Tailored Audio Augmented Environments for Museums

Lucia Terrenghi
Fraunhofer Institute for Applied Information Technology
Schloss Birlinghoven
Sankt Augustin, 53754, Germany
+49 (0) 2241 / 14 - 2561
Lucia.Terrenghi@fit.fraunhofer.de

Andreas Zimmermann
Fraunhofer Institute for Applied Information Technology
Schloss Birlinghoven
Sankt Augustin, 53754, Germany
+49 (0) 2241 / 14 - 2561
Andreas.Zimmermann@fit.fraunhofer.de

ABSTRACT
The paper deals with the design of an intelligent user interface augmenting the user experience in a museum domain, by providing and immersive audio environment. We focus on the issues concerning multimodal interaction, by taking into account aural-visual perception principles. In addition we highlight the potential of augmenting the visual real environment in a personalized way, thanks to context modeling techniques. The LISTEN project, a system for an immersive audio augmented environment applied in the art exhibition domain, provides an example of modeling and personalization methods affecting the audio interface in terms of content and organization.

Categories and Subject Descriptors
H.5.1. [Multimedia Information Systems].

General Terms

Keywords
audio interfaces, augmented environment, user interface, modeling techniques, personalization, museum.

1. INTRODUCTION
One core idea of ubiquitous computing [8] is that everyday life objects and environments will acquire computational capacities, e.g. sensing data in the context, elaborating them and adapting the system to the context of interaction, thus providing new functionalities. By naturally moving in the space and/or by manipulating physical objects in our surroundings we will act upon information in the virtual layer. This enables new augmented user experience and funds the basis for new challenges in user interface design.

In this paper we describe the LISTEN project as example of tailored audio augmented environment for the museum domain: the aim is to highlight the potential of augmenting the user’s experience with the employment of personalized audio information. We attempt to enrich the visitor’s experience providing audio content combined with the art pieces and tailored on the basis of parameters such as focus, speed or motion style.

2. PRINZIPALS OF AURAL-VISUAL INTERACTION
To date, perceptual and cognitive aspects have mostly been researched separately for each modality, without taking into account possible interactions between different sensory modalities. Current workload models suggest that humans have independent resources available for the processing of information on each sensory modality. Results in cognitive psychology have provided evidence, however, that extensive multimodal links actually influence multimodal information processing.

Bringing together aural experience and visual domain offers an interesting context of interface design and cognitive psychology exploration. There is an inherent integration of the human senses [4, 5]: in everyday perception of sound in the environment, people tend to visualize a sound by associating it with the source they believe produced it. In synaesthesya, the most extreme form of sense integration, a person exposed to stimulation of one sense also experiences physical sensations of another sense. Chitowic [2] has documented cases in which people actually see colors when hearing music and feel shapes while tasting food.

Michel Chion’s [1] thesis about the role of sound in films is that “The audio-spectator considers the elements of sound and image to be participating in one same entity or world. The result of the audio-visual contract is that one perception influences the other and transforms it. We never see the same thing when we also hear; we don’t hear the same thing when we see, as well”. According to Chion, people possess a natural urge to merge sounds and images into a cohesive whole as a strategy for making sense of the world. He discusses many uses of sounds including its use for punctuation, for unification of images which would otherwise seem unrelated, and for creating a sense of anticipation about what will follow. In his discussion on how sound influences our perception of time, including our perception of movement, speed, rhythm, and pacing, he claims that “The eye perceives more slowly because it has more to do at once; it must explore in space as well as follow along in time”. He says that people are more capable of acutely tracking the details of motion with the ear, than with the eye, and the people who posses both sight and hearing can usually understand spoken language faster than they can read.

3. LISTEN: A TAILORED AUDIO AUGMENTED ENVIRONMENT
The project here presented develops a tailored, immersive audio augmented environment for the visitors of art exhibitions. The Listen project takes the challenge to provide a personalized immersive augmented environment, which goes beyond the guiding purpose, and is based on the combination of aural and visual perception. In October 2003, an installation of the LISTEN application has been applied for the visitors of the August Macke art exhibition at the Kunstmuseum Bonn [7] in a Macke...
The users of the LISTEN system move in the physical space wearing wireless headphones, which are able to render 3-dimensional sound, and listen to audio sequences emitted by virtual sound sources placed in the environment. The visitors of the museum experience personalized audio information about exhibits through their headphones. The audio presentation is indeed adapted to the users’ contexts providing an intelligent audio-based user interface.

In the following sections we focus on the goals to achieve, on what can be adapted within an audio augmented environment, and on the possibilities of combining the order, the source, and the content of audio pieces so as to retrieve a tailored presentation and provide a customized user experience.

### 3.1 The Concept

The key idea of the LISTEN concept [3] is to place the individual perception of space – of visual, auditory and imaginary space and their relationships – at the centre of the interface design so as to convey an immersive user experience. By moving through real space, users automatically navigate an acoustic information space designed as a complement or extension of the real space. Virtual acoustic landmarks play an equally important role than the visual ones for the orientation of the users in this augmented environment. Acoustic labels are attached to visual objects, thus affecting the soundscape and its perception. The viewers become interactors as soon as they move in space and the computer system reacts by presenting sounds via the headphones. A repertoire of distinctive audio information is available, that has been selected by curators and artists, stored in the system and can individually be retrieved. As a difference to a traditional audio guide, the augmented, intelligent, and interactive audio environment has a memory: It registers the repetition of actions and reacts immediately with offering new audio sources. Fine-grain motion tracking is essential for the LISTEN approach because full auditory immersion can only be reached if the binaural rendering process takes into account the rotation of the user’s head.

