In Proceedings of International Conference on Intelligent User Interfaces IUI2004. Madeira, January 2004.

The Museum Visit: Generating Seamless Personalized Presentations on Multiple Devices

C. Rocchi, O. Stock, M. Zancanaro
ITC-irst
Via Sommarive, 18
I-38050 Povo Trento Italy
+39 0461 314 444
{rocchi,stock,zancana}@itc.it

M. Kruppa and A. Krüger
DFKI GmbH
Stuhlsatzenhausweg
D-66123 Saarbrücken Germany
+49 681 302 4137
{mkruppa,krueger}@cs.uni-sb.de

ABSTRACT
The issue of the seamless interleaving of interaction with a mobile device and stationary devices is addressed, in a typical situation of educational entertainment: the visit to a museum. Some of the salient elements of the described work are the emphasis on multimodality in the dynamic presentation and coherence throughout the visit.

The adopted metaphor is of a kind of contextualized TV-like presentation, useful for engaging (young) visitors. On the mobile device, personal video clips are dynamically generated from personalized verbal presentations; on larger stationary screens distributed throughout the museum, further background material and additional information is provided. A virtual presenter follows the visitors in their experience and gives advice on both types of devices and on the museum itself.

ACM Classification Keywords
H.5.1 Multimedia Information Systems, H.5.2 User Interfaces

Keywords
User interfaces, multimedia, life-like characters, adaptive systems.

1. INTRODUCTION
In this paper we mainly address the issue of the seamless interleaving of interaction with a mobile device and stationary devices, in a typical situation of educational entertainment: the visit to a museum. We present some novel developments in the PEACH project (http://peach.itc.it, [6]), dedicated to the exploitation of cultural heritage. The project's goal is to go one step further in the development of location-aware adaptive systems similar to the multimodal approaches presented in [1] and [8].

The mobile system is intended to combine the dynamically adapted language-based output with a dynamically produced visual documentary. While the first part is an improvement of well established techniques, the second is based on cinematic techniques. The input to the system comes from the locations of the visitors and observations the system itself has made about their behaviour and presumed interests and what they have been exposed to and presented so far. All the material is presented coherently throughout the visit.

We describe work aimed at a kind of contextualized TV-like presentation, so that a young visitor is hooked. On the mobile device, personal video clips are dynamically generated from personalized verbal presentations; on larger stationary screens distributed throughout the museum (so-called *VirtualWindows*), further background material and additional information is provided. A virtual presenter follows the visitors in their experience and provides advice on both types of devices and on the museum itself.

2. THE SITUATIVE CONTEXT OF THE VISITOR
The situative context of the system is derived from different sensors, which are connected to the mobile device. The position of the mobile device is determined by the use of long-life infrared beacons that are installed throughout the museum. Several beacons with different sending ranges, which are installed in the same location, allow to roughly distinguishing the distance of the mobile device to that location.

Accelerometers provide the 3D-orientation of the device. This allows (a) to estimate the orientation of the visitor, and (b) to determine whether the user is looking at the screen of the device (the device is held within a certain range of vertical angle range).

All the visitor interactions with the mobile device are recorded and send to a central server, where the visitor's situative context is constantly updated.

3. TV-LIKE PRESENTATIONS ON THE MOBILE DEVICE
Although many research projects are exploring the new possibilities offered by Personal Digital Assistants (PDA) in a museum setting (see for example, [2] and [3]), usually these multimedia guides use static images, while others employ pre-recorded short video clips about museum exhibits. In our approach, we have focused on automatically produced video clips to be played on the small screen of the mobile device and using a life-like character either as an anchorman or a presenter.

3.1 Personalized Video Clips

In the Peach project, we use information about discourse structure for automatically producing video clips [8]. At presentation time, a sequence of pictures is synchronized with the audio commentary and the transitions among them are planned according to cinematic techniques. The language of cinematography [5], including shot segmentation, camera movements and transition effects, is employed in order to plan the animation and to synchronize the visual and the verbal parts of the presentation. In building the animations, a set of strategies similar to those used in documentaries were thus employed. Two broad classes of strategies have been identified. The first class encompasses constraints, imposed by the grammar of cinematography, while the second deals with conventions normally used in guiding camera movements in the production of documentaries. For instance, a strategy in the first class would discourage a zoom-in immediately followed by a zoom-out, while a different strategy in the second class would recommend the use of sequential scene cuts, rather than a fade-out effect, to visually enumerate different characters in a scene. In order to formally use discourse structure, we employ the Rhetorical Structure Theory [4].

