The Environment as a Medium: Location-aware Generation for Cultural Visitors

Elena Not, Daniela Petrelli, Oliviero Stock, Carlo Strapparava and Massimo Zancanaro
ITC-irst, Istituto per la Ricerca Scientifica e Tecnologica,
I-38050 Povo TN, ITALY.
URL: http://www.itc.it/irst/
{ *not* | *petrelli* | *stock* | *strappa* | *zancana* }*@irst.itc.it*

Abstract

Multimedia generation for a location-aware system has to take into account the fact that moving in a physical space and seeking for information is a cognitive as well as a psycho-motor experience. The surrounding environment attracts, distracts, provides evidence, reinforces understanding, conveys information. It plays -we propose- the role of a medium, to be deliberately allocated and coordinated with the others. In this respect, multimodality plays a key role in emphasizing the prominenence of this medium and in getting round its staticity. Reference is made to two location-aware prototypes we implemented for individualised presentation of cultural sites.

1 Location-aware systems: when the environment enters into play

Recently, many works have investigated the potential of location-aware systems for augmenting the functionalities of the environment to support users in their everyday activities or leisure (HCIMD, 1998). By using sensors, these systems are able to locate the user's position and to react properly, e.g. by activating/deactivating electronic devices, transferring data for ubiquitous computing, or by proposing information helpful to the user's current task by means of portable devices. The field, in Europe especially, is considered of strategic importance (see for instance the EU 5th Framework Programme). We consider this as one element of an extended view of multimodal presentations, as discussed in (Stock, 1999).

In this paper we discuss how the surrounding environment influences the way in which information should be presented to a user for an effective *in situ* communication. Our proposal, discussed in section 3, is that a general way to treat the physical context is to consider it as one of the available media (even with its peculiarities and limitations), to be appropriately coordinated with the others. These ideas have emerged within the HyperAudio project (Not et al., 1998) and the HIPS European project[1] (Benelli et al., 1999), in a scenario where a flexible hypermedia is integrated in a portable electronic guide. In our discussion we will focus on the museum domain, but many of the issues discussed have general validity in the field of location-aware computing where objects and situations represent the main focus of the visitor's task (e.g., cultural tourism).

When designing a multimedia generation component for a location-aware system like a portable electronic guide to a museum, designers should bear in mind that the experience of a user moving in a physical space and seeking for information is twofold:

- a cognitive experience: visitors move around and get information about the objects exposed (reading/hearing descriptions contained in the accompanying labels, trying to relate what they are seeing to their own experience, ...). They need to accommodate and interpret the information source to their own pace and interests (Oberlander et al., 1998); here the interaction scenario is richer than for the exploration of a virtual museum.
- a psycho-motor experience: the architecture of the building, how the exhibition is organised, exhibits' appearance, crowding, distances and architectural barriers are all elements that can improve or hamper the visitor's enjoyment. Users' behaviour in the environment differ in many aspects, either in how they move in the space and in the type and amount of information they seek (see also (Marti et al., 1999) and (Petrelli et al., 1999)).

More in general, the surrounding environment attracts, distracts, provokes emotions, provides evidence, reinforces understanding, and, particularly in a cultural scenario, conveys information. Therefore, multimedia information generated and provided to the user in this richer interaction context is influenced by the following factors:

[1] HIPS is a joint effort of the following consortium: University of Siena (Italy, coordinator), CB&J (France), GMD (Germany), ITC-irst (Italy), SINTEF (Norway), University College Dublin (Ireland), and University of Edinburgh (Scotland).

1. where the user is and how she moves in the space (e.g., whether she is in front of an object or whether she is simply walking around a room);
2. how the surrounding physical space is structured (e.g., whether objects are close or not, whether the room is crowded) and it provides visual feedback that helps understanding and orientation;
3. how the available media/modalities (e.g., written text, audio messages, graphics projected on glasses) allow users to concentrate on the real objects and support the system in guessing the information assimilated by the user and her interests.

