USING AN ANIMATED TALKING CHARACTER IN A WEB-BASED CITY GUIDE DEMONSTRATOR

Georg Fries, Stefan Feldes, Alfred Corbet

Deutsche Telekom Berkom GmbH
64307 Darmstadt
email: {fries, feldes, corbet}@berkom.de

ABSTRACT

Taking into account that the user acceptance of an animated agent is influenced by different criteria, including appropriateness of the application domain, quality of application design, and quality of character design, we have developed a demo application in a domain where we found the agent substantially helpful – a web-based city guide. The animated character interactively guides the user through some sights of the city of Darmstadt. It can display and explain how to get to different places of interest by moving around and pointing at locations on a map. The system allows input via mouse clicks, speech and typed text. Output modalities of the agent are speech, gesture, text (cartoon word balloons) and some facial expression. Further we have designed two new 3D characters. We describe experiences gained during system development and discuss design aspects concerning application as well as character animation.

1. INTRODUCTION

One arising research topic in the field of spoken dialogue systems is the enhancement of speech interfaces by adding further modalities, e.g., the integration of animated talking characters in web-based applications [1]. Such a multimodal interface can potentially be more user-friendly and user-centered compared to common graphical user interfaces [2]. By using gaze, turn-taking and pointing gestures, the character can provide a link between the visual information and the accompanying speech as well as emulate some aspects of a more human-like communicative behavior [3]-[6].

The acceptance of applications with animated agents depends on many factors, including the quality of the underlying speech technology, such as the intelligibility and naturalness of the speech synthesis and the accuracy and features of the speech recognition. In this paper, however, we focus on further important criteria which may influence the user acceptance: Appropriateness of the application domain, quality of the application design, and quality of the character design and animation. We start with describing the city guide application we have chosen from the domain of tourist information, and why we found the animated talking agent substantially helpful. We discuss the advantages and limitations of the underlying animation software as well as some general design aspects for web-based applications using animated agents. Finally, we present some suggestions how to design the character resulting from our experience during several design cycles for our new three-dimensional characters, Harry and Joe.

2. INTERACTIVE CITY GUIDE DEMONSTRATOR

2.1. Application Description

Our Darmstadt City Guide is a demo application where an animated talking character offers orientation help to tourists visiting the city of Darmstadt. It is realized with web-based technology and thus, in principle, accessible via the Web. The service may also be installed as a kiosk at places of interest for tourists, e.g. train station, visitor center. The language of the service (text and speech) is English, except for particular situations in the dialogue where a side character speaking German enters the scene.

The animated agent, Harry (Figure 1), is as an interactive city guide who presents information on the city highlights and shows users how to get there by foot or public transport. He displays pictures and maps for illustration, uses pointing gestures to emphasize the current focus and uses speech synthesis to explain aspects that are not visually present. For instance, on the museum page, the animated agent uses the picture of an outside view of the museum to present which exhibitions the visitor may find in which part of the building and where exactly the entrance may be found. For this purpose, he moves around the page pointing at the different parts of the building while orally explaining what is currently on exhibit.

Three main categories of information are presented:

- a guided city tour which acts as an overview
- a section with pages on individual city sights such as Landesmuseum, Future Lab, Wedding Tower
- a section where the user can ask for route description between different places

The guided city tour is mainly based on a detailed map of the center of Darmstadt, whereas the individual city highlights are presented with overview and detailed pictures of the sight. In both cases, the character moves over the map or pictures to the points of interest presenting them with pointing gestures and oral explanations.

