What Are They Going to Talk About? Towards Life-Like Characters that Reflect on Interactions with Users

Patrick Gebhard, Michael Kipp, Martin Klesen, Thomas Rist

German Research Center for Artificial Intelligence (DFKI)
Stuhlsatzenhausweg 3
66123 Saarbrücken, Germany
{gebhard, kipp, klesen, rist}@dfki.de
http://www.dfki.de/crosstalk/

Abstract. We first introduce CrossTalk, an interactive installation with animated presentation characters that has been designed as an interactive installation for public spaces, such as an exhibition, or a trade fair. The installation relies on what we call a meta-theater metaphor. Quite similar to professional actors, characters in CrossTalk are not always on duty. Rather, they can step out of their roles, and amuse the user with unexpected intermezzi and rehearsal periods. From the point of view of interactive story telling, CrossTalk comprises at least two interesting aspects. Firstly, it smoothly combines manual scripting of character behavior with an approach for automated script generation. Secondly, the system maintains a context memory that enables the characters to adapt to user feedback and to reflect on previous encounters with users. The context memory is our first step towards characters that develop their own history based on their interaction experiences with users. In this paper we briefly describe our approach for the authoring of adaptive, interactive performances, and sketch our ideas to enrich conversations among the characters by having them reflect on their own experiences.

1 Introduction

Animated conversational characters are increasingly used in a wide range of different application areas, including virtual training environments [9], [17], interactive fiction [10], [11] and storytelling systems [18], as well as in e-commerce applications where computer agents play the role of product presenters and sales assistants. Our work at DFKI builds on prior work on embodied conversational agents [6] and presentation agents [1], [2], [3], [4]. One of the conversational characters developed at DFKI is Cyberella, a virtual receptionist which provides visitors with information about staff members and projects [7]. Cyberella assumes a setting in which the agent addresses the user directly like in human face-to-face conversations. However, there are situations in which direct agent-user communication is not necessarily the most effective and most convenient way to present information. Inspired by the evolution

of TV commercials over the past 40 years, our group has discovered role-plays with synthetic characters as a promising format for presenting information. We have therefore proposed a shift from single character settings towards interactive performances by a team of characters as a new form of presentation [2]. The use of multiple characters allows to convey social aspects such as interpersonal relationships between emotional characters [13], [16].

Seeking for potential exhibits that could be demonstrated at the CeBIT 2002 computer fair, the idea arose to have our pre-existing Cyberella character present another the so-called Inhabited Market Place, another pre-existing system that conveys information about products to the user in the form of a simulated sales conversation. The combination of the two systems resulted in the interactive CrossTalk installation, [5], [15]. Figure 1 shows the installation and a close-up of the three screens.

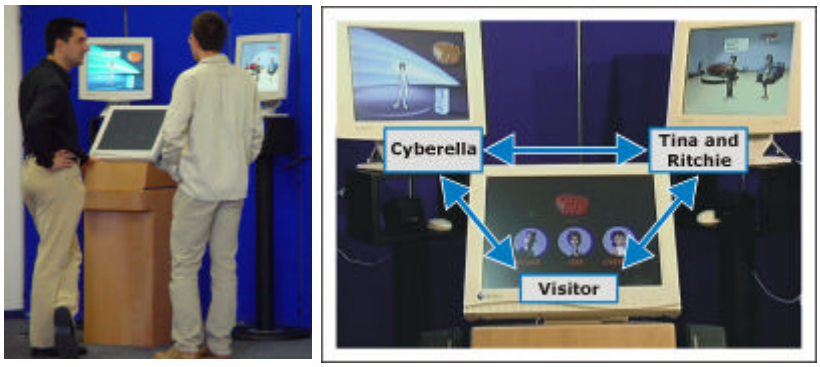

Fig. 1: The CrossTalk interactive installation

A major conceptual contribution of CrossTalk is the introduction of a meta-theater metaphor, taking forward the "computers-as-theatre" paradigm that has been originally presented by Brenda Laurel [10] and since applied by others too, e.g., [8]. Our motivation for introducing the meta-theater metaphor is a practical one. It also allows us to emulate small talk between characters which then becomes yet another performance or "meta-theater" [5]. The purpose for this "off-duty" activity, quite natural for humans, is twofold: (1) It attracts and binds the attention of passers-by, and (2) gives our agents the authenticity of real human actors, conveying the impression that they are permanently alive.