2 A sample application scenario

Within the HyperAudio and HIPS projects we have implemented a flexible hypermedia generation system where the environment plays a relevant role. The developed application scenario is that of a portable electronic guide to a museum (more in general, a cultural site) that provides contextual and personalised information about exhibit objects, by detecting physical movements and maintaining a model of user's interaction and interests. Information is mainly conveyed through audio messages dynamically composed and sent to the user's headphones. Additional links to further information and related images are available for pen pointing on a palmtop screen. See figure 1.

The following sample text is generated by the HIPS system for a visitor facing the fresco "La Maestà" by Simone Martini, placed in the Palazzo Pubblico in Siena (for more detail on the generation process see section 4).

At your left, you can admire the great fresco "La Maestà", depicted by Simone Martini in 1315. The fresco is located in the main part of the hall, the central point that gave the orientation of the Sala del Mappamondo. On the contrary the "Guidoriccio" fresco, on the opposite side of the room, was a sort of great 'poster', glorifying the power of the Siena Republic. It was a sort of historical documentation more than an artwork to be judged for its artistic value.

This text exemplifies some of the peculiarities of information presentations generated within a physical environment: (i) inclusion of references to what the visitor is seeing; (ii) support for orientation in the physical space; (iii) suggestions for further physical exploration.

In the following section, we discuss the general principles followed by our presentation Composer to build effective presentations contexualized in the physical environment. We propose to consider the environment as another medium (with its own characteristics and constraints) available to the Composer: taking the environment into account is thus an instance of the general problem of media selection (see for example,(Reiter, 1997)).

3 Integration of the physical environment with other media for a coherent presentation

According to (Bordegoni et al., 1997), a medium is a physical space in which perceptible entities are realized. Indeed, in a museum (as well as in a cultural city, an archaeological site, etc.) the most prominent medium is the environment itself. A multimedia generation component has take into consideration this special status of the environment so to build coherent and more effective presentations. The aim is that of *integrating* the 'physical' experience, *without competing* with the original exhibit items for visitor's attention.

- From a multimedia point of view, additional uses of the visual channel have to be carefully weighted. In HIPS, we prefer to exploit the audio channel (mainly for language-based presentations, even if the role of non-speech audio, like music or ambient sounds, has been investigated). Yet we use images on the PDA to support the visitor in the orientation task (in HyperAudio, we experimented oriented maps for physical space navigation and, in HIPS, 3D or 2D images to support linguistic reference to physical objects). In this latter case, the visual channel is shared between the PDA and the environment but the goal is still to provide support to environment-related tasks[2].
- From a multimodal point of view, other modalities have to be employed to focus the visitor's attention on specific objects or to stimulate interest in other exhibits. For example, in HIPS the linguistic part of the presentation (speech audio) makes large use of deictic and cross-modal expressions both with respect to space (such as "here", "in front of you", "on the other side of the wall", etc.) and to time ("as you have seen before", etc.). At a deeper level, presentations are planned to contain elaborations on visual details or comparisons to other exhibits[3].

The peculiarity of the environment as a medium is its staticity: the system cannot directly intervene on

[2] Actually, the PDA screen is also employed to suggest further navigation in the virtual space of information; an alternative realization for this would have been the use of auditory icons (Mynatt et al., 1998)

[3] The extent to which the system inserts elaborations and comparisons depends on the user model.

