Building smart embodied virtual characters

Thomas Rist, Elisabeth André

Angaben zur Veröffentlichung / Publication details:

Rist, Thomas, and Elisabeth André. 2003. "Building smart embodied virtual characters." In *Smart Graphics: Third International Symposium on Smart Graphics, SG 2003 Heidelberg, Germany, July 2–4, 2003*, edited by Andreas Butz, Antonio Krüger, and Patrick Olivier, 123–30. Berlin [u.a.]: Springer. https://doi.org/10.1007/3-540-37620-8_12.

Building Smart Embodied Virtual Characters

Thomas Rist[1] and Elisabeth André[2]

[1]DFKI GmbH, Stuhlsatzenhausweg 3, Saarbrücken Germany
rist@dfki.de
[2]Laboratory for Multimedia Concepts and Applications, Universität Augsburg, Germany
andre@informatik.uni-augsburg.de

Abstract. Embodied conversational characters are autonomous, graphically embodied virtual creatures that live in a 2D or 3D virtual environment. They are able to interact intelligently with human users, other characters, and their digital environment. While for decades research has concentrated on geometric body modelling and the development of animation and rendering techniques for virtual characters, other qualities have now come in focus as well, including the provision of conversational skills as well as the simulation of believable behavior including affect and peculiarities induced by individual personality traits. As a consequence, the domain of virtual characters has become much more diverse and now encompasses a wide range of disciplines, from computer graphics and animation to AI and more recently also psychology, sociology as well as design and arts. The current paper discusses a number of design issues that arise when building an application with one or more embodied characters. By means of selected sample applications we also illustrate a yet ongoing development of animated presentation agents starting with TV-style information presenters to highly interactive multi-character scenarios in which information is conveyed to the user in the form of multi-party conversations.

1 Introduction

The creation of virtual humans is an old dream of mankind – indeed much older than computer science, AI, computer graphics and animation. With the advent of powerful but nonetheless affordable multimedia workstations, in the early nineties some groups have started research on animated virtual characters in order to deploy them in information presentation tasks, and in some cases, to promote them even as a general and ultimate metaphor for human-user interaction. 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 with less native interaction techniques. Furthermore, well designed characters show great potential for making interfacing with a computer system more enjoyable.

The development of a presentation agent or a user interface with conversational embodied characters requires a number of design decisions and implementation tasks.

2 Character Embodiment

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 for their visual realization while recorded voices or synthesized speech and sound determine their audible appearance. In our own projects we have also experimented with a wide range of different character types and different graphical realizations. Fig. 1 shows sample characters arranged in a two-dimensional design space. So-called video-agents or video actors are located in the lower left-hand corner. Their production usually requires a human actor that performs gestures which are recorded and stored in a library of video clips. By concatenating these clips, complex behaviors can be assembled and played back.

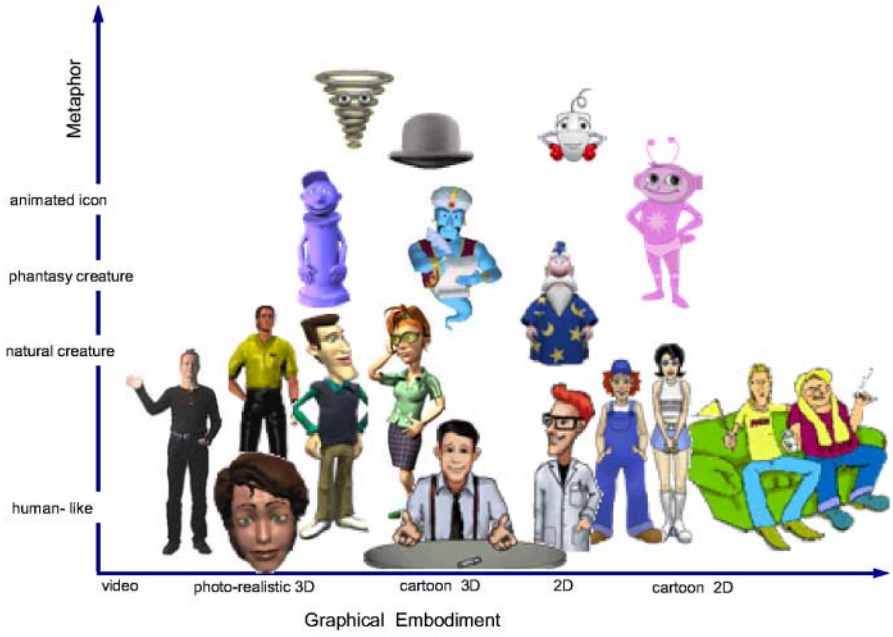

Fig. 1. Different kinds of characters[1] and different forms of graphical embodiment.