2 Authoring Interactive Scenes

Using the theater as a metaphor we equated our agents with human actors. But where does the script come from which defines the verbal and nonverbal behavior? There are basically two approaches: Either a human author writes extensive libraries of scripts that can be executed by the characters. In this case, the quality of the resulting performances is just a function of skills and ingeniouity of the script writer.

Likewise, a system can retrieve presentation content from a database or any other source, and deploy some mechanisms for automated script generation. In this case, the system acts itself as a play writer, but in contrast to human authoring, script generation is much more flexible since it can be done on-the-fly at runtime [3]. Ideally, an authoring system supports both scripting approaches. To this end, it should entail automated script generation but at the same time provide creative content authors (often non-programmers) with tools for the creation of rich, compelling content. In addition, the system should provide a platform that enables to coherently interweave manual scripted scenes with scenes that have been composed automatically.

Scenes that are automatically generated at runtime rely on a representation of the subject-matter domain and on dialogue model that is compatible with the particular genre, e.g., car sales dialogues. In addition user-defined parameters, such as the roles and personality traits of the characters are taken into account in the generation process. In CrossTalk all conversations that relate to the discussion of a product are generated automatically. For instance, when in "on-duty mode", the two virtual actors Tina and Ritchie (living in the left-hand screen shown in the close-up of Figure 1) perform a typical customer-buyer sales talk about a particular car. Eventually all questions and answers that relate to the car under discussion are grounded in the entries of a database that holds a representation of the car attributes.

In CrossTalk a pre-scripted scene entails a some dialogue moves on a certain topic. In addition to a specification of the corresponding verbal utterances, a scene specification may also include special annotations to control the agent's non-verbal behavior, such as gaze, gesture, and body posture, as well as system control commands (see Figure 2). For the CrossTalk installation a large corpus of pre-scripted scenes (more than 220 for English and German each) has been assembled. A pre-scripted scene captures some dialogue moves on a particular topic. In addition, a scene specification may include special tags to control the agent's non-verbal behavior, such as gaze, gesture, and body posture, as well as system control commands (see Fig. 2). The scenes can be written using an ordinary word processor and are then translated by our DialogueCompiler into scene plans. This allows us to interweave such pre-scripted scenes with scenes automatically generated at runtime.

Scene: OFF-Chat stage-direction

... ...

Ritchie: [TINA AS_LookLeft] Ok, if you are interested leave me your number. [V_LookToCy]

Tina: Well, <Pau=300> ok. [RITCHIE V_LookToActor] Sounds ... great. [AS_Glasses] I'll think about it.

Cyberella: [GS_Chide] My agent will contact you.

Ritchie: Yeah. Sure. [GS_DoubtShrug] All right.

Fig. 2: Pre-scripted scene.

We realized that apart from the time-consuming scripting effort [15] it is much more of a challenge to create, maintain and extend the *structure* of the story. Technically speaking, every story contains a logical scene flow that defines the transitions

