Proceedings of ICAD 04-Tenth Meeting of the International Conference on Auditory Display, Sydney, Australia, July 6-9, 2004

CONNECTING VIRTUAL SOUND AND PHYSICAL SPACE IN AUDIO-AUGMENTED ENVIRONMENTS

Joachim Goßmann

ZKM – Center for Art and Media
Institute for Music and Acoustics
Lorenzstraße 19, D-76135 Karlsruhe, Germany
gossmann@zkm.de

ABSTRACT

This paper describes strategies to create multimodal coherences between sound and space in Audio-Augmented Environments. Visitors to AAEs wear motion-tracked wireless headphones displaying a location-aware, adaptive audio presentation augmenting the physical space with virtual sound-scenes. The work underlying this paper was done by the author during the course of the LISTEN project at Fraunhofer IMK[1][2]. In the first half, the concept of the 'perceptual device' is introduced, which allows attaching audio content to a physical space. In the second part this is supported by practical examples from the publicly accessible AAE "Macke Labor". The "Macke Labor" was installed at the Kunstmuseum Bonn, opening in October 15th 2003. It was developed by the author with an interdisciplinary team of collaborators. The paper is rounded off with the description of the sonification technique used in the evaluation of the recorded tracking and event data of the "Macke Labor".

1. INTRODUCTION

This paper refers to research activities during the LISTEN project. The LISTEN project is a project supported by the 5th IST (Information Society Technologies) framework programme of the European Union. The project explores a novel form of multi-sensory content: the 'Immersive Audio-Augmented Environment'[1]. Within the project, technological components, concepts and demonstrators were developed as collaborations between coordinating partner Fraunhofer IMK in Bonn, Germany with the Kunstmuseum Bonn, the IRCAM in Paris, France, and the Institute for Material Science (IEMW) at the Technical University of Vienna, Austria. As an industrial partner, AKG Acoustics Vienna participated in the project. The developed technologies include a high-resolution electromagnetic tracking system, a wireless digital headphone system, as well as infrastructures and authoring-tools for the creation and presentation of spatial sound-scenes and non-linear time-based narratives. Two full-scale applications were produced as public exhibitions in summer and fall of 2003, both of which were installed at the Kunstmuseum, Bonn.

1.1 The pre-condition of an 'Audio-Augmented Environment' in the sense of LISTEN

The visitors to such an environment wear motion-tracked wireless headphones. This allows the creation of coherences between two per-se independent time-based processes: The spatial behaviour of a visitor in an exhibition room, and a headphone-based presentation of audio.
The use of a high-resolution tracking system in combination with binaural rendering primarily allows to populate the physical space with virtual sound-sources and room acoustic models: An 'augmentation' created along physical and perceptual paradigms (interactive sound spatialisation) [3][4]. However, it also yields new questions about structure and progression in the sounding content: Next to enabling a spatial coherence between virtual sound-scene and physical space, a tracked wireless headphone system as suggested by the LISTEN project enables coherence between the temporal progression of the sound presentation and the spatio-temporal behaviour of the visitor. The following paragraphs will focus on the use of the resulting layers of coherence between sound, motion and space in combination or dissociation to create new kinds of auditory-spatial experiences for the visitor.

1.2. The visitor timeline

The concept of an exhibition space is in general that of a static physical presence which does not change its appearance or structure over time. However, the subjective timeline of the visitor dissolves this static concept of space into a stream of changes and events organized along the evolving path of the visitor's motion exploring the space. The user may enter and leave subdivisions of the space - spatial features or objects in the room may be encountered and receive special attention – the visitor may focus outward on objects placed in the space, or inward, on inner images. The second sight of an object in the space may see something new and different than the first glance. A user may explore the room in a random manner, or follow a fixed path (e.g. strictly along the walls). In this way, the space as subjectively perceived by the visitor indeed changes over time and along the way generates a stream of spatial and behavioral visitor contexts. Depending on the type of room to be augmented (the preceding description describes a typical visit to an exhibition space within a museum context), there might be typical behaviours for visitors to the spaces, resulting in a specific set of possible activities or contexts of the visitor. These contexts can be ascertained from a location model [5] evaluating the visitor's position within the space and the visitor's field of vision, and also from applying filters to the generated tracking data. This allows to extract events of changes to the position, behaviour and history of the user, resulting in a multi-dimensional timing substrate of the visit to the augmented space. The timing substrate can in turn be used to adapt and shape the content of a running audio presentation.