Audio-visual attractiveness, however, does not make characters smart. Rather, the success of an animated character in terms of user acceptance and interface efficiency very much depends on the character's communication skills and the impression of its overall behavior. Moreover, empirical studies show that there are dependencies between a character's appearance on the one hand and its skills on the other hand

[1] Characters shown in this diagram have been designed by Peter Rist except the Microsoft characters Genie and Merlin, the 3D talking head Greta which is a courtesy of Catherine Pelachaud, and the "video avatar" which is based on video recordings of Andreas Butz.

[5,9]. One extreme are photo-realistic humans (video or animated 3D models). In this case, humans who interact with such characters often have high expectations in the character's skills, especially if spoken language is used as the lead interaction modalitiy. To avoid frustrating the user, it is therefore often advisable to rely on cartoon-style characters or use a fantasy-style character design as these forms of embodiments may raise lower expectations.

3 Types of Conversational Settings

The choice of domain, tasks, and conversational setting imposes constraints on any prototype development. For instance, in the area of intelligent information presentation with animated characters we observe an ongoing evolution of systems as illustrated in Fig. 2. The first setting refers to applications in which a single character is deployed to present information. From the point of view of the user a generated presentation appears quite similar to watching a TV-news speaker or to the display of a video clip because no interaction is foreseen at display time. In contrast, the second setting is typical for applications with characters that are able to converse with a user in some sort of a dialogue (e.g., via spoken or typed natural language, or based on dynamically configured menus). Moving on to the third setting actually means a shift from a face-to-face character-user setting to a user-as-observer setting. That is, two or more characters talk to each other on the screen 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 an open multi-party dialogue setting which allows for both reactive and proactive user participation. Technically speaking the fourth scenario is quite challenging as one has to resolve on an operational level the conflict between predestination and freedom of interaction. To complicate things even further, one can think of multi-party settings with multiple characters and multiple users. However, up to now such settings remain a great challenge since in this case the characters must also be able to overhear and understand conversations among the human partners.

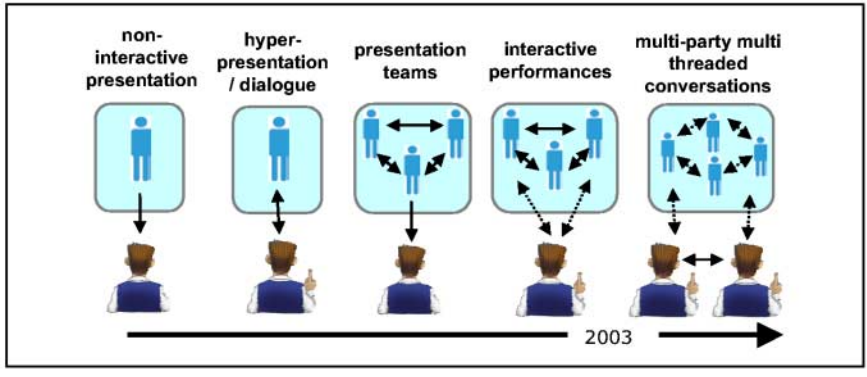

Fig. 2. Different kind of conversational settings for systems with animated characters

4 Behavior Control

Most of the current systems with animated characters distinguish between a character's embodiment and a behavior control component. Some relate this distinction to the biological body/brain dichotomy. Others take a more technically oriented view and associate embodiment with an animation engine (often called *character player*) while behavior control is related to some sort of automated behavior generation, often based on AI techniques, such as task-oriented hierarchical planning, or the simulation of certain aspects of human-like cognition. Following the latter distinction, observable behavior of a character can be regarded as the execution of a script in the character player. Thereby, a script 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. 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 build on top of an existing general-purpose script-based programming language. For instance, the Microsoft Agent characters 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 character scripting languages have been be defined, such as VHML (www.vhml.org) or MPML (www.miv.t.u-tokyo.ac.jp/MPML/). In any case, the script may be seen as a kind of an application programming interface (API) for the character player that allows to specify the agents behavior at a certain level of abstraction.

5 Approaches to Automated Character Scripting

A particular 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 characters 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 own 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 [2]. This approach has been successfully applied to build a number of applications in which information is conveyed either by a single presenter 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.

