LISTEN
Augmenting everyday environments through interactive soundscapes

Olivier Warusfel, IRCAM, Paris, France.
[olivier.warusfel@ircam.fr],
Gerhard Eckel, Fhg-IMK, Bonn, Germany.

Abstract— **LISTEN is a European funded project aiming at creating a new medium: the immersive audio-augmented environment. The concept consists in superimposing a virtual soundscape to the real environment users are exploring. Wireless tracking technology, 3D audio rendering and interactive scenario design allow composing new relationships between sound, body and space. LISTEN developed the concepts and technology necessary to approach this vision and produced audio-augmented reality applications dedicated to media art, museum pedagogy and product marketing. A complete realization of a large-scale audio-augmented reality system for in-door application has been demonstrated and evaluated in public exhibitions at the Kunstmuseum of Bonn. The present document describes the main components of the system architecture, the main features of the artistic and pedagogical demonstrators together with some comments on the design of the augmented world taken from the musical and scientific standpoints.**

I. Concept

LISTEN is a European funded project coordinated by the Fraunhofer Institut für MedienKommunikation in Germany [4][5]. The aim of LISTEN was to create a new medium: the immersive audio-augmented environment (IAAE). The project was driven by the vision of composing interrelationships between sound, space and the motion of users. LISTEN developed the concepts and technology necessary to approach this vision and produced audio-augmented reality applications dedicated to media art, museum pedagogy and product marketing. LISTEN has produced a complete realization of a large-scale audio-augmented reality system for in-door applications. Several demonstrators were developed and evaluated in public exhibitions at the Kunstmuseum of Bonn.

Throughout the project the LISTEN appeared as providing a new type of medium in its unique way of relating body, sound, and space. With LISTEN-medium, a person wearing a wireless motion-tracked headphone is immersed in a virtual auditory scene, which augments the real environment the person explores through locomotion. Although LISTEN does not define how the real environment is augmented, one of the driving forces of the project was to make the headphones "transparent" in the sense that the user would not be aware of exploring two environments at the same time (the physical one and the virtual auditory one). Using techniques of binaural auditory rendering and room acoustic simulation, a virtual auditory scene can be synthesized, which seamlessly augments the real space in the sense that it provides perceptually plausible cues related to the objects existing in the real space (e.g. by rendering virtual sound sources in an acoustic environment which shares the acoustic signature with the real environment or by positioning virtual sound sources at locations coinciding with objects in the real space). It also appeared that realism, not only on a semantic level, but also on a pure acoustic level was not at stake for the acceptance of the Listen-medium. On the contrary, this acceptance strongly depends on the congruence, possibly arbitrary, between acoustic and idiothetic cues (sensations linked to self-motion, including vestibular and proprioceptive cues) [12].

A possible working definition of an IAAE is: a multi-sensory form of content which augments real (mainly visual) content with virtual auditory content where the form of augmentation may range from "realistic" to "abstract." The former would reflect the real room acoustic situation (location of sounds, room response) and create perceptually plausible cues whereas the latter could create situations, which are perceptually implausible (with respect to our experience in real acoustic situations) but consistent with respect to the scenario. An example for such a more "abstract" way of relating the virtual soundscape with the real space could be invisible spatial zones, which are characterized by different room acoustic signatures, which, in a physical environment, are hard or impossible to produce. A typical LISTEN application would combine "realistic" and "abstract" elements of augmentation for forming the right mix needed for the particular use of the IAAE.

II. Architecture

The global architecture of the LISTEN system together with the partners involved in the different research and development

This work was supported by the Information Society Technology program of the European Commission under the R&D contract IST 1999–20646 http://listen.imk.fraunhofer.de.

tasks are depicted on figure 1. In order to relate the synthetic virtual auditory environment to the real space, the position and orientation of the user's head needs to be known. The user is equipped with a wireless tracking device mounted on the headphone. The tracking data, position and orientation of the head, is sent to a program that maintains the description of the world, the history of user's behavior and the interactive scenario. According to these parameters a soundscape is composed in real time. It consists in a series of sound objects (sound files, audio effects, etc...) together with the description of their spatial organization, updated in real-time with respect to the listener's position and orientation. These data are transmitted to a real-time audio rendering software dedicated to 3D sound spatialization, and which synthesizes the sound scene using the binaural format dedicated to headphone listening. The binaural audio signals are then transmitted to the headphone using wireless communication.