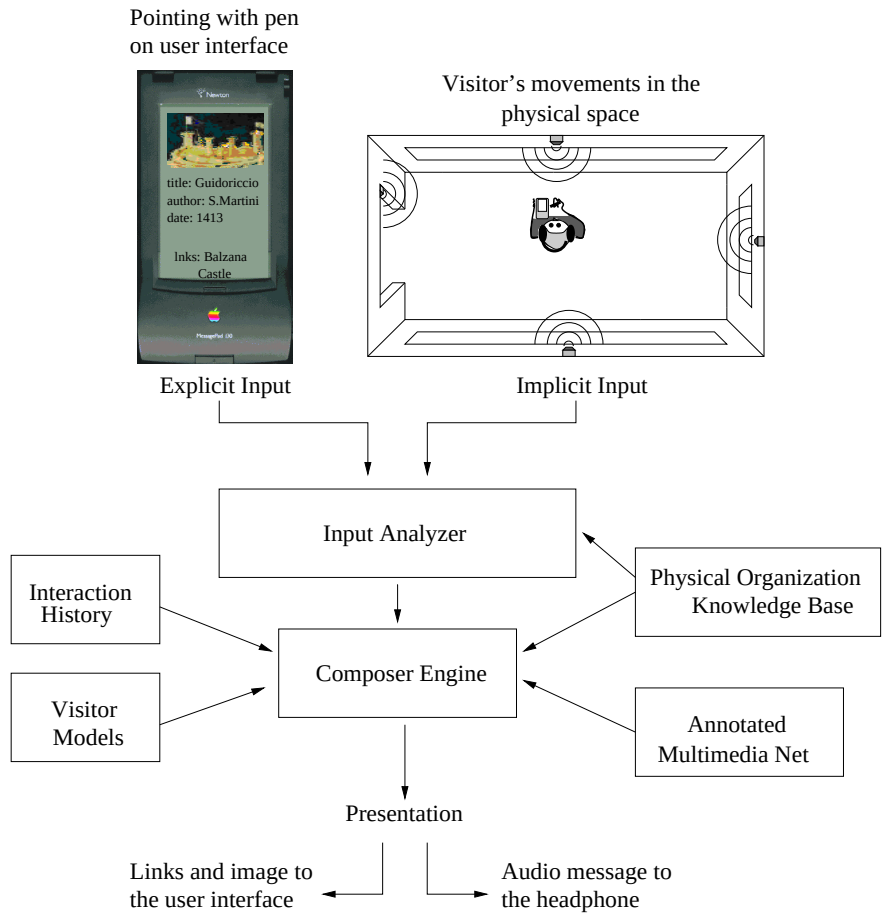

Figure 1: General architecture for the HyperAudio/HIPS Presentation Composer

this medium (i.e. the system cannot move or hide exhibits nor change the architecture of a room, at least without considering technology-based futuristic extensions.) Therefore, it may appear that the main task of a presentation system is to adapt other media in service of the environment. Yet a multimodal approach can get round the staticity constraint at least in two ways:

Dynamically changing the user's perception of the environment: exploiting augmented reality techniques (for example as described in (Feiner et al., 1997)) it is possible to overlay labels or other images on what the visitor is actually seeing. In this way, for example, the system can plan to highlight some relevant exhibits in the environment or shadow other less relevant ones. 3D audio effects or selection of characteristic voices or sounds for audio messages can stimulate user's curiosity and attention (Marti et al., 1999). A similar effect can be obtained exploiting the power of language: language-based presentations can be carefully planned to attract the visitor's attention to more important exhibits and shadow less relevant ones. The simplest example: when in HIPS a visitor enters in a room for the first time, she usually receives a general presentation of a room followed by the presentation of what the system hypothesises is the exhibit most interesting for her.

Changing the user's physical position: the system can induce the user to change her physical position either by a direct suggestion (e.g. "go to the other side of the room, the big fresco you'll see on wall is La Maestà") or indirectly by introducing a new topic (e.g., "La Maestà, one of the absolute masterpieces of European Gothic painting, is located in the wall behind you").

Ultimately, the goal of a location aware system for cultural visits is to support visitors to adapt their visiting experience to their own interests; but in some cases a visitor should be encouraged not to miss some particular exhibits (for example, you cannot visit the Louvre for the first time and miss the Gioconda). Sometimes this task can be accom-

plished by direction giving, but there are other ways to promote exhibits: for example, by providing at the beginning of the visit a list of hotspots, or by planning a presentation that, in a coherent way, links the exhibit in sight to other ones through reference to the visitor's interest. More in general, further research is needed in the direction of pedagogically-motivated systems possibly with meta-goals to pursue, educational strategies to follow and intentions to satisfy. In this respect, the interaction between the visitor and the system must evolve from simple interaction to full-fledged collaboration (for a discussion on this topic applied to cultural tourism see (Stock, 1999)).

