 13:383–414.
- Rickel, J., and W.L. Johnson. 1999. Animated Agents for Procedural Training in Virtual Reality: Perception, Cognition, and Motor Control. *Applied Artificial Intelligence* 13:343–382.