The input for the Video Clips planner is a text annotated at discourse level, made of non-overlapping spans (*segments*), where each segment has a *topic* (the entity the text is about) and the *rhetorical relation* that links it to the previous text span. Besides the annotated text the planner takes in input a repository of images. The annotation schema provides general features of each image (height, width and source file) as well as information about details, relevant portions of an image illustrating one or more topics.

3.2 The role of life-like characters during presentations

While the dynamically arranged video clips are a basic element of our dynamic presentation, we have also experimented with a life-like character that plays the role of an accompanying agent, ready to move on the mobile device or to jump on the Virtual Windows, in order to provide continuous assistance and continuity to the presentation. The character helps in solving problems like how to reach a certain exhibit, and yielding explanations. User evaluations [7] have shown that the introduction of a life-like character makes presentations more enjoyable and attractive (something that we regard as very important to keep younger visitors engaged).

The use of life-like characters on portable devices has to be carefully weighted because of the small dimension of the display. Nevertheless, there are specific roles that a properly designed character can play on a mobile device to improve the level of engagement with the presentation. In particular, following the TV metaphor, two main roles can be recognized: the *presenter* and the *anchorman*. When playing the role of the presenter, the character introduces new media assets and uses pointing gestures. When playing the role of the anchorman, the character just introduces complex presentations without interfering with them any further. Although simpler than the presenter, the role of an anchorman can help the visitor understand many different presentations, providing a context in which they all make sense. In its role of an anchorman the character also supports the seamless integration of the mobile devices' small screen and the large screen of the Virtual Window. Similar to a TV-presenter who walks around the studio to present different content, the character is able to move between the mobile device and the Virtual. Besides the specific role that the character may play, it is also a metaphor for the actual interests of the visitor. By providing different characters and giving the visitor the choice between them, the different views on the exhibits are transparently conveyed and selected.

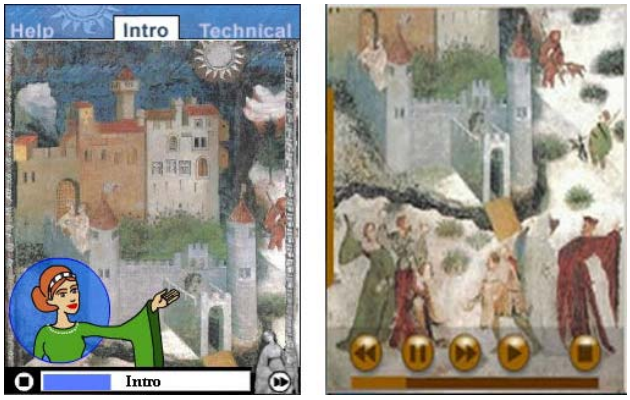

Figure 1. Screen shots from a running presentation: the life-like character first presents a static graphics and then announces the presentation of a video clip.

For example, in our demo scenario we chose an aristocratic woman (see Figure 1) for the generally interested visitor and an artist for visitors who are more interested in explanations on artistic techniques.

The mobile device receives the multimedia presentation from a presentation server over a wireless network. During that presentation, the visitors have several means to control the flow of the presented material, i.e. they are able to interrupt or to jump to other parts of the presentation.

4. PRESENTATIONS ON THE VIRTUAL WINDOWS AND TRANSITIONS BETWEEN DEVICES

The Virtual Window is the primary medium to provide the visitors with in-depth information on interesting topics. It has enough resolution to allow the full use of graphics, animations and video clips of all kinds.

If visitors approach a Virtual Window, their personal presentation agent will transit to the Virtual Window, where it appears fully sized. When visitors approach a Virtual Window for the first time, the presentation agent, in its role of an anchorman, proactively informs them about the Window and how to make use of it.

If the visitors are close enough, the presentation agent starts to disappear from the mobile device and to reappear on the Virtual Window. This *beam*-effect is used to guide the visitor's attention towards the Virtual Window, where they find the personal presentation agent continuing the presentation. Once the presentation agent is on the *VirtualWindow*, the visitors can continue to coherently interact both with the agent and the presentation. The presentation agent is playing a more active role while guiding the visitor through the presentation on a *VirtualWindow*. Sophisticated gestures and animations thus lead to a much more believable appearance.

Another functionality that we make use of is the possibility for the visitors to choose a different presentation agent before leaving the Virtual Window. Since each character represents a special interest group (e.g. in our scenario a neutral character and an artist), the newly chosen character changes the stereotype that is used to classify the visitors and hence influences the future presentations generated by the server. Finally, when leaving the *VirtualWindow* the presentation agent follows the visitors and after another transition automatically reappears on the mobile device.