Figure 1: The user may select his/her choice either by speaking or by clicking the option on a pop-up menu. Once the choice is made, the agent starts "walking" the required route on the map

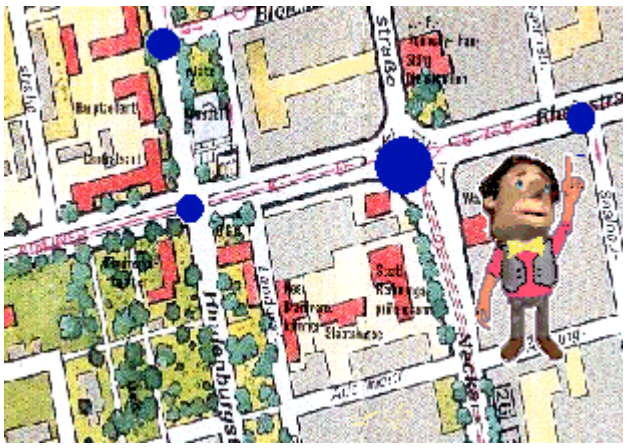

Figure 2: While "walking", the agent orally explains when important points for orientation are reached and marks all direction changes at crossings with a circle

The street guide section, where the animated agent explains and displays how to get from one place to another, exhibits a higher degree of interaction with the user. Harry asks for different parameters, like desired start and destination locations, and whether public transport, walking or driving routes should be shown. He also offers options such as a preview of the destination on the map. The user may select his/her choice either by speaking or by clicking the option on a pop-up menu (Figure 1). Once the user's choice is made, the guide starts "walking" the required route on the map. He orally explains when important points for orientation are reached, e.g.: *"When you've reached the gas station turn right ..."* and marks all direction changes at crossings with a circle (Figure 2). He may also suggest interesting sights on the way. This functionality is realized not in a hard coded manner, but is calculated online based on a predefined network of crossing points. The algorithm used finds the next crossing point in the direction of the target relative to the current point. Thus, the process of adding new routes is significantly facilitated.

2.2. Underlying Technology

In the following section, we describe and comment on the technology we used for this application.

We implemented our application as a series of Web pages. The animated characters were realized on the basis of the Microsoft Agent technology [7]. Programming the interactive behavior of the characters was embedded in the HTML pages in the form of Visual Basic scripts. JavaScript could be used in principal. However, it was missing support for events at the time of our development. The Microsoft Agent itself is accessed via an ActiveX control. This approach implies some obvious limitations. ActiveX, as well as the requirement that the MS Agent server must be installed on the system displaying the Web page, in effect limits the usage to MS operating systems.

Due to the standardized speech interface of the MS Agent technology it was possible to integrate German speech synthesis and German speech recognition with our application without having to make changes to the animation part of the software. For this we used the support for dynamic activation of vocabularies which depend on the current dialogue situation. The level of unnaturalness inherent to today's speech synthesizers seems to be less critical, because the voice appears to belong to the cartoon-like character in this environment.

The animation software allows the integration of new characters into the system. We designed and developed two new 3D full-bodied cartoon-like characters each having a repertoire of 50 gestures. Realization is done in 128x128 pixel graphics. Thus, computational load for graphical animation is kept low and can be run on any up-to-date PC.

A major advantage of MS Agent technology is its availability and its usability on today's standard PCs due to fast efficient graphics and easy script-based programming. However, with respect to the inherent properties of the current version of animation software we found the following limitations during application development:

- A character cannot speak and gesture at the same time. This is a major handicap as it slows down the animation and is perceived as unnatural by the user.
- Different characters may be present on the screen. However, only one can be active, i.e. can speak, move or gesture at a time.
- In contrast to virtual environments like VRML worlds, the agent is not part of an object scenery, but is placed and moved on top of the 2D plane of the screen. Positioning can be programmed in terms of screen coordinates relative to the upper left corner of the browser window. However, a desirable script on a symbolic level, like "Go to top picture on this page" is not possible.

3. APPROPRIATE APPLICATION DOMAIN

There is an ongoing discussion about the effects on the user that the integration of animated characters into interfaces may have, e.g., if the character improves the usability of the application or if the user may be distracted by it's actions. In an evaluation study of a presentation agent reported in [8], there was no support for the assumption that using a life-like character could improve user comprehension of information presented. On the other hand, that study reports that most subjects rated their learning tasks presented by a character as less difficult than those without an agent, i.e., using an agent was perceived as an improvement. We have not yet applied any formal evaluation to our system. However, our preliminary observations during system design and dialogue programming seem to confirm the latter result of the study in [8]: Most users subjectively perceived our agent as helpful and attractive.

