KUNGL TEKNISKA HÖGSKOLAN

CID-180 ● ISSN 1403-0721 ● Department of Numerical Analysis and Computer Science ● KTH

A Series of Three Mixed Reality Workshops: Sounding Out – Sound in Mixed Reality Environment

eRENA ESPRIT Project 25379 Workpackage 7 Deliverable D7b. 1-2

Ramon Gonzales-Arroyo, Heike Staff, John Bowers

CID, CENTRE FOR USER ORIENTED IT DESIGN

Ramon Gonzales-Arroyo, Heike Staff, John Bowers

A Series of Three Mixed Reality Workshops:
Sounding Out – Sound in Mixed Reality Environment

Report number: CID-180, eRENA ESPRIT Project 25379 Workpackage 7 Deliverable D7b. 1-2
ISSN number: ISSN 1403 - 0721 (print) 1403 - 073 X (Web/PDF)
Publication date: August 2000
E-mail of author: bowers@nada.kth.se

Reports can be ordered from:

CID, Centre for User Oriented IT Design
NADA, Deptartment of Numerical Analysis and Computer Science
KTH (Royal Institute of Technology)
SE- 100 44 Stockhom, Sweden
Telephone: + 46 (0)8 790 91 00
Fax: + 46 (0)8 790 90 99
E-mail: cid@nada.kth.se
URL: http://cid.nada.kth.se

Deliverable 7b.1-2

A Series of Three Mixed Reality Workshops: "Sounding Out – Sound in Mixed Reality Environment"

ABSTRACT

The second workshop in the series of practical mixed reality workshops, defined in Workpackage 7b.1, focusses on sound. This is the topic with which most of the eRENA partners have to deal: finding technical and aesthetic solutions for sound, within an image oriented mixed environment. Virtual (and mixed) reality environments often lack a convincing solution for all aspects of sound.

The workshop took place at the ZKM, on December, 2nd and 3rd 1999. Five lectures and demonstrations, presented by eRENA partner's collaborators, alternated with discussions among the participants.

The deliverable documents the content of the workshop structured in four parts: sound in relation to its form of data, in relation to space, to image, and to interaction. One of the achievements of the workshop was in showing how many different relationships are existing around these four aspects. Artistic and pragmatic use of sound in mixed reality environments were regarded as well as the different accesses to visual and audio matters.

Document ID	D7b.1-2
Type	Deliverable
Status	Final
Version	1.0
Date	25 August 2000

Author(s)	Ramon Gonzales-Arroyo (ZKM), Heike Staff (ZKM) John Bowers (KTH)

Contents

1. Introduction

The Workpackage 7b “Mixed Reality Environments” covers the series of three workshops. The second workshop was dedicated to the topic of “Sound in Mixed Reality Environments”. According to the eRENA Project Programme for years 2 and 3, the focus was on further development of current research work of the eRENA partners, especially in the field of real and screen-mediated performances.

In the beginning of 1999, we decided to postpone the workshop from year 2 to year 3, in order to make it more effective as a tool for all partners to evaluate and reflect on work-in-progress. Although the workshop was originally dedicated primarily to the experiences of multimedia opera and networked theatre, we changed the focus and concentrated more on the different eRENA partners’ projects. These were not only finished projects, but also works-in-progress. There are neither standard techniques nor standard protocols for sound in or the audification of mixed environments. The solutions always seem to be rather individual, not only in aesthetic but also in technical and conceptual terms. In the general frame of virtual reality research, which is mainly image oriented, sound and audification is a neglected topic. For these reasons, we became convinced that we needed to put the emphasis on the exchange between the experts from the eRENA partners.

The Workshop “Sounding Out – About Sound in Mixed Environments” was planned as a two day expert meeting, taking place at ZKM on December, 2 and 3, 1999. Half of the time was for lectures and demonstrations of 5 different projects from the eRENA partners (ZKM, GMD, Nott, KTH). The remaining time was allocated to discussions and personal exchanges about questions that are important to the partners.