between scenes. This has to be modeled by the author and interpreted by the system at run-time. Usually, the scene flow is hard-wired and not reusable for other performances. Also, since stories are created by authors, possibly non-programmers, they depend on others to implement the story's logical framework. For the specification of the scene flow we use a cascaded finite state machine in which states refer to scenes that have been either pre-authored or which are generated automatically on-the-fly. Cascaded finite state machines allow the shared use of modules (part of a scene flow), similar to subroutines in a programming language. This simplifies the modeling process in case of, for instance, repeated patterns of agent-user interactions like simple yes/no questions. Also, such modules can be reused for other applications. Figure 3 shows a simple example in the CrossTalk scene flow: The launched system will start in *off-duty* mode modeled as a supernode with no scene attached (upper box). In a next step, the sub-node *idle0* that is declared starting node will be processed. After performing its attached scene, node *idle1* or node *idle2* will be processed with probability 0.5 each. If a visitor enters the CrossTalk installation, the currently processed sub-node of supernode *off-duty* will be interrupted using the interrupt edge *visitor_detected* and the scene *intro* in supernode *on-duty* will be performed. At the end of the scene, Cyberella asks whether the user wants to provide parameters for the ensuing demo. To handle this simple yes/no question (blocking event) we use a conditional yes-edge *c(yes)* and a no-edge *c(no)*. A third conditional edge *c(t>20), feedback-reminder* is triggered if the user does not answer within a certain amount of time (20 seconds). In this case the scene *feedback_reminder* is performed. During the presentation the user can give positive or negative feedback. This is realized using an interrupt edge *c(feedback), feedback_scene* which handles the feedback event. This event interrupts the generated sales dialogue and invokes the associated *feedback_scene*. Afterwards, the sales dialogue will be resumed. If the visitor leaves the CrossTalk installation, the interrupt edge *visitor_gone* immediately stops all ongoing activities in on-duty mode and activates the off-duty supernode.

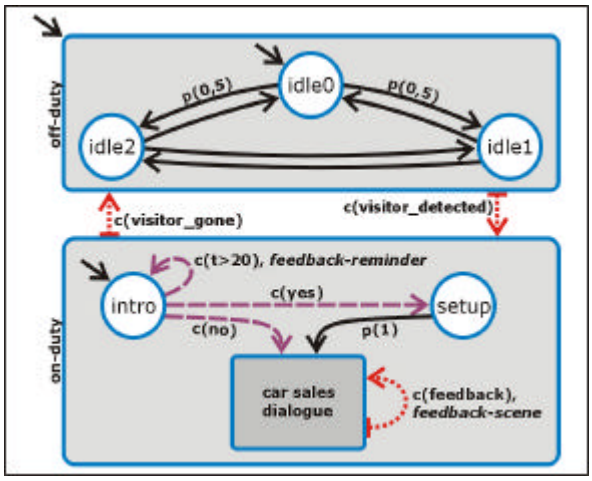

Fig. 3: Simplified scene flow in the CrossTalk scenario.

3 Exploiting CrossTalk's Discourse History

Our aim to "attract and bind" the user is based on the two operational modes of the system: on-duty and off-duty. In off-duty mode the user sees three actors supposedly engaged in small talk and rehearsals, an unusual activity meant to "attract" a potential user. Having succeeded thus, the system enters on-duty mode and the user sees a performance as the result of the observed rehearsals. Now a systematic stepping out of character refers back to the agents' "real life" as actors that had been explicitly established by the off-duty mode. To make the conversations among the characters more interesting and believable we anticipate that the characters should make more reference to: (a) feedback of the current user, and (b) previous encounters with other users.

At a number of occasions CrossTalk has been exposed to several hundreds of people most of which also interacted with Cyberella. A first step to account for (a) and (b) is to collect information in a discourse history and to extract from these data key measures like the user's interaction frequency, the user's average response time to questions, types of interaction, quality of feedback (positive, negative, ask for help) as well as the variance in feedback within a session and across users. Preliminary log file analysis showed that it is actually possible to distinguish between several stereotypical user types based on observed interaction patterns. Possible user categories include: critical user (many negative feedbacks), active/passive/slow user (many/few interactions, frequent interaction time-outs), collaborative/obstinate user. Also, it is possible to co-relate feedback with particular characters and their actions. For instance, if a user provides more often negative feedback on contributions made by the character Tina but applaud on Ritchie, it can be guessed that this user likes Ritchie more than Tina.

In turn, stereotypes as well as unusual interaction patterns can inform the selection of pre-scripted scenes as well as the automated generation of scenes. Moreover, we decided to provide script authors in the editing phase with a number of macros that refer to history contents. At runtime, these macros expand to access functions that read out referenced variables of the discourse history. In the following, we will provide some examples how the discourse history can be used to enrich the conversations.