We believe that employing animated characters can enrich an application if one or more of the following services that an agent can provide are required by or useful for the respective application domain:

- explaining details of graphical objects or pictures by combining speech with appropriate pointing gestures and related movements on the screen
- giving the information on the screen a temporal and spatial structure by communicating it sequentially
- directing the users attention to different parts or objects of the visual information
- simulating human behavior so that the user feels more familiar, e.g., allow for the emulation of conversation styles common in human-human communication [8], [9]
- including an emotional component

This will then result in benefits for the application such as

- improving expressiveness of the application,
- reducing the users barrier to get acquainted with and use the system [10],
- attracting a special user target group, e.g. children or novice users,
- making the information presentation more lively or just adding more fun.

In this sense, the street guide section of our city guide, where Harry shows how to get from one place to another, is a tailor-made task for an animated agent. Here, the agent acts like and emulates the behavior of a human expert, such that he communicates the information not as a whole, but sequentially in handy pieces.

Other appropriate domains include scenarios which simulate a real life social situation, e.g., the character plays the role of a sales agent, a teacher, a travel guide, a product presenter or a personal assistant.

4. APPLICATION DESIGN ASPECTS

Unfortunately, the design criteria of spoken dialogue systems cannot directly be applied to speech interfaces with animated agents. The integration of animated talking characters reveals new problems, such as deciding which modality to use for agent input and output. In our system, the user can decide when to use speech or graphical menus as input modality. Further, we let the agent speak accompanying explanations, but use text display for longer passages of informational text. In the following sections we discuss two other important aspects of application design: timing and employing multiple characters.

4.1. Appropriate Timing

Appropriate timing is important for all interaction levels, especially the duration of utterances, gestures, movements, and complete scenes. One way to be efficient is to allow the agent to talk in an abbreviated conversational style using pronouns and contractions. However, we found it difficult to judge the appropriate length of the agent's utterances. For instance, some users rated an utterance as natural and efficient, while others perceived the same utterance as rigid or unnaturally brief. In another case, the agent was perceived as polite by some, and ponderous and long-winded by others.

Because of the limitation of the underlying technology our character cannot speak and gesture at the same time. For this reason we had to keep the gestures and movements short to avoid longer unnatural breaks between his utterances.

Apparently, users have significantly differing preferences on the behavior of animated characters. One solution is to let the user customize the agent's performance individually, allowing him/her to chose long or short explanations, politeness or efficient commands, extensive gesturing and moving or minimal body action. Finally, the user should be able to switch the agent off without loosing application control.

4.2. Using Multiple Characters

The simultaneous or alternative use of multiple characters is a good means towards realizing different roles. In our application, Harry, the main character, meets one of two other agents who have different characteristics: a professor, and a young guy named Joe. The professor serves as an expert with deeper inside knowledge, and explains background details of the sights. Joe is capable of speaking German. He appears when the user prefers to get explanations in German. Another possible scenario is a discussion between an information seeker and an expert, or a roundtable discussion on a special topic, where each character represents a different point of view. When designing such multiple character scenes, it is important to retain interactivity with the user, i.e., to ask him/her questions and to avoid the generation of long video-like scenes.

5. CHARACTER DESIGN

In this section we discuss character design aspects which increase the agent's ability to communicate with the user.

5.1. Face And Body Size

Because the animated character is a means towards communicating information but not an integral part of the visual information which is presented to the user, the default size of the agent is kept relatively small. To emulate a kind of three-dimensionality in the on-screen scenery, we can change the agent's size to let him appear far or near in relation to other visible objects. To ensure that the user can clearly recognize the character's actions, we have carefully designed the parts of the body which are important for the communication task: head, mouth, lips, eyes, eyebrows and hands. Since we have chosen to design cartoon-like characters, it was not necessary to retain the natural proportions of human physique, e.g., we gave Harry an oversized face compared to his body.