We invited William Gaver to fill the role of the academic expert who would structure and chair the discussions. He is a senior researcher at the Royal College of Arts, London, and a well known scientist in the field of audio perception and interaction in social and / or virtual environments.[1]

In order to make the best possible use of this expert meeting, co-workers (especially those working on sound) from all eRENA partners were invited, as well as experts from other scientific institutions. The number of participants was limited to 20 in order to encourage open and fruitful discussions.

1.1 Ideas of the workshop

Four main ideas guided the planning of the workshop.

- Sound in mixed or virtual reality environments seems to be something that we might call an oft-neglected topic. Finding technical and aesthetic solutions for sound, within an image oriented mixed environment usually means starting with the images and working on the aspects of sound later. Both areas have underlying different physical, technical and aesthetic rules and treatments. The areas are shared between different experts on signal processing who have to collaborate in order to build a complete environment.

- There seems to be no standard form of access to solutions for sound in mixed environments. Therefore we wanted to promote the personal exchange in the workshop: to learn other person's thinking on tasks and problems as a means of finding new solutions for their own problems. Discussions should deal more with failures than with own successes, because one usually learns more from failures.

- We wanted to encourage the interdisciplinary exchange in all directions and on all levels: sound – image, artistic – scientific, conception – realisation.

- We also wanted to foster personal networks among experts from different institutions. On the other hand, we provided a broad range of topics and aspects, because we believe that creative technical developments are often fostered by examining neighbouring research fields. Thus the whole entire domain of sound was, where not covered fully, at least touched upon.

1.2 Description of the workshop

The workshop took place on December 2 and 3 at ZKM Karlsruhe with the following schedule:

Thursday, 2	
10.00 Introduction and first discussion	William Gaver (Royal College of Arts, London): What is Sound for?
11.30 Lecture	Milena Radencovic (University of Nottingham): Using The Spatial Model of Interaction to control the sound in Collaborative Virtual Environments
12.15 Discussion	Functions for Sound
15.00 Lecture	Gerhard Eckel (GMD Bonn): Sound Rendering in Real and Virtual Environments
15.45 Discussion	Real and Artificial Sounds / Pragmatic and Artistic Use of Sounds
17.00 Lecture	Pierre Dutilleux (ZKM Karlsruhe): AML – Architecture and Music Laboratory
17.45 Discussion	Space or Atmosphere / Localisation of Sounds or Awareness of Space
18.30	Studio-Tour "ZKM Institute for Music and Acoustics"
Friday, 3	
9.00 Lecture	Torsten Belschner (ZKM Karlsruhe):

	Integrating Sound and Image in "SonoMorphis"
9.45 Discussion	Sound and vision 1
11.00 Lecture	Michael Saup (HfG Karlsruhe): Sound and Vision in his Works from the early 90's
11.30 Discussion	Sound and vision 2
13.00 Lecture	John Bowers / Sten Olof Hellström (KTH Stockholm): Intersensory Interaction and Information Display
13.45 Discussion	Performance v. participation: is watching worthwhile?
15.00 General discussion and close	

Twenty-one participants were involved in the workshop. They came from different fields like: composition and/or scientific research; multimedia, computer graphics or dance choreography; pedagogy and/or science.

1.3 Structure of the deliverable

In this deliverable the content of the lectures and discussions are documented, not in the temporal timetable sequence of the workshop, but now organised by three main topics:

- sound and space
- sound and image
- sound and interaction

In the second chapter "Sound and Data in General", short summaries of the main presentations are given, with connections to other related work from the eRENA project noted when relevant.

The final chapter includes possible outcomes and a reflection on the extent to which the main goals of the workshop were reached. We also explicitly note three specific items of work which have been conducted in Year 3 of eRENA with a clear influence from the workshop.

2. Sound and Data in General

William Gaver pointed out in his "Conclusions" lecture that the workshop demonstrated clearly to all participants the enormous magnitude (amazing vastness) of the field of sound in multimedia. In his introductory lecture, he had already remarked on the different functionalities which sound may have, some of which were expressly mentioned in one or other of the lectures / demonstrations. What became more obvious during the workshop was the wide range of different approaches that any one single "use" of sound could embrace.