5. ADAPTING PRESENTATIONS TO THE SITUATIVE CONTEXT OF THE VISITOR

One goal we had in mind when designing the concepts for our project, was to transparently combine mobile and stationary output devices. In addition, it is necessary to adapt the style and content of the presentation to the situative context of the users in order to provide a coherent presentation throughout the visit, leaving the underlying technology as unnoticed as possible and thus emphasizing the contents of the presentation.

Given a visitor specific situative context, the presentation server first selects the appropriate content and the degree of adaptation that is necessary. For this purpose, we make use of different strategies that adapt the presentation not only to the location and the interest of the visitor but also to the available modalities. Since we are able to distinguish between users who are actually only listening to the presentation and visitors who are looking at the devices, the system can decide when to provide video clips and when audio-only. The strategies also take into account technical resources of the output media, i.e. the screen resolution and display size.

The content for presentations at the Virtual Window is selected according to the visitor's interests during the visit. Instead of providing only additional material according to the stereotype (e.g. general vs. artistic view) the system provides further detailed information on the exhibits that were of specific interest to the visitor (according to the *visiting history*). Meta-strategies allow providing the visitors with information that helps to change their situative context if necessary. The system could for example advise the visitors to look at an image that is displayed on the mobile device. One specific strategy even allows the system to guide the visitors to the next Virtual Window, where the content may be presented more appropriately. After having determined the content and structure of the presentation, the server starts to plan the behavior and role of the life-like character and where appropriate also plans the structure of a video clip.

The behavior of the life-like character is captured in its own set of strategies, helping the system to decide for example, which of the two roles (presentation agent or anchorman) the character should play during a piece of presentation.

Finally the server renders the overall presentation with material retrieved from a multimedia database that contains graphics and text. At this point the video clips are generated from static graphics and the text for the character is transformed into spoken language using a speech synthesizer. The final presentations are then delivered either to the mobile devices (via wireless network) or to the Virtual Windows.

6. CONCLUSIONS

In this paper we have described an intelligent interface project aimed at making a museum visit of interest to the young. The overall modality is reminiscent of the concept of mobile, personalized TV, with a presenter, that gives continuity, video clips and verbal presentation. The main feature is that all this material is personalized and produced in a context-dependent manner. We believe this may lead to further developments in the area of educational entertainment. There are many themes that are completely open. Perhaps the biggest challenge is concerned with keeping the attention of the users high and granting a long term memory effect. We need to be able to design presentation techniques that hook the visitors and continuously build the necessary anticipation and release tension. The "story" must be entertaining, and, in the future, it should include mechanisms of surprise. The expectation sometimes must be contradicted and this contrast will help in keeping the attention and memorizing the situation. Another aspect is that the users are responsible for what they do and hence for the material that is presented to them, but yet through the presentation some specific goals of the museum curator can be submitted for adoption. The way is also open to new modalities of visit, particularly important with children.

7. ACKNOWLEDGMENTS

The work has been conducted in the context of the PEACH project funded the local government of the Autonomous Province of Trento.

8. REFERENCES

[1] J. Baus, A. Krüger, and W. Wahlster. A Resource-Adaptive Mobile Navigation System. In *IUI2002: International Conference on Intelligent User Interfaces*, pages 15– 22, New York, 2002. ACM Press.

[2] K. Cheverst, N. Davies, K. Mitchell, A. Friday, and C. Efstratiou. Developing a context-aware electronic tourist guide: Some issues and experiences. In *Proceedings of CHI 2000*, Amsterdam, 2000.

[3] R. Malaka and A. Zipf. Deep map – challenging it research in the framework of a tourist information system. In D. Buhalis (eds) D. R. Fesenmaier, S. Klein, editor, *Information and communication technologies in tourism*. Springer-Verlag, Wien, New York, 2000.

[4] W.C. Mann and S. Thompson. Rhetorical structure theory: A theory of text organization. In L. Polanyi (ed.), editor, *The Structure of Discourse*. Ablex Publishing Corporation, 1987.

[5] C. Metz. Film Language: a Semiotics of the Cinema. Oxford University Press, New York, 1974.

[6] O. Stock. Language-based interfaces and their application for cultural tourism. *AI Magazine*, 22(1):85–97, 2001.

[7] S. van Mulken, E. André, and J. Mueller. The persona effect: How substantial is it. In *Proc. Human Computer Interaction Conference*, pages 53–58. Springer, 1998.

[8] Zancanaro, Stock, and Alfaro. Using cinematic techniques in a multimedia museum guide. In *Proceedings of Museums and the Web 2003*, Charlotte, NC, March 2003.