It has to be noted that depending on the intended application of the AAE, the location model does not necessarily have to adhere to the perceptual structure of the physical space, but may also create an invisible 'virtual architecture' within the space. This can generate relationships between the resulting audio presentation and the visual space that range from the audio being an obedient location-aware accompaniment to the visual experience, providing the visitor with fitting information, narration, atmosphere or other 'suitable' presentations, to a presentation completely dissociated from the visual structure of the space, creating a world of its own - virtual objects, architectures, sound-environments or invisible interactive sculptures – or simply a completely detached 'personal stereo' presentation. In a dissociated relationship, the interaction may become intransparent to the user, as visual motivations to changes in the sound are missing – the user might feel left alone in connecting the 'invisible dots'. (This might nevertheless be an interesting strategy to challenge visitors in a playful immersive AAE).
Audio-Augmented Environments are multi-modal by their nature, as the visitor roams a physical, visual space, while listening to an audio presentation augmenting the space. If the environment is supposed to create an intuitive experience, the relationship between space, sound and motion/interaction should be obvious, and the structuring of the visual space should accordingly play an important role in the design of the environment.
The question arises how to create relationships between the sounding content and the visual elements in the physical space of an AAE.

1.3. Semantic coherence

In audio presentations using spoken word, semantic coherence can be used to guide the attention of the visitor: A voice can point out a certain object or landmark within the space, creating a semantic network between the spoken word, the feature of the physical space referred to and any audio content that might be displayed simultaneously or in temporal proximity to the speech. This way of creating coherence does not necessarily require a tracking system or an AAE – it is in fact a standard in contemporary audio guide systems or even used by personal museum guides. The shortcomings of this method however are obvious: Pointing the visitor toward a certain feature of the space or demanding the user to behave in a certain way with words may result in a cumbersome and clumsy audio presentation. Based on the possible coherences between the visitor position and motion and the audio presentation mentioned above, spatial and temporal convergences can replace speech in networking audio content to features and objects of the visual space, utilising perceptual qualities of audio (spatial perception and temporal structure) to convey similar information in a much more subtle and functional way:

1.4. Spatial coherence

The coherence that first comes to mind when thinking about AAEs is the creation of a virtual sound-source at a specific location within the space (spatial convergence). Depending on the quality of the binaural rendering and the nature of the visual object, this allows creating a perceptual grouping of the sound displayed to the visual object present in the same location – without necessarily verbalising the connection. This strategy may also be used to guide the attention of the user into a certain direction. The sound-design becomes crucial, encompassing the creation of the signal to be spatialised as well as the properties of the added virtual room acoustic rendering. Considerations such as suggested by Lennox et al. [6] about the design of 3D Audio Environments can be applied. The attachment of content to the space by means of spatial coherence works best with clearly visible landmark objects that dominate the space. Informal tests during the project showed that synchronising auditory events with a visual cue such as a synchronous lighting effect or an event on a video monitor (added temporal convergence) greatly enhances the perceptual localisation of the sound. For best results, the spatialised signal should be familiar to the visitor, have a broad frequency spectrum and seem plausible within the current context or surrounding.

1.5. Spatio-temporal coherence

The second strategy specific to AAEs is the creation of auditory-spatial coherence through temporal convergences between distinguished perceptual events occurring in the spatio-temporal substrate during the room exploration and notable events in or changes to the audio presentation, creating a resulting perceptual event that could be described as "synaesthetic synchronicity". For example, employing simultaneity between the occurrence of a perceivable process in the interaction of the user - such as the appearance/disappearance or motion of objects in the visitor's field of vision through motions of the head, or the entering or leaving of a visually marked subdivision of the exhibition space – and direct auditive feedback, sounds can be attached to objects and subdivisions within a space. The feedback can be designed as changes to the audio presentation like the appearance or disappearance of sound or more subtle, like changes in sound parameters such as color or volume. The grouping becomes stronger, the more pronounced, clearly marked or impulsive both of the grouped perceivable processes are. In other terms, this can also be described as a sonification of events occurring in the temporal substrate mentioned in (1.2). This does not necessarily require the sound to be rendered binaurally.
Converging to events in the field of vision (orientation) and events derived from the structure of the space to be augmented (position) can be regarded as primitives, each implying their own demands on the design of the space and the sound. Informal test scenarios showed that attaching a sound to the appearance of an object in the user's field of vision works best if the visual object has a high contrast to the background, and the sound has an impulsive decaying envelope, possibly sustaining the sound until the object leaves the user's field of vision. This effect can be enhanced by combining this strategy for coherence with the localization of the sound at the visual object.
The attachment of sounding content to a subdivision of the space requires the subdivision to be clearly marked. In order for the sound to attach itself to an area within the space rather than coming from a designated direction or object, the sound needs to have a rather constant character and a "diffuse" (or decorrelated) characteristic, avoiding the generation of a specific localization. If specific spatialisation is however used in the sound, it needs to surround the user evenly rather than pointing at a specific direction.