2.1 Sonification of data

One of the most extreme cases of usage of sound is the so-called sonification of data, that is, the acoustical representation of some data, which needs not have any direct relationship to acoustic parameters – a sonic form of visualisation. Gerhard Eckel showed an example of *Vector field Sonification and Visualisation* done with the *CyberStage* at GMD[2]. The *CyberStage* is GMD's CAVE-like audio-visual projection system integrating a 4-side visual stereo-display and 8-channel spatial auditory display. In this frame a virtual environment for the exploration of vector field data from airflow simulations of a car air-conditioning system was developed for Daimler Benz. In a real-size model of the interior of a car, the spatially distributed static vector field of air velocity data is visualised by particles injected into the field, which then float along the streamlines, and, what is of central interest for us now, this conventional approach was refined by sonifying the velocity data with low-pass filtered noise portraying wind-like sounds. The velocity data was mapped on the noise amplitude and the cut-off frequency of the filter. The result is certainly very expressive, in the sense that one gets a very clear representation of the air stream velocity through this simple mapping algorithm. (In the eRENA project, similar techniques for data sonification were explored and evaluated in Deliverable D4.3/4.4 in Year 2.)

2.2 Artistic use of data audification

Another approach of sonification of data, but within a completely different approach, was that of Torsten Belschner. In *SonoMorphis*, an interactive installation with genetic graphics and sound, done in collaboration with Bernd Lintermann at ZKM Karlsruhe[3], the goal is rather artistic than scientific, and therefore the sonification doesn't aim to give a clear representation of the underlying data, in this case genome's structures, but rather to provoke the creation of acoustic gestalts through the mapping algorithms which relate the data to the acoustic parameters. In other words, the parameters of the graphics must be interpreted from acoustical viewpoints in such a way that a musical structure emerges from them.

Both sound and projection relate in equal parts to the same underlying abstract structure that they try to make palpable to the user. The sound acoustically represents selected properties of the genomes, i.e. their structure, position, and behaviour in a non-arbitrary way. In this way automatic compositions arise, the results of which are functions of their components and are variable in the details of their contours, complexity, and their behaviours, so that the overlapping of visual levels and sound levels may produce an open structure.

Figure 1: Image from the computer graphics of *SonoMorphis* (Bernd Lintermann & Torsten Belschner)

2.3 Sound data controlling computer graphics

At the opposite side of this usage of sound we find what we might refer to as the "datafication of sound" – a mapping of acoustic signals into data. Being, in any case, a more common practice, since sound has anyway to be represented as some form of data in order to be manipulated in a technological media environment, the workshop confronted us with some special approaches that depart from normal usage. Typically, the mapping of an acoustic signal into data has as a goal the recovering at a later stage of the signal or of some processed result derived from it. Michael Saup, a multimedia artist who currently is professor of media art at the Hochschule fuer Gestaltung (HFG, a neighboured institution of the ZKM), has had as main focus of his work for the last years the control of images through audio-input. He uses some representation of an acoustic signal, which could come from a wide range of different sources, as a graphic signal, be it as modulating or interference signals on an image environment. For example, in his piece *a666*, a musical instrument controls a hard-disk recorder so that the video is not cut any longer but instantly controlled in time, or, in *tilt-it*, the sound-track of a video is used for the direct-creation of 3d-computergraphics.[4] (This approach can be compared to the use of a musical instrument to parameterise computer graphical algorithms in the *Lightwork* performance reported in eRENA in Deliverable D2.2, Year 1.)

2.4 Sound in spatial simulations

To some an extent, the work of Pierre Dutilleux, *AML: Architecture and Music Laboratory*[5], could also be seen from this view-point. In his work, which crystallises as a museum-exhibit giving the visitors an opportunity to experience the relation between sounds and performance spaces, he measures the impulse responses of real rooms, and further processes these data to enter it as parameters of a software simulator – the Spatialisateur6 - to imitate these spaces.