When developing a system with more than one character one has the choice between a centralized script generator (which might rely on an AI planner or another generative approach, such as Lester's sequencing engine [4]) on the one hand, and a distributed approach on the other hand. Taking a centralised approach, the generator determines the behavior of all involved characters. Such a scripting approach facilitates the generation of well-structured and coherent presentations, however, 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. Examples include the presentation of live data as well as presentations which 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 specify the agents' behavior to the last detail, but give them instructions instead that they may refine and work out presentation runtime. Table 1 provides an overview of the different systems from script-based approaches to interactive performances.

Table 1. Overview of different scripting approaches

	Metaphor	**Scripting Time**	**Script Producer**	**Structuring Principle**	**Degree of determinism**	**Technical Realization**
tv-style presentations	scripted talk	prior to presentation, offline	separate system component	script-centered	fixed script, no interaction	centralized planning component
hyper-presentations & query /answer dialogs	scripted talk, interactive narration	switching between scripting / displaying	separate system component	script-centered	pre-defined choice points, expandable script	centralized planning component
Non-interactive pres. teams	script-based theatre	prior to presentation - offline	separate system component	plot-centered	fixed script, no interaction	centralized planning component
Reactive Presentation teams	improvisational theatre	during presentation- online	involved characters	character-centered	open-ended	distributed reactive planners
Interactive performances	improvisational theatre	during presentation- online	involved characters and user	character-centered	open-ended	distributed reactive planners

6 Architectures for Systems with Characters

Most systems that deploy life-like characters make a concrete commitment to one of the conversational settings illustrated in Fig. 2 and reflect this decision in a particular system architecture. However, if later on the desire emerges to support other conversational settings as well, an almost full re-implementation of the application often becomes unavoidable.

In a recent project, called MIAU [7], we wondered whether it is possible to develop a single platform which (a) can be used to construct a broad range of character applications, (b) even allows to switch on-the-fly between director- vs. character-centred scripting approaches, and (c) supports a clear separation between the specification of scripting knowledge (being a knowledge-engineering task), and the required computational machinery for behaviour generation (being an implementation task).

The architecture of the resulting MIAU platform is shown in the upper part of Fig. 3. We adopt the metaphorical distinction between a character's brain and a character's body which is typically reflected in an architecture by a separation of components for behavior planning on the one hand, and a character player component on the other hand. We further assume that the player will receive commands for direct execution from the superordinate behavior determining part. The MIAU platform itself abstracts from the player technology used for character animation, speech synthesis, and receiving user input. Rather, the platform consists of the following components:

For each character $C_1 \ldots C_n$ MIAU foresees so-called *character components* containing a separate behavior planner as well as a separate interaction manager. The behavior planner has the task to decompose complex discourse goals into basic acts that can be executed by the character player. The interaction manager, in a certain sense, corresponds to a dialogue manager as found in NL dialogue systems since it is responsible for keeping book on interaction states and the interaction history. However, in the MIAU platform the interaction manager realizes a character's internal interface for communication with other system components by means of formalized communication acts.

To allow a user to alter settings for the performance, to take an active part in a performance, or even to intervene in the role of a director or co-director, the platform also incorporates an U box, the so-called *user component*.

However, since this time the user decides on what to do, we don't foresee a planner, but an input analyzer for mapping user input onto formalized communication acts. The internal communication with other system components is handled by the interaction manager similar to the structure of a character component. In case the user is represented in the scenario by an embodied (and possibly animated) avatar, the avatar may be employed to audio-visually indicate his or her input activity. For instance, if a text widget is used for acquiring user input, the user's avatar may speak the input sentence. Currently, we restrict ourselves to a single participating user only. Nevertheless it seems possible to extend the architecture for multi-user, multi-character scenarios by adding more user components.

In addition to character components and the user component, MIAU incorporates also a D-box, the so-called *director component*. In contrast to the characters, the director does not participate in performances and therefore has no embodiment. Rather, this component is foreseen to enable some degree of centralized control on the

overall interactive performance. While the director also comprises a planner, this time the planner is used in order to influence the course of the performance depending on the degree of centralized control wanted for a certain application. Internal communication with other components is again handled via the interaction manager.

Finally, the MIAU platform comprises a message board which is shared among the different components for the exchange of internal communication acts.