5.2. Technical Design And Animation Aspects

Each animated gesture or movement is composed of a timed sequence of frames consisting of bitmap graphics. So, in principle, it would be sufficient to draw each required picture for an animation sequence separately. Nevertheless, we preferred a more sophisticated two step design process. The first step was the development of a three-dimensional polygon model of the character. The second step was the extraction of bitmap sequences after deforming the 3D model in the desired manner. Using a 3D model in this way speeds up the generation of new gestures. Further, it is easier to realize clear, unambiguous pointing gestures and turning movements as compared to a 2D model.

The user should easily and clearly locate the agent on the screen to follow and understand the gestures, actions and movements. For that purpose it is important to ensure that the colors and the shape of the agent's body are in rich contrast to the pictures, maps and other graphical objects presented on the screen. Because the character may appear on top of these visual objects, it is better to avoid anti-aliasing the outside edge of the characters body.

6. CONCLUSION

We developed an application in a domain where we found the use of an animated character substantially helpful. Our next step will be a formal evaluation of the system, to confirm our preliminary results which indicate that users perceive the character as helpful and attractive. Further we intend to investigate if the agent can improve the application task in an objective manner, e.g., if the agent can speed up the information transfer to the user and if he can help the user to better understand and remember the presented information compared to a similar application without agents.

Finally, to improve our system, we are currently developing a platform independent VRML character. This approach will allow the agent to become part of an object scenery, rather than being placed and moved on top of the 2D plane of the screen. Further, it will facilitate the simultaneous generation of multiple gestures and speech.

7. REFERENCES

1. E. André, J. Müller, T. Rist. WebPersona: A life-like Presentation Agent for the World-Wide Web. *Proceedings of the IJCAI-97 Workshop on Animated Interface Agents: Making them Intelligent*, Nagoya, 1997.
2. P. Wyard, et. al.. A Combined Speech And Visual Interface To The BT Business Catalogue. *ESCA Workshop on Spoken Dialogue Systems*, Vigsoe, Denmark, 1995.
3. J. Cassell, M. Steedman, N. Badler, C. Pelachaud, M. Stone, B. Douville, S. Prevost, B. Achorn. Modelling the interaction between Speech and Gesture. *Proceedings of the 16th Annual Conference of the Cognitive Science Society*, Georgia Institute of Technology, Atlanta, 1994.
4. J. Beskow, K. Elenius, S. McGlashan. Olga - A dialogue system with an animated talking agent. *Proceedings Eurospeech '97*, Rhodes, Greece, 1997.
5. G. Ball, D. Ling, D. Kurlander, J. Miller, D. Pugh, T. Skelly, A. Stankosky, D. Thiel, M. Dantzich, T. Wax. Lifelike computer characters: The persona project at Microsoft. In J.M. Bradshaw, Editor, *Software Agents.* Cambridge, MA: MIT Press, 1996.
6. K. R. Thórisson. Gandalf: An embodied Humanoid Capable of Real-Time Multimodal Dialogue with People", *Proceedings of First ACM International Conference on Autonomous Agents*, Marina del Rey, CA, 1997.
7. Developing for Microsoft Agent, Microsoft Press Redmond, Washington, 1998. See also http://www.microsoft.com/workshop/imedia/agent/default.asp.
8. E. André, T. Rist, J. Müller. Integrating Reactive and Script Behaviors in a Life-Like Presentation Agent. *Proceedings of the 2nd International Conference on Autonomous Agents* (Agents '98), Minneapolis/St. Paul 1998.
9. J. Lester, J.L. Voerman, S.G. Towns, C.B. Callaway. COSMO: A life-like animated pedagogical agent with deictic believability. *Proceedings of the IJCAI-97 Workshop on Animated Interface Agents: Making them Intelligent*, Nagoya,1997.
10. B. Adelson. Evocative Agents and Multi-Media Interface Design. *Proceedings of the ACM UIST 92* Monterey, CA, 1992.
11. G. Fries, A. Wirth. Felix - A TTS System with Improved Pre-Processing and Source Signal Generation. *Proceedings Eurospeech '97*, Rhodes, Greece, 1997.