In this approach, in a certain sense akin to that of "Ecological Acoustics"[7], an attempt is made to derive from some data representing a sound signal some features which might directly relate to physical aspects of the source environment where the signal was produced. It is not the signal itself which interests, but properties of it, observable only in some data representation of it, which link perception to relevant features of the physical space where some source signal was reproduced and captured. We shall come back later to this work, but what interests now is that the sound, has to be represented into some kind of data, which is further transformed into parameters of a software device in order to allow the user the feeling that he can manipulate aspects of the surrounding environment where a source signal is reproduced. (In the eRENA project, we have also explored simpler, less computationally intensive approaches to sound spatialisation, see Deliverables D4.5 and D7a.3.)

2.5 Data control of sound and space in CVEs

One of the most special approaches to sound coming up in this workshop, and one of a very special interest to the eRENA project, was that presented by Milena Radenkovic and related to her work with Steve Benford and Chris Greenhalgh at the University of Nottingham[8]. The topic of her lecture was the control of sound in Collaborative Virtual Environments (CVE) by means of the *Spatial Model of Interaction*. This model proposes a flexible and scaleable distributed architecture for real-time voice mixing which is specifically tailored to support very dynamic sessions, rapidly varying requirements and very high levels of participation, and which might include crowd audio synthesis and environment sonification to minimise network traffic in the audio medium, and thus further increasing its scalability. This item of scalability is of extreme importance in this context since one of the newest and most challenging deployments of CVEs is *Inhabited TV*[9]: a new and exciting medium for entertainment and social communication, which combines CVEs with broadcast TV. The potential size of an *Inhabited TV* application (hundreds of thousands of simultaneous participants) challenges scalability of CVEs and in particular scaleable integration of different media.

In her lecture she presented the concepts on which the *Spatial Model of Interaction* is based, and later showed a number of videos, to give a practical feeling of how the different concepts linked to the model could be experienced in some particular situation. The first idea she presented was that of layered participation as a mechanism for describing Inhabited Television and defining associated terminology. Each layer provides different possibilities for navigation, interaction, mutual awareness and communication between participants, and is supported by a distinct combination of interface and transmission technologies. The different layers are composed by professional performers, online inhabitants, viewers who may have interactive devices, and finally the production team – camera operators, directors and technical support who works behind the scenes.

But the most important concepts for the model are those of: *Medium, Aura, Awareness, Focus and Nimbus*.

- *Medium* is the way by means of which information is communicated; be it video, audio, graphics, text or whatever.
- *Aura* represents the scope of interest of the people that participate.
- *Awareness*, quantifies the degree, nature and quality of communication.
- Focus is the field of interest of one character in the environment, while
- *Nimbus,* the complement of *Focus,* is the interest awakened by this character on others.

The concepts of *Focus* and *Nimbus* serve to organise a number of levels in a CVE, such as selectivity of what is shown, following the interests and preferences or urgency or importance of some spot, the kind of realisation, with field over space rules or negotiating agents, or even security considerations. *Awareness*, probably the most important concept, combining those of *Focus* and *Nimbus*, attempts to quantify interest or QOS. Its result may serve directly to control such relevant aspects as audio encoding, volume of sound, video frame-rate and its resolution, the graphical level of detail, and so on.

Related to these concepts, Milena Radenkovic presented the *Third Party Objects*, as a mean to represent context via two mechanisms, namely those of 'effect' and 'activation'. For example, a virtual room, its effect could be to isolate, its activation ruled by the state of being inside or outside; or other more abstract examples where activation could be ruled by common focus or awareness, or membership to a crowd, etc.