4 The solution adopted in HyperAudio/HIPS

The approach adopted in HyperAudio/HIPS to generate presentations for a user moving in a physical space aims at a high reuse of existing information repositories. Presentations are built concatenating material selected from a repository of information units appropriately annotated, which is designed so that its content and its structure can be used in an adaptive way. For each unit, the annotation encompasses the description of the information contained, the relations with other units and the different ways in which the unit can be presented to the user (see fig. 2 for a sketch of a sample macronode). In our formalism, an atomic piece of data is called a *macronode* and, for textual information, it typically corresponds to a paragraph. Indeed, a macronode can represent data in different media, for example text, audio, images, or a combination of these. At the interaction time, the annotated data is processed by a Composer Engine, a rule-based system using discourse strategies and linguistic rules to introduce flexible content selection and organization and control over the linguistic realization (Not and Zancanaro, 1999). The rules make reference to external knowledge resources, like a model of the physical environment and user position, a user and visiting style model, a domain model, and work to guarantee that the composed message displays i) a coherent rhetorical structure; ii) a topic flow compatible with discourse and interaction context; iii) references to material already presented and to user interests; iv) cross-references to the environment; v) a linguistic realization emphasizing cohesion.

The user's movements are traced by means of infrared sensors, allowing to identify user's position and orientation. An internal model of the physical structure of the environment and of meaningful objects placement is kept in a Physical Organization Knowledge Base (POKB), which also contains information about the average time spent by visitors in front of each exhibit (Bianchi and Zancanaro, 1999). The information in the POKB is used by the system to reason on objects' proximity and to derive the user's interests (Specht and Opperman, 1999) and preferred visiting strategy (Marti et al., 1999).

Cross-references to the environment are introduced to emphasize the prominence of the surrounding space as a medium integrated in the current presentation:

- the actual exhibits are referred whenever possible to support system's descriptions with a visual feedback: *On the contrary the "Guidoriccio" fresco, on the opposite side of the room, was a sort of great 'poster', glorifying the power of the Siena Republic....*
- orientation is supported through language-based messages and images: *On the north-east wall, opposite to the windows, portraits of saints are located on each column. On the top of the wall there are two paintings representing famous battles in the history of the Sienese Republic.*
- suggestions on where to go next are provided according to the discourse context, the user's interest model, and the physical effort required to reach the new spot[4].

If the visitor walks away during a presentation, the audio message and the information displayed on the screen are properly adjusted to avoid inconsistency: for example, if the visitor moves away while an exhibit presentation is being played, the presentation is stopped; yet a general description of a room will not be stopped while the visitor is in the room.

The discourse rules that build presentations and coordinate the use of the different media and modalities treat the environment much in the same way as the other media: there is a representation of the information the environment conveys (partly in the POKB and partly in the domain concept hierarchy) and reasoning services use the visitor's current position and interaction history to identify the portion of the environment which is in focus for the current presentation[5]. A uniform formalism is adopted in the discourse rules to represent the features of the various media and of the discourse context to be tested.

From the implementation point of view, currently in HyperAudio/HIPS most of the decisions on media selection are realized by first selecting the environment spot relevant for the current presentation and then by coordinating the content of the other media accordingly. Indeed, an interesting research goal

[4] The physical effort might depend on distance, obstacles, user's handicaps, crowding, ...

[5] For example, for the same user's position a general description of the room or a specific description of the closest exhibit might be generated, according to which is the portion of the environment considered as the most salient.

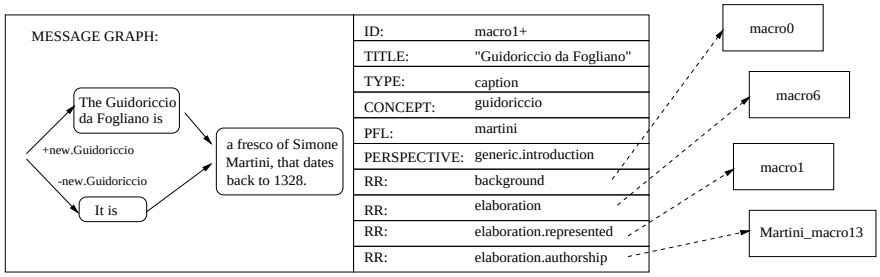

Figure 2: Sketch of the contents of a sample macronode for the artwork "Guidoriccio da Fogliano".

to pursue is that of getting a completely interleaved media selection and coordination.