All of the mentioned strategies to create coherence can be used in combination, strengthening the attachment of the auditory content to the visual object. We call these combinations of spatial cues with direct interaction between sound and motion that create new structures of coherence

between sound and space in AAEs 'perceptual devices'. Examples are given below.

2. EXAMPLES

The following examples were used in the publicly accessible AAE "Macke Labor" that was installed at the Kunstmuseum, Bonn from October 15th to November 23rd, 2003. Within the context of didactic mediation for an art museum, the "Macke Labor" applies a radiophonic strategy of sound utilization [7] expanded by aspects of 'open form', and the methods of sound processing, spatialisation and interaction provided by the infrastructures developed during the LISTEN project. It was produced by the author with an interdisciplinary team of collaborators from the field of museum pedagogy, architecture, radiophonic art, media studies and VR-research. The environment was focused on the Bonn early expressionist painter August Macke – whose works became the visual 'objects of interest' within the AAE. The physical site of the "Macke Labor" (architect: Ingo Bergmann) was architectonically structured into four subdivisions: An "Entrée" situation and three zones, creating a spatial topography of the three important places of work in the life of the painter August Macke: Bonn, Germany, Hilterfingen, Switzerland (Schweiz), and his famous journey to Tunisia (Tunis). Each subdivision was equipped with a set of works from the respective period of Macke's oevre.

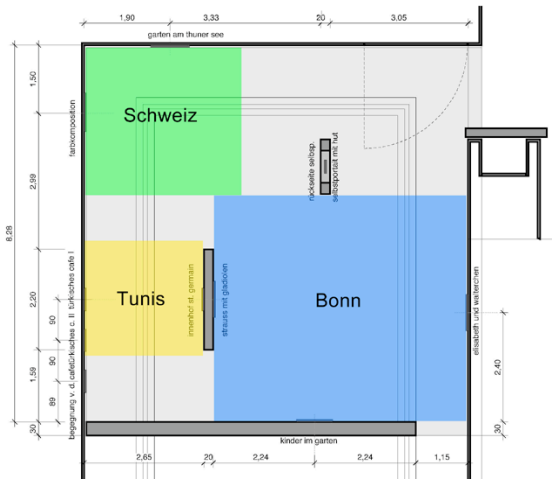

Figure 1: The floor pattern of the "Macke Labor", displaying the three spatial zones.

As a starting point for the design of user interaction and navigation, the observed interaction of visitors to non-augmented museum spaces was used as described in 1.2. The audio content in the "Macke Labor" is either attached to an 'object of interest' – in this case the paintings of August Macke – or to one of the subdivisions of the space.
The audio content is modeled into 'modules' that on a common timeline define the processes of sound generation, spatialisation and possible interactions with sound parameters, generating the audio-behaviour of the environment in all aspects for a certain amount of time.
The temporal behaviour of the AAE was created by chaining together these time-modules into a continuous presentation in reaction to the changes happening in the spatio-temporal substrate (1.2.). This process will be topic of another paper.
The audio content in the "Macke Labor" has multiple tasks to fulfil:

- a semantic task, to conveying information, 'meaning' and narrative content
- the attachment of this meaning to the objects and subdivisions of the exhibition room
- creating necessary functional feedback about the interaction of visitor with the AAE.

In the "Macke Labor", a semantic layer of spoken word was continuously present: During the discussions with the curators at the Kunstmuseum Bonn, the use of solely non-verbal sound elements was dismissed as being too 'unspecific' for the context of mediation in an art museum. It was decided that the use of non-verbal elements for the sonification of system functionality (such as to signify the system has detected that the visitor gazes on one of the paintings) was also not appropriate to be used in the art-museum context of the "Macke Labor". As a consequence, the non-semantic functionalities of the sound were subtly integrated into the modules bearing the narrative content.