3. Sound and Space

The formal space, the imbedding space, the feeling of depth, the perspectives and then the atmosphere, contrasts of lights, the game with colours, transparencies and textures. Plastic art has always had to pay a very special attention to space. Time was an implicit factor in the apprehension of a piece of plastic art. The pieces seemed to be static, and all their traits, illusions, power had to emerge from that static plane. Now, ever since movement entered into the potentiality of their characteristics, the so-called visual arts have had to deal with time in an explicit way; and the space has had to attain a qualitative change; space and time, as we all know by now, are intimately connected. Further, all processes which involve time in an explicit way will interact with, and, conversely, be affected by the surrounding environment where they are present. The outer, “physical” flow of time, the external space and the sound environment. Sound and image, therefore, have to intimately collaborate, since they are intimately linked, in all mixed reality environments. In this section we shall pay special attention to the relationship between space and sound, whilst in the following section we shall concentrate on the relationship between image and sound.

3.1 Aural signatures of real rooms

Pierre Dutilleux presented his work *AML: Architecture and Music Laboratory* (see Footnote 5). As was mentioned above, the final step of this work was a museum installation, currently exhibited at ZKM's Media Museum, where the visitor is given an auditory grasp of an architectural environment. For that purpose the visitor can select a room between twelve different possibilities (a concert hall, a roman theatre, a church, an apartment, etc.), and he can also choose the type of sound that will be projected in the virtual room selected. All the source sounds, amongst which one could find musical excerpts, speech or other type of natural sounds, were recorded dry for the purpose. It should be added that the visitor could also record and later use his voice as a source sound. Further, the visitor who wishes to better understand how a volume can modify a sound, may select also as an aural spatial environment between a number of basic shapes. Amongst these abstract environments, one can find the free field, a cylinder, a cube, a sphere, etc.