5 Conclusion

In this paper, we have proposed that in designing a multimodal generation component for a context-aware system the environment is treated as a medium. We have talked about the prominence of the environment with respect to the other media when it is the main focus of the user's task, as in the case of visitors in a physical museum. We have also discussed how multimodality can get round the staticity of the environment. Reference has been made to generation in HIPS and HyperAudio, two location-aware prototypes for individualised presentation of cultural sites.

References

G. Benelli, A. Bianchi, P. Marti, E. Not, and D. Sennati. 1999. Hips: Hyper-interaction within physical space. In *Proceedings of IEEE Multimedia Systems '99, International Conference on Multimedia Computing and Systems*, Florence, June 7-11.

A. Bianchi and M. Zancanaro. 1999. Tracking users' movements in an artistic physical space. In *Proceedings of I3 Annual Conference*, Siena.

M. Bordegoni, G. Faconti, M.T. Maybury, T. Rist, S. Ruggeri, and M. Wilson P. Trahanias. 1997. A standard reference model for intelligent multimedia presentation systems. *Computer Standards and Interfaces*, 18:477–496.

S. Feiner, B. MacIntyre, T. Hollerer, and A. Webster. 1997. A touring machine: Prototyping 3d mobile augmented reality systems for exploring the urban environment. In *Proceedings of ISWC 97 (Int. Symp. on Wearable Computing)*, pages 74,81, Cambridge, MA, October.

HCIMD, editor. 1998. *Proceedings of the First Workshop on Human Computer Interaction with Mobile Devices*, 21-23 May. http://www.dcs.gls.ac.uk/ johnson/papers/mobile/HCIMD1.html.

P. Marti, A. Rizzo, L. Petroni, G. Tozzi, and M. Diligenti. 1999. Adapting the museum: a non-intrusive user modeling approach. In *Proceedings of UM99*, Banff, June.

E.D. Mynatt, M. Back, R. Want, M. Bear, and J. Ellis. 1998. Designing audio aura. In *Proceedings of the 1998 ACM Confenrence on Human Factors in Computing Systems (CHI'98)*, pages 566–573, Los Angeles, CA.

E. Not and M. Zancanaro. 1999. Reusing information repositories for flexibly generating adaptive presentations. In *Proceedings of IEEE International Conference on Information, Intelligence and Systems*, Washington DC, November 1-3.

E. Not, D. Petrelli, M. Sarini, O. Stock, C. Strapparava, and M. Zancanaro. 1998. Hypernavigation in the physical space: Adapting presentations to the user and to the situational context. *The New Review of Hypermedia and Multimedia*, 4:33–45.

J. Oberlander, M. O'Donnell, A. Knott, and C. Mellish. 1998. Conversation in the museum: experiments in dynamic hypermedia with the intelligent labelling explorer. *The New Review of Hypermedia and Multimedia*, 4:11–32.

D. Petrelli, A. De Angeli, and G. Convertino. 1999. A user-centered approach to user modelling. In *Proceedings of the Seventh International Conference, UM'99*, pages 255–264.

E. Reiter. 1997. Discussion paper: Choosing a media for presenting information. *ETAI News Journal on Intelligent User Interfaces*. http://www.dfki.de/etai/statements/reiter-nov-97-responses.html.

M. Specht and R. Opperman. 1999. User modeling and adaptivity in nomadic information systems. In *Proceedings of the i3 Annual Conference*, Siena, October.

O. Stock. 1999. Was the title of this talk generated automatically? Prospects on intelligent interfaces and language. In *Proceedings of IJCAI99, the 16th International Joint Conference on Artificial Intelligence*, pages 1412–1419, Stockholm. Invited talk.