Including discourse context in pre-scripted on-duty scenes

To enable a coherent embedding of pre-scripted dialogue fragments into an overall performance during runtime it is often useful to make reference to actual entries in the discourse history. For instance, an author may wish to trigger a certain pre-authored script in case a user continues to provide negative feedback on the performance for a longer time. Or an author may use access references in a scene specification as a placeholders for a counter, for instance, to have the characters talk about the actual number of interactions at runtime. Further examples are (U means user, X is either Tina or Ritchie, C is Cyberella):

- U seems to like X ? X says: "Oh, a real fan!"

- U seems to X ? X makes nervous gestures or does not react at all
- U seemed to disliked X but now provides positive feedback
  ? X says: "That was about time!"
- U seemed to like X but now provides negative feedback
  ? X says: "Oops, just a little accident"
- Lazy user ? C says: "You don't like talking, do you?"
- High feedback variance ? C says: "Why don't you just let it roll for a while" or start ignoring feedback.
- Liking could trigger some scene extensions, e.g. Tina/Ritchie giving out profuse thanks, side remarks by Cyberella about good vibrations, or by Ritchie on the topic of bribery...
- The assumed user class could guide the amount of options given to the user. If the user is active you could provide him/her with more frequent choices. If not, not.

Including discourse context in pre-scripted off-duty scenes

A great deal of CrossTalks pre-scripted scenes concern the character's off-duty behavior. Inspired by back-stage conversations among actors as well as by off-duty chats among stand personnel during exhibitions, these scripts encode small-talk conversations about everyday belongings, genre-specific jokes etc. Since it is quite natural to talk about personal experiences gained while performing professional activities, such as serving stand visitors, we are especially interested in having our characters reflect on their interaction experiences with real users. While the basis for dialogue contributions may be exactly the same entries taken from the discourse history, their verbalization will be more colloquial in the off-duty mode as compared to the on-duty mode. For instance, they may call the previous user a "lazy guy" and even make a joke about him (i.e., insert the definite reference expression into a certain joke template that may be applied to any lazy users).

Including context in generated scenes

If the generated content is semantically tagged (e.g., interesting vs. less interesting), one can guide the selection of plans (dialogue strategies) with the help of context. For instance, reacting to less interested users by playing more spectacular scenes or reacting to a long session duration by playing shorter scenes. The way you can use context in generated scenes is highly application-specific. If the agents are equipped with models of personality and emotions, the feedback can be processed by adapting the agents' mental state which in turn effects their behavior [1].

3 Conclusions

Designing applications with animated characters can become a complex task, especially if multiple characters are involved and user interaction has to be accommodated. The presented CrossTalk system is an example of such a relatively complex

system. A particular challenge in CrossTalk concerns the scripting of scenes. We have sketched an approach that combines pre-authored scripting with automated script generation. We also sketched how a discourse history can be exploited to enrich the conversational skills of the involved characters. Related scripting approaches include the systems IMPROV [14], SCREAM [14], and the mission rehearsal application by [17].

Originally CrossTalk was developed for the CeBIT'02 convention with the objective to attract CeBit visitors and motivate them to enter the DFKI booth. Since the installation has been demonstrated to a public audience at various occasions (including COSIGN'02 in Augsburg, IST'02 in Copenhagen, and in-house events at DFKI). We conclude that the installation suffices its purpose (i.e., attract stand visitors and amuse them). However, there are many opportunities to extend and improve CrossTalk further. In our future work on CrossTalk we plan to further exploit the discourse history. One idea is to create statistical models of behavior from the session log files, that can be used to predict future behavior. This extension will enable our characters to speculate about the a visitors next choices suggest ("Let me guess which issues would interest you most for the next demo"), and in the off-duty mode also about the kind of next visitors. Another objective is to arrive at characters that develop their own back-history from one exhibition to another.

Acknowledgement

The work presented in this paper is a joint effort with contributions from our colleagues Stephan Baldes and Markus Schmitt. We would also like to thank our graphics designer Peter Rist for providing us with the virtual actors Cyberella, Tina, and Ritchie. CrossTalk has been built upon contributions from the MIAU project funded by the German Ministry for Education and Research and from the EU-funded IST projects NECA, SAFIRA, and MAGICSTER.