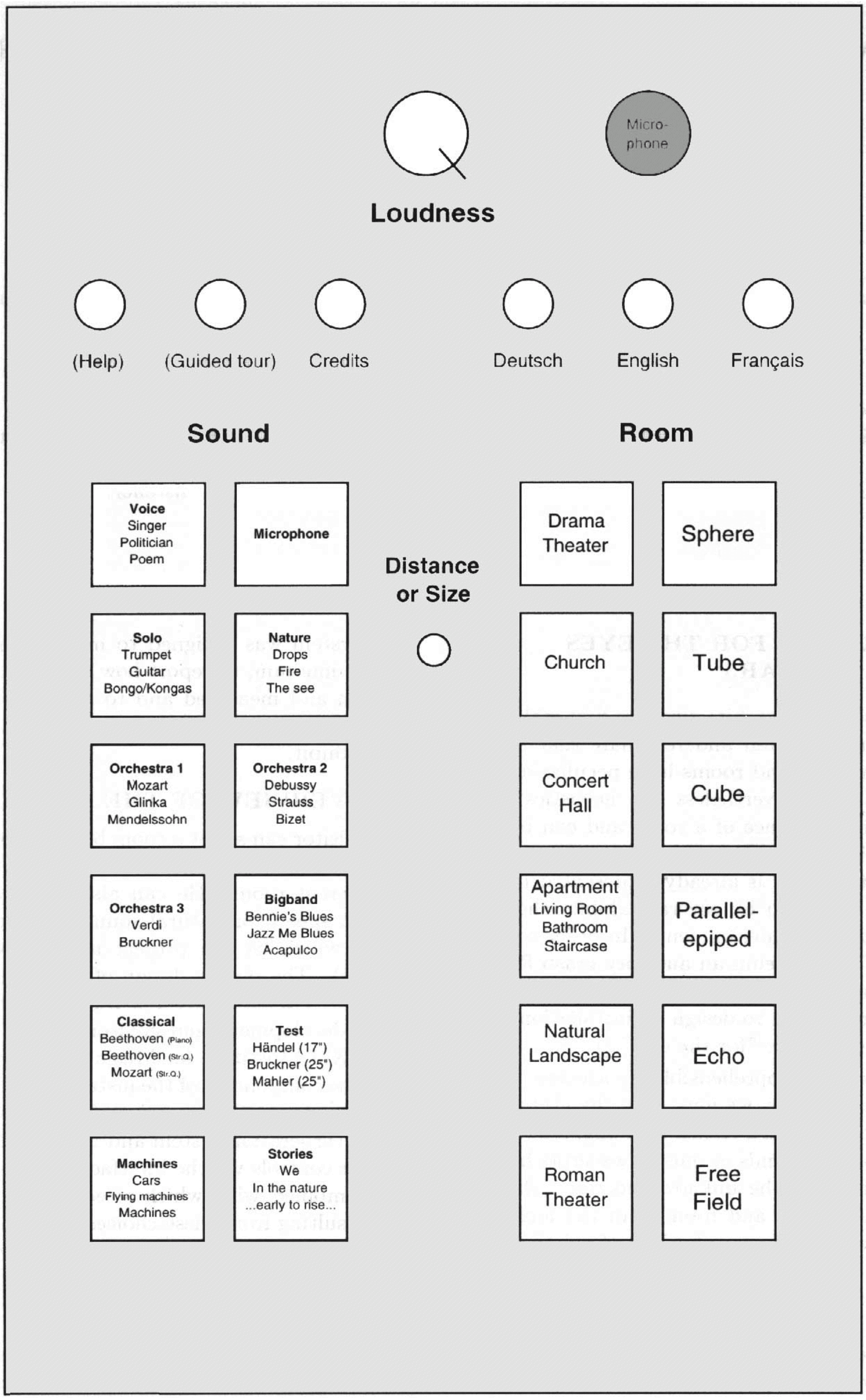

Figure 2: The installation *AML* at the ZKM Media Museum

His lecture gave rise to a good number of interesting questions. To begin with, he included in the lecture some practical examples of how different basic-forms do modify a sound. This he did by means of a small speaker through which some white noise was projected reproduced, and a small set of hard-plastic objects in the form of a sphere or a cylinder, with which he could cover the speaker. This experience could have been called auditory representation of form, or better, spatial-form conveyance through an auditory scene. A fascinating question which provoked quite a bit of discussion. Some remarks might be worth mentioning. Some training seems to be needed in order to be able to relate spectral-modification and form of the object, and, what is more important, the evolutions of the sound spectra happening at the transitional moments of exchange of one object for another, seemed to give a better grasp of the form of the object that was involved. One could infer from this effect that dynamic evolutions in the source sound would help enormously to get a clear picture of the type of object.

A second aspect of his lecture was dealing with what we could call the aural signature of a given room; indeed an important question also. For this purpose he had to measure and evaluate a number of pre-selected sites. Measurements of the room's impulse response were done by means of the MLS (Maximum Length Sequence) method, where a pseudo random noise is played in the room through a loudspeaker and the sound pressure in the room is picked up by a microphone. A computer compares the signal that has been sent to the loudspeaker to the signal measured by the microphone using an algorithm called cross-correlation. In theory, the output of this last operation is an impulse response that characterises the room, but only for the unique set of points where loudspeaker and microphone were placed. The selection of the impulse response that will actually serve as representation of the aural characteristics of it, is absolutely subjective. As he actually states, the process of selection can only derive from listening to the dry sounds on the room in order to get a clear mental representation of how the room sounds, then wander in the room to spot the audible particularities and to select the positions to be measured, and then measure systematically. With these derived impulse responses later, the researcher must try to select which one(s) represents better the given room, by comparing it with the memorised mental image he had at the location.

Notwithstanding the interest of this installation, highly well done, as proofed by the good performance of the exhibit in the museum, it might well be stated that the problem it raises is so interesting that it transcends this single application. In any case this problem seems also to be an extremely complex one. Real rooms have normally various and different geographical accidents, so that no one single representation might embrace its complete aura; and then, binaural audition with the aid of the frequent though imperceptible head-movements we humans perform even in a quiet situation, probably serve to build up an image in place which can be extremely hard not to say impossible to achieve in a virtual environment. However it is most challenging to imagine a virtual sound space, created by using multi-channel technology, where one could create the image of a non-uniform surrounding space.