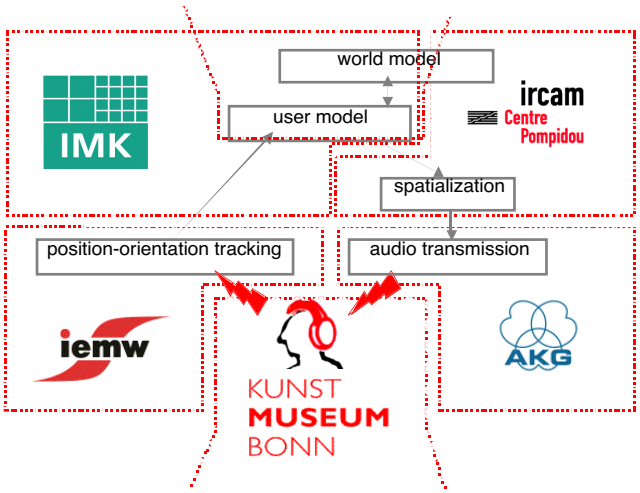

Figure 1 *Architecture of LISTEN system and partners involved in the project*

A. Tracking system

A **tracking system** based on RF-burst signals was designed for acquiring wirelessly the position and orientation of several simultaneous users (8 in the demonstrators). This task was developed by the *Institut für Industrielle Elektronik und Materialwissenschaften* from the Technical University of Wien- IEMW [1][2]. Transmission of short RF-bursts is a smart technique to resolve the indoor multi-path problem combined with high ranging accuracy.

The radio tracking system comprises transmitter /receiver units on the one hand as well as a central signal-processing and post-data-processing unit on the other hand. With regard to multi–user requirements (i.e. multiple persons, low power and low weight at the user's headset), a design was selected where the user is carrying the wireless navigation transmitter in contrast to concepts where the object to be tracked is receiving information (figure 2). The navigation signal transmitted from the user's headset is received by navigation receivers spread through the serving area at precisely known locations. The central signal processing unit collects the data accumulated by the tracking receivers and calculates the absolute position (x, y, z) and the orientation (azimuth) of each navigation transmitter.

Figure 2 *Tracking transmitter integrated on headset*

B. World and User modeling

Tracking data are sent to a **world model** (i.e. a model of the real environment), which is used to localize the user in the real and virtual environment, i.e. interpret his or her movement with respect to the modeled real or virtual objects of interest (e.g. user enters into a zone, looks at an object, etc.). This component may be extended by a **user model**, which interprets the user's behavior with respect to various criteria and eventually data acquired from previous visits or from visits of other users. In running mode, these components compose interactively the sound scene according to the author scenario and to the position and behavior of the listener. In editing mode, these components allow defining the sound objects and events of the virtual soundscape, their spatial organization and time scheduling. Obviously, as based on listener interaction, this edition tool must not only integrate a collection of predefined spatial and time interaction functions (time and spatial triggering, fading, etc...) but should preferably propose a scripting environment in order to allow the composition of more complex interaction and scenario. Different authoring tools were studied and developed during the project at Fraunhofer–IMK and at Ircam [8][6][7].

Figure 3 *Experimenting on-site authoring*

C. Spatialisation rendering

The audio rendering component, which synthesizes the virtual auditory environment was based on Ircam's Spatialisateur® [9], here used with binaural rendering and room simulation techniques. Audio augmented reality, as designed in the context of the LISTEN project, intends to

superimpose virtual sounds to a physical space. Since the audio message is presented over headphones, this implies the use of the binaural technique, which provides a powerful approach for creating highly immersive and interactive soundscapes. The rendering component is controlled by the world model / authoring tool component and produces the binaural signal the LISTEN users will receive through wireless headphones. One of the difficulties when applying binaural rendering is the need for individual adaptation linked to the use of audio filters that should respect individual acoustical cues associated to the head morphology (HRTFs). This problem can represent strong limitations when addressing public VR applications since it is not yet possible to measure or estimate accurately these individual filters in daily environments. The automatic adaptation or selection of individual HRTFs is still on progress research [10][11]. However, in LISTEN-medium context, the integration of interaction between auditory and idiothetic cues seems to be a factor allowing to relieve this constraints[3]. This part was developed by IRCAM.

D. *Audio device*

A wireless headphone system, which is integrated with the tracking system (typically includes a tracking emitter or target – depending on the type of tracking system used). This part was developed by AKG.