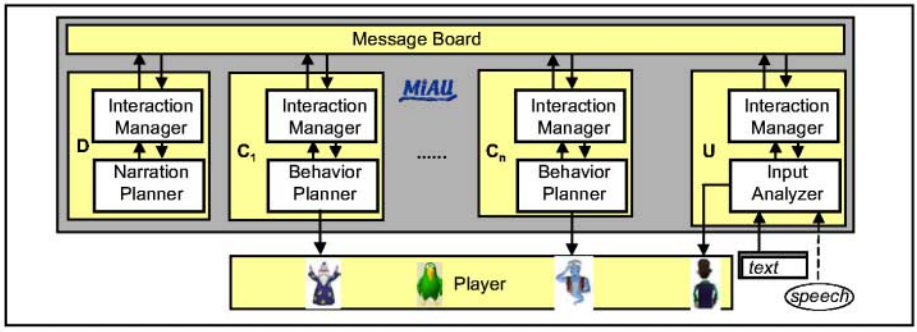

Fig. 3. Architecture of the MIAU platform for the set-up of interactive applications with a flexible number of conversational characters.

The MIAU platform is currently tested in a number of different projects (cf. Fig. 4) dealing with life-like characters including two variants of the eShowroom [1,2] the interactive CrossTalk installation [3], and Avatar Arena [8].

Fig. 4. From left to right: the eShowroom, its interactive version, CrossTalk, and Avatar Arena

The eShowroom (first screenshot from left in Fig. 4) is an electronic car showroom in which either a single character or a team of characters are deployed to inform a user about the features of a certain car. In the interactive version of the eShowroom, a user can also take part in a "car-talk" with one or more characters (second screenshot from left in Fig. 4). The interactive CrossTalk installation (third screenshot from left in Fig. 4). has been designed for public spaces and was shown during the CeBit 2002 and the IST 2002 exhibitions. The basic idea is that a virtual stand hostess invites the user to watch a car-sales performances given by the virtual actors Tina and Ritchie who live on the opposite screen. Finally, Avatar Arena is a test-bed for the simulation of changing inter-personal relationships during negotiation dialogues. The simulation is based on socio-psychological theories of cognitive consistency [6] and captures some aspects of group dynamics.

7 Conclusions

In this paper we addressed a number of design issues that arise when building an application with one or more embodied characters. While powerful character players have become available there is no an increasing interest in research on modelling behavior and conversational skills. We argued that the choice of the conversational setting places constraints on both, the scripting approach and the architecture of a character system. We then sketched the MIAU platform that is flexible enough to switch back and forth between different conversational settings. Such a flexibility may pave the way for new engaging presentation formats. However, building an interesting character application that will be appreciated by their users is foremost a challenging design task. For instance, the number of characters, their roles as well as the foreseen role for the user need to be carefully chosen from one application to another.

References

[1] 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, J., Sullivan, J., Prevost, S. and Churchill, E. (eds.): *Embodied Conversational Agents*, 220–255, Cambridge, MA: MIT Press.

[2] André E, and Rist T (2001) Controlling the Behavior of Animated Presentation Agents in the Interface: Scripting vs. Instructing. AI Magazine 22(4):53–66.

[3] Gebhard P, Kipp M, Klesen H, and Rist T (2003) What Are They Going to Talk About? Towards Life-Like Characters that Reflect on Interactions with Users, In Proc. of Technologies for Interactive Digital Storytelling and Environment, TIDSE 2003, Fraunhofer IRB Verlag, pp. 380–387.

[4] Lester, J, Voerman, JL, Towns, SG and Callaway, C (1999). Deictic Believability: Coordinated Gesture, Locomotion, and Speech in Lifelike Pedagogical Agents. in: Applied Artificial Intelligence 13:383–414, 1999

[5] McBreen H, Shade P, Jack MA, and Wyard PJ (2000) Experimental assessment of the effectiveness of synthetic personae for multi-modal e-retail applications. In Proc. of Agents 2000, pp. 39–45

[6] Osgood C and Tannenbaum P (1955) The principle of congruity in the prediction of attitude change. Psychology Review, pp. 62, 42–55.

[7] Rist T, André E, Baldes S (2003) In: Proc of IUI'03, A Flexible Platform for Building Applications with Life-Like Caracters ACM Press, pp 158–165.

[8] Rist T. Schmitt M, Pelachaud C, and Bilvi M (2002) Towards a Simulation of Conversations with Expressive Embodied Speakers and Listeners. In Proc. of Computer Animation and Social Agents, CASA 2003.

[9] van Mulken S, André E and Müller J (1998) The Persona Effect: How Substantial Is It? In Proc. of Human Computer Interaction Conference, Berlin, Springer, pp. 53–66.