References

1. André, E., Klesen, M., Gebhard, P., Allen, S., and Rist, T. Exploiting models of personality and emotions to control the behavior of animated interactive agents. In: Proc. of the Agents'00 Workshop on Achieving Human-Like Behavior in Interactive Animated Agents, 2000, 3-7.
2. André, E., and Rist, T. Presenting through performing: On the use of multiple animated characters in knowledge-based presentation systems. In: Proc. of IUI'00, ACM Press, 1-8.
3. André, E., Rist, T., van Mulken, S., Klesen, M., and Baldes, S. The automated design of believable dialogues for animated presentation teams. In: Cassell, J., Sullivan, J., Prevost, S., Churchill, E. (eds.). Embodied Conversational Agents, Cambridge, MA, The MIT Press, 2000, 220-255.

4. André, E., Rist, T., and Müller, M. WebPersona: A life-like presentation agent for the world wide web. In: Proc. of the IJCAI'97 workshop on Animated Interface Agents: Making them Intelligent, Nagoya, 1997.
5. Baldes, S., Gebhard, P., Kipp, M., Klesen, M., Rist, P., Rist, T., and Schmitt, M. The interactive CrossTalk installation: Meta-theater with animated presentation agents. In: Proc. of the PRICAI'02 workshop on Lifelike Animated Agents, 2002.
6. Cassell, J., Sullivan, J, Prevost, S., and Churchill, E. (eds.). Embodied conversational agents. The MIT Press, Cambridge MA, 2000.
7. Gebhard, P. Enhancing embodied intelligent agents with affective user modelling. In: Proc. of UM'01 (doctoral consortium summary). Springer, Berlin, 2001.
8. Hayes-Roth, B., van Gent, R.: Story-Making with Improvisational Puppets. In: Proc. of Autonomous Agents'97, 1997. 92-112.
9. Johnson, W.L., Rickel, J.W. and Lester J.C.: Animated Pedagogical Agents: Face-to-Face Interaction in Interactive Learning Environments. International Journal of Artificial Intelligence in Education 11:47-78, 2000.
10. Laurel, B. Computers as theatre. Addison-Wesley, Reading MA, 1993.
11. Murray, J.H. Hamlet on the holodeck: The future of narrative in cyberspace. The MIT Press, Cambridge MA, 2000.
12. Perlin, K., and Goldberg, A. Improv: A system for scripting interactive actors in virtual worlds. Computer Graphics, 29 (3), 1996.
13. Prendinger, H. and Ishizuka, M. Social role awareness in animated agents. In: Proc. of the Fifth Conference on Autonomous Agents, ACM Press, 2001, 270–377.
14. Prendinger, H. and Ishizuka, M. SCREAM: Scripting emotion-based agent minds. In: Proc. of AAMAS'02, ACM Press, 2002, 350-351.
15. Rist, T., Baldes, S., Gebhard, P., Kipp, M., Klesen, M., Rist, P., and Schmitt, M. CrossTalk: An interactive installation with animated presentation agents. In: Proc. of COSIGN'02, 2002.
16. Rist, T., and Schmitt, M. Applying Socio-Psychological Concepts of Cognitive Consistency to Negotiation Dialog Scenarios with Embodied Conversational Characters. Proc. of AISB'02 Symposium on Animated Expressive Characters for Social Interactions. 79-84. 2002. (Extended version submitted for publication in 2003.)
17. Swartout, W., Hill, R., Gratch, J., Johnson, W.L., Kyriakakis, C., LaBore, C., Lindheim, R., Marsella, S., Miraglia, D., Moore, B., Morie J., Rickel, J., Thiébaux, M., Tuch, L., Whitney, R., and Douglas, J.: Towards the Holodeck: Integrating Graphics, Sound, Character and Story. Proc. of Autonous Agents'01, 409-416, 2001.
18. Ryokai, K., Vaucelle, C., Cassell, J. Virtual peers as partners in storytelling and literacy learning. In: Journal of Computer Assisted Learning (in press).