### 3.2 Audio Augmented Environment Modeling

The physical environment is augmented through a dynamic soundscape, which users experience over motion-tracked wireless headphones for 3-dimensional spatial reproduction of the virtual auditory scene. A sophisticated auditory rendering process takes into account the current position and orientation of the user’s head in order to seamlessly integrate the virtual scene with the real one. Speech, music and sound effects are dynamically arranged to form a personalized and situated soundscape, offering information related to visual objects placed in the scenery.

In Figure 1 the application of LISTEN in a museum environment is sketched. The zones, which are connected to the visual objects placed within the environment, are divided into object zones and near fields. The object zones are associated with specific sound information about their connected visual object from a general perspective. The near fields are connected to smaller parts of the visual objects and contain more detailed sound information. Within the near field the orientation of the user’s head plays a more important role than in the general object zone. The right hand picture shows a sound emerging directly from the painting.

### 3.3 The Context Modeling Process

To date, adaptive systems mostly adapt information selection and presentation to the user’s goals, preferences, knowledge, interests, etc. In most cases, the user model acquisition is driven by monitoring the activities of users with the user interface or by an analysis of their connection and device characteristics. Additionally, nomadic information systems utilize localization technologies to adapt to a richer context model. These technologies allow a tracking of the user’s movements in physical space. In the HIPS project [6] these aspects are modeled to achieve a better fit of information to the users’ current context.

The project LISTEN goes beyond this in having a rich context model combined with fine-grained localization technologies. Based on a complete model of the user’s visual environment, the LISTEN user modeling component can build hypotheses about the user needs based on the time spent listening to certain audio streams or even from the user gazing at a certain part of a visual object. This enables the LISTEN system to build better hypotheses about the user’s interests and preferences. By using sensor data, the users are tracked in the information space in the same style a location tracking system tracks users in the physical space. The enrichment of information items with significant meta-data enables the personalization and customization of information offers. Our approach divides the personalization process into four steps: information collection, modeling, controlling and rendering. Each step fulfills a certain role within the user modeling process. The next subsections describe these steps in more detail.

#### 3.3.1 Information Collection

A network of sensors is placed in the environment and connected to variable parameters of the domain. These sensors are used for recognizing changes within the environment and especially for the perception of the users’ interaction with this environment. An observation module receives all incoming events sent by the application. These event descriptions are pushed into a database. Thus, an event history for every user is saved and an implicit user profile is recorded.

#### 3.3.2 Modeling

By the means of statistical models, the implicit user profile already allows the deduction of valuable information that can be used for standard adaptation activity (e.g. the more time the user spends with a specific visual object, the more s/he likes it). This
implemented domain-dependent methods directly change variable involved, impressions and refinement issues were brought out.

5. CONCLUSIONS AND EVALUATIONS
rules on the level of possible connection categories.

forced to design a complete sequence of information presentation speech. In this way the curator of an interactive experience is not sound items on several channels, i.e. typically music, effects, and

This multidimensional classification of the sound items allows a classified in a category system with several dimensions describing
Macke exhibition to enable the description of these small

A first positive result was the success of the synaesthetic experience: the visitors enjoyed the combination of audio-visual perception and felt as the interaction with the real visual objects was augmented. At the same time, curators appreciated the possibility to deliver content concerning the artworks in an innovative, enriched, and less descriptive way.

Critical points in the domain model were noticed: the zones of interaction surrounding each artwork were sometimes too small, thus forcing the visitor to approach the artwork very closely. Especially in the case of overlapping zones, the user could hardly localize the boundaries of the object zones. In order to overcome these problems, we are working on making the zones more flexible through “breathing zones” in which a sound is more attached to the user’s behavior. In this sense auditory icons, providing some landmarks in the virtual environment navigation, were inserted in the audio presentation in order to make the user “feel” the interaction with the environment.

Further effort needs to be put into the recognition of the user’s real focus as well: the tracking system senses the visitor’s position in the space, but his focus can be on an object belonging to another sound zone. Besides, some visitors could not realize whether the changes in the audio virtual environment were due to their movements in the space or were part of the audio sequence. A minority did not always realize the personalized character

A main criticism in the evaluation was the selection process of stereotypes for content adaptation. The visitors neither do like to be clustered or classified, nor state what class they belong to in public: the request to select what kind of art visitors they are generated irritation. Due to this fact and due to the decision of not providing any input devices, we intend to gather more significant data about the user. The new test scenario will provide stereotypes that are hidden from the user and that represent the user’s motion data about the user. The new test scenario will provide stereotypes for content adaptation. The visitors neither do like to be clustered or classified, nor state what class they belong to in public: the request to select what kind of art visitors they are generated irritation. Due to this fact and due to the decision of not providing any input devices, we intend to gather more significant data about the user. The new test scenario will provide stereotypes that are hidden from the user and that represent the user’s motion style like sauntering, goal-driven or standing still, because they have less social impact, are easy to detect and to revise.

6. REFERENCES

4. AUTHORING THE SOUNDSCAPE
The main entities for structuring and generating presentations are not the artworks as such, but rather the parts of environment and background descriptions. We designed a domain ontology for the Macke exhibition to enable the description of these small information items on a variety of dimensions. Sound items can be classified in a category system with several dimensions describing the sound items technically and from a stylistic point of view. This multidimensional classification of the sound items allows a variety of sequences and presentation styles even combining sound items on several channels, i.e. typically music, effects, and speech. In this way the curator of an interactive experience is not forced to design a complete sequence of information presentation but can combine resources in a collage style or define sequencing rules on the level of possible connection categories.

5. CONCLUSIONS AND EVALUATIONS
In order to evaluate the LISTEN interface some demonstrations took place at the Kunstmuseum in Bonn. The outcomes of these tests enabled a preliminary evaluation: in two workshops in which testing visitors, art curators, sound designers, and artists were involved, impressions and refinement issues were brought out.