2.1. Creating perceptual networking between content and space in the "Macke Labor"

The design and use of 'perceptual devices' is not only a question of technology, it also needs to be discussed with its implications on the dramatic and choreographic aspect of the visitor experience in mind. Depending on how the audio content 'addresses' or interacts with the exhibition space and the motion of the visitor, its perception may shift from being functional or instructive to becoming playful and whimsical, from sharpening the visitor's sense for the experience of sound and movement, possibly distracting attention from the paintings, to becoming completely obscure. It determines to what degree the visitor needs to make a conscious effort in the use of the environment, or is able to forget about the tracked headphone altogether. It was therefore a challenge in the creation of the "Macke Labor" to design the interaction as a balancing act between animation and playfulness on the one hand, and the desired intuitiveness and 'invisibility' of the AAE on the other – a compromise between the functional gathering of information about the displayed art object (comparable to an audio-guide) and the sensitisation of perception by the employment of a mutli-sensory environment which is a part of the overall design question posed by the "Macke Labor": What can a modern-day form of mediation 'look', 'feel' and 'listen' like at all?

2.2 Example1: Encountering the art-historian. A standard 'perceptual device', utilizing a localised sound-source and amplitude based interaction in combination

The basic layer of narration in the "Macke Labor" is carried by a female 'art-historian' and a male 'trainee'. The visitor encounters these voices in several occasions in the room. They are conveying the art-historic portion of the narration. The perceptual device applied to group the perception of the voice to the visual perception of a painting is twofold:
First, the virtual sound-source is located at the painting referenced by the voice (spatial coherence). This results in a panning of the sound-source in coherence with the motion of the painting in the field of vision. A feature to be noticed here is that because sound sources in front are not easily localised, the location of the voice is 'somewhere in the middle' as long as the visitor looks straight at the painting. The attachment of the voice to the painting becomes more

apparent when the visitor turns the head, resulting in a panorama shift that allows a better localisation of the sound source.
Second, the amplitude by which the voice is played to the visitor changes with the angle in which the head of the user points. When the user is pointing straight at the painting, the level is highest, decaying when the user transgresses a certain angular tolerance (amplitude mapping to head orientation). This is a feature not inherent to sounds in physical reality, and works very well in terms of signifying the visitor that the presentation is about to stop when the gaze is taken further away from the painting. The same amplitude shift happens when the visitor increases the distance to the painting (amplitude decreasing with distance).
In this example, all layers work towards the creation of the same networking between the audio content and the painting. As the attachment of the audio content to the visual object is clarified so well, the voices themselves may digress more freely from describing the painting without loosing perceptual coherence to the picture (reducing semantic coherence).

Figure 2: A visitor to the "Macke Labor" wearing the LISTEN headphone system

2.3. Example2: An exchange of letters between Elisabeth Macke and German museum staff in the Nazi aera – Applying amplitude-based interaction only.

In another module, letters from the time of third Reich Germany are quoted, illustrating Elisabeth Macke's engagement in protecting the works of her deceased husband from the hands of the curators of the exhibition "Entartete Kunst". An introduction by the art-historian is followed by the recitation of three letters, each of which composed into a specific spatial atmosphere 'surrounding' the user: The sound-sources, apart from the art-historian, are not placed at the painting, some are even located behind the visitor. The grouping of the content to the painting is created by the interaction with the amplitude only, as described above.

2.4. Example3: The "Voyage to Tunisia"

The "Voyage to Tunisia" represents a content which is not attached to a single work of August Macke, but to a subdivision of the exhibition space, displaying works related to or produced during August Macke's famous journey to Tunisia in 1914, in which he was joined by his painter friends Paul Klee and Louis Moillet. The visitor is engulfed by a spatial soundscape composed of on-location recordings from Tunisia (radiophonic artist: Sabine Breitsameter). This sound-scene coheres with the physical space in orientation, however, the sound-events within it are not directly attached to physical objects or works of August Macke. The soundscape is superimposed by two voices, representing August Macke and Paul Klee. While the soundscape coheres with the surrounding space, the two voices remain positioned helf-left and half-right behind the head of the visitor, creating an intermediate layer between the visitor and the images and invisible scenes evoked by the soundscape. The soundscape itself is obviously dissociated from the exhibition space and its architecture, but is, nevertheless, part of a semantic network involving the sounds, the voices and the works created from the journey surrounding the visitor. A second level of coherence is created by means of interaction: The soundscape is only heard within a designated zone in the exhibition space, visually marked by the architecture. When the visitor exits this zone, the audio presentation fades out and pauses, continuing as the visitor re-enters. In this way, the "Journey to Tunisia" employs a combination of semantic- and interaction-based networking to connect audio content and exhibition space.