III. APPLICATIONS

In the context of the LISTEN project, different types of applications have been explored : media art, museum pedagogy, product marketing and scientific experiments on spatial cognition and investigation for application in psychopathology rehabilitation. With the installation "Raumfaltung," LISTEN has been used in the artistic domain, where it has been proven to be a very powerful tool for exploring new ways of combining different artistic disciplines such as visual arts, music, and literature. Media art is an interesting field for exploring new ways of producing multisensory content before being adapted to other application domains such as entertainment, pedagogy, or marketing. With the environment "Macke Labor," LISTEN has been used as a tool for museum pedagogy and was designed as an audio-visual environment on artwork by August Macke. "Raumfaltung" was on display for 10 weeks (July 6th - September 14th 2003). During this period 4310 visitors walked into the installation using LISTEN-headphones. The "Macke Labor," environment was on display for 6 weeks (October 15th – November 23rd 2003) and welcomed more than 1600 visitors.

IV. SPATIAL INTERACTION

During the presentation, we will comment on different spatial interaction that were studied during the elaboration of the exhibition content and / or scientific experiments. Different type of elementary interactions have been tested especially in order to trigger sound events according to the spatial behavior of the user. A typical case is when desiring to attach a sound object to an existing visual anchor in the real environment. Various strategies were investigated by Fhg and exhibition curators from KunstMuseum Bonn in order to find a sonification paradigm allowing to symbolize the link between a real object and a sound source or event. The main approaches range from triggering sonic icons or "ambient sounds" according to the object-listener relative position (gaze direction, distance), to the application of different realistic or symbolic sound spatialisation to the audio content. More elaborate interactions take into account the time dimension, slaving the scheduler of a sound object or event to the detection and analysis of the listener's motion (speed, activity, ...). This introduces the body as an interface to the content and the listeners showed fast recognition and intuitive use of the proposed interaction mechanisms. The introduction of temporal interaction and control is mandatory in order to organize the content according to narrative dependencies or musical constraints.

Although still needed to be explored, the Listen-medium showed the consistency of public VR applications based on the composition of interactive soundscapes. Listeners immerged in such interactive audio-augmented environments assessed very positively the concept and found their navigation experience very engaging and interesting. This confirms the important contribution of spatial audition for improving navigation tasks [12] and for eliciting the *presence* in VE and its potential for a variety of public applications.

REFERENCES

[1] Alois M.J. Goiser, About the Robust Nature of Spread-Spectrum Systems, IEEE-ISSSTA 2002

[2] Alois M.J. Goiser, Wideband Wireless Digital Communications, Editor: A. Molisch, Prentice-Hall.

[3] Blum A., Katz B., Warusfel O., "Eliciting adaptation to non-individual HRTF spectral cues with multi-modal training", submitted to 7èmeCongrès Français d'Acoustique / 30. Deutsche Jahrestagung für Akustik, Strasbourg, March 2004.

[4] Eckel, G., "Immersive Audio-Augmented Environments - The LISTEN Project", ed. B. Banissi, F. Khosrowshahi, M. Sarfraz and A. Ursyn, Proceedings of the 5th International Conference on Information Visualization (IV2001), IEEE Computer Society Press, Los Alamitos, CA, USA, 2001.

[5] Eckel, G., "The LISTEN Vision", Preconference Proceedings of ACM SIGGRAPH and Eurographics Campfire on Acoustic Rendering for Virtual Environments, pp. 55-58, May 26th-29th, 2001, SnowBird, Utah, USA, 2001.

[6] Delerue O., Warusfel, O., "Authoring of virtual sound scenes in the context of the LISTEN project", proceedings of the AES 22nd International Conference on Virtual, Synthetic and Entertainment Audio, Helsinki, June 2002.

[7] Fink J., Kobsa, A., "User modeling for personalized city tours". Artificial Intelligence Review, 18(1):33–74, 2002.

[8] Goßmann J., Specht M., "Location Models for Augmented Environments", Workshop Proceedings of Location Modelling for Ubiquitous Computing (Ubicomp2001), pages 94-99, 2001.

[9] Jot, J-M., Real-time spatial processing of sounds for music, multimedia and interactive human-computer interfaces, Multimedia Systems, v.7 n.1, p.55-69, Jan. 1999

[10] Katz B., "Estimation of individual inter-aural time delay (ITD) using statistical based spatial functions and morphological data", submitted to JASA, Journal of the Acoustical Society of America.

[11] Rio E., Vandernoot G., Warusfel O., "Optimizations of Multi-Channel Binaural Formats Based on Statistical Analysis", Forum Acusticum, Sevilla, September 2002.

[12] [Viaud-Delmon 2004a] Viaud-Delmon I.,Seguelas A., Rio E., Jouvent R., Warusfel O., "3-D Sound and Virtual Reality: Applications in Clinical Psychopathology", Cybertherapy 2004, San Diego January 2004.