3.2 Spatial sound by VR technologies

Another lecture that touched in some of its aspects the problems of space and sound was that of Gerhard Eckel. We shall mention here two of these aspects, concretely his presentation of *Camera Musica* and of his sound installation *Stele*.

Camera Musica is a work-in-progress music installation project that was first presented as a speculative paper in a German contemporary music journal in 1994[10]. The first sketch of the installation, a film of which was presented at the workshop, was realised in 1997 and presented at the Eighth International Symposium for Electronic Arts in Chicago[11]. This first sketch was realised with the *CyberStage* system. As indicated above, the *CyberStage* is GMD's CAVE-like audio-visual projection system, the design of the hardware and software components of which resulted from *the Integrated Simulation of Image and Sound* (ISIS) research project.

The concept of *Camera Musica* comes from the idea that new media technology may offer new perspectives upon the utopia of open form in electronic music composition. From this viewpoint, complex and potentially open musical topologies could be explored by a broad public through virtual architecture, since musical spaces become legible and inhabitable for non-musicians in an intuitive way. Architectural organisation of musical space allows the creation of meaningful places in the network of relationships laid out in a composition. The visual scene used in the sketch consists of a building-like structure composed of free-floating walls of various dimensions and colours. Some of the walls reach out into open space thus mediating between inside and outside.

The music in *Camera Musica* is conceived as a family of various, interrelated musical situations composing in their interplay what we may call a musical space. It is this space which is to be made accessible in the installation. The audience is enabled to move from one situation to another within this space and to slowly explore its features through the relations between individual situations. Each situation is characterised by certain possibilities of choosing the musical material and arranging it, thereby determining the particularity of the situation - its mood, atmosphere, form and air. Depending on the position and orientation of the user, these choices are taken by a program whose development is part of the composition. The spatialisation of the sounds in *Camera Musica* is abstract in the sense that it does not refer to an acoustical but an artistic model. The direction and distance of the sound sources used in *Camera Musica* are controlled by compositional parameters.

A strong interest in the interaction between visual art and music, led G. Eckel to the idea of building a sculpture with a visual and a sonic appearance. The *Stele* was produced in spring 1998 at GMD and ZKM and was presented in public for the first time in June 1998 at GMD Birlinghoven Castle. The main goal of *Stele*, as well as that of the *Viola Spezzata*, a later project intimately related to this one, was "to assemble sound objects into a sound sculpture generating a strong presence in space". For this purpose he designed a kind of acoustical antenna built from eight Genelec 1030 loudspeakers positioned one on top of the other such as to form a 2,5m high column. The sculpture's static visual aspect and its dynamic acoustic appearance form thereby a powerful contrast, which mutually intensifies the two components.


Figure 3 and 4: The *Stele* at the Kubus with Gerhard Eckel during his workshop demonstration

The acoustic radiation pattern of this antenna can be controlled dynamically by means of a DSP program realised with Max/FTS running on an SGI computer with 8 audio ports, each of which drives a speaker of the sculpture. The originality of this approach lies in the independent control of the 8 loudspeakers using varying time delays in the ms range. This permits to independently direct or distort several first wave fronts and thus simulate movements (e.g. varying inclination) of the sculpture. The resulting complex acoustic excitation of the room is characterised by very ambiguous localisation cues and moving phantom sources. Another aspect of the original at the basis of the *Stele* project was to create an experience of the intimacy and complexity of a certain kind of instrumental gestures created with string instruments. For that purpose, he conceived and recorded in close collaboration with Vincent Royer, a professional alto player, the whole sound material. The processing of this material is rather conventional and comprises transposition, filtering and some stochastic processes to build clouds of sounds and layered structures. The main compositional work consisted in mapping the movements inherent to the recorded and arranged sound material to the sonic movements of the *Stele*, which as he states, becames a process that proved to be highly empirical and could only be done in situ, i.e. with the *Stele* installed in a proper room. This process has been carried out at the big recording studio of ZKM, the so-called Kubus.