3. EVALUATING THE VISITS

During the installation of the "Macke Labor", the tracking and event data generated by the visitors was recorded and stored. For the evaluation of the visits, a tool for sonification and visualisation of the visit's progress was developed, allowing to recapture each of the ca. 700 visits as time-lapse. The sonification consists of three layers, one layer displaying the motion of the visitor through the zones of the room, a second layer containing a detailed display of the motions of the user in respect to the paintings, and a third layer which contains a cartoonified version of the actual audio content heard by the visitor. The zones are sonified as keynotes, softly toggled by the entrance and exit of the user to the respective zone. The user's motion is audible via discrete frequency spectra, one per painting, which are controlled in amplitude by the distance between the visitor and the respective painting. For each of the paintings, the degree by which the visitor faces them is illustrated by a lowpass filter whose cut-off frequency is modulated by the angle between the visitor's nose and the respective direction of the painting. The relative motion of the visitor is also audible in panoramic shifts of the spectrum attached to a painting in accordance with the suggested motion of the painting in relation to the visitor. The audio content of the environment was cartoonified using samples of the respective voice actor and granular synthesis.
The evaluation of the tracking paths show that the grouping of the content to the picture with the amplitude-based interaction alone is enough to group the visual perception of the image to semantic content that is rather associative in nature. Concerning the 'unlocalised' presentation attached to the zones, the tracking data suggests that many visitors

rested in front of one of the paintings (visual landmarks) for most of the presentation of the soundscape. Obviously, the attachment of the soundscape to the delimited 'Tunisia Zone' that allowed free motion within the zone was indeed not understood by all visitors. This may be a general problem of mixing content attached to objects and content attached to spatial zones, but points at the difficulty to create spatio-temporal content structures that can be intuitively understood by the visitor.

Figure 3: Two tracking paths as displayed by the evaluation tool: The time is denoted by a color-change from blue to red, lighter color fans denote head orientation, the size of the dot denoting the visitor's position becomes smaller the greater the visitor speed.

4. CONCLUDING REMARKS

The feedback gained from the public evaluation of the "Macke Labor" clearly shows that the interaction strategies described were successfully comprehended intuitively by most visitors. The "Macke Labor" was certainly able to profit from the familiarity of visitors to traditional audio guides. Also the intentional concentration of all visual stimuli in the room on the artworks displayed – common to exhibition design in an art museum – created suitable natural reference points for the visitor interaction and navigation. The "Macke Labor" has also shown that it is possible to create a responsive structure for user navigation through an augmented exhibition space not even requiring the content to be rendered binaurally in order to create a spatial sound scene. The types of interactions tapped into during the course of the production of prototypes and applications, especially the possibilities yielded by the mapping of tracking data to parameters of the sound generation provide a territory for auditory-spatial interrelations that yields room for many discoveries that we still hope to make in the future.

5. REFERENCES

[1] G. Eckel "Immersive Audio-Augmented Environments – The LISTEN Project," in *Proc. of the 5th International Conference of Information Visualisation (IV2001),* IEEE Computer Society Press, Los Alamitos, CA, USA, 2001

[2] The LISTEN website: http://listen.imk.fraunhofer.de

[3] I. Viaud-Delmon, L. Sarlat, O. Warusfel "Localisation of Auditory Sources in Virtual Reality," *submitted to 7ème Conrès Français d'Acoustique / 30. Jahrestagung für Akustik,* Straßbourgh, France,alborg, March 2004

[4] O. Warusfel, I. Viaud-Delmon "Contribution of interactive 3D sound to presence," *Presence 2003, 6th International workshop on Presence,* Aalborg, October 2004

[5] J. Goßmann and M. Specht, "Location Models for Augmented Environments," *Personal and Ubiquitous Computing*, vol. 6, no. 5-6, pp. 334-340, Dec. 2002.

[6] Peter P. Lennox, John M.Vaughan MSc and Dr Tony Myatt "3D Audio as information-environment: Manipulating perceptual significance for differentiation and pre-selection," in *Proc. of the 2001 International Conference on Auditory Display,* Espoo, Finland, July 2001, pp. 155-160.

[7] U. Zindel, W. Rein (ed.), *Das Radio-Feature*, UvK, Germany, 1997.

6. ACKNOWLDGEMENTS

I would like to thank the "Macke Labor"-Team for the wonderful collaboration. In alphabetic order:

Reinhard Behrenbeck, Sabine Breitsameter, Ruth Diehl, Julia Hageberg, Frank Hasenbrink, Annelie Kürsten and, last but certainly not least the coordinator of the LISTEN project, Dr. Gerhard Eckel.