4. Sound and Image

The relationship between sound and image is certainly of a complex nature. While our everyday experience consists of many natural correlations (the sight of a flowing stream, the sound it makes), it is extremely hard to state general relationships in a satisfactory way – and certainly not in a way that can be formalised and readily embodied in interactive computational systems. From a psychological point of view, it can be argued that the way visual and auditory perceptual mechanisms analyse and process their corresponding stimuli in order to create an internal image of the external scene, are very different. Underlying structures have to be very carefully mapped into relevant parameters of vision or of sound in order to be able to potentially apprehend the coherence of the structure in each one of them. The relationship between the senses is, certainly, one of the most fascinating questions with which art has confronted itself in this century. Forms of art with a tendency to abstraction confront these questions most severely as natural correlation between the senses cannot always be capitalised upon.

Contributors to the workshop presented a number of strategies for addressing these underlying difficulties. In each case, relationships between sound and image were being explored where there was a requisite flexibility in the coupling between the two. Contributors seemed to have set themselves the task of avoiding the dual problems of (i) an excessively close correlation of all visual and sonic details (a situation which choreographers sometimes refer to as ‘Mickey Mousing’ after the manner in which visually available gestures are directly reflected in music in cartoons, something typically to be avoided in setting dance to music) and (ii) a relationship so obscure that it is un-obvious to performers and audience, so that any possibility for interaction is forfeited. Each of the contributions to the workshop addressing image-sound associations were trying to steer a path between these two perils.

4.1 Sound controlling image

Michael Saup addressed the issue of image-sound relations in the context of his own multi-media pieces. As we have suggested, his approach is, in general, rather of a transgressive nature concerning this problem. He uses sound, some representation of sound-signals, as image signals, or as signals to control devices. Interesting, but very personal and extreme, in order that his approach could have triggered a more “general” question concerning the relationship between sound and vision.

4.2 Movements controlling images and sounds (VR environment)

Another interesting case, was that of *Camera Musica*, the interactive music installation of Gerhard Eckel. In this piece, an attempt is made to relate architectural structures with musical structures. The whole environment is rather abstract. The visual scene shows a static object, a kind of labyrinth, or an apartment with walls and other geometrical objects, which can be explored by the user, moving around, approaching or putting distance. The movements of the user, what he sees and from where, have a causal relationship to the musical structures that can be heard. Indeed, he has posed himself a rather difficult and interesting problem. What we saw at the workshop was only a video from a sketch of this work in progress, with an attractive atmosphere, undoubtedly, but from which it is difficult, not to say impossible, to know how this relationship between architectural and musical structures is working.

4.3 Layers of links between sound and image (*SonoMorphis*)

The lecture/demonstration which touched more directly the relationship of image and sound was that of Torsten Belschner and Bernd Lintermann, co-authors of *SonoMorphis*. We have already described this interactive installation, but it might be interesting to comment now on their decision concerning this relationship. To begin with, it is important to state clearly that there is an underlying data structure to which both sound and image are related. The result is that the relationship between the sound structures and the visual structures is rather loose; both domains are interpretations of the underlying structure, and therefore, there is not a direct, but an indirect link between them. This is a very interesting decision from their part. However, if there had being no other connection between the domains of sound and image, the relationship would have risked being too loose and endanger the sense of coherence of the whole environment. Of course, it must be said that there is no simple rule of thumb, since the way the underlying data is read by each domain can be so different, that finally the result could have been anything from complete separation to absolute co-ordination. They consciously chose this middle way, in order to allow the user to experience and search for herself all possible links. It is in this half way between total co-ordination and breaking separation, where further hints of coherence are of extreme importance. On the one hand, they created a direct link between sound and visual objects through the means of spatialisation. There is a virtual camera that sees in perspective the evolution of the visual structure, and this spatial information is mapped into a simple spatial audio rendition of the musical structures. This is a direct link of a “superficial” feature, by which the mind is pushed to connect the two domains without imposing very strongly. Another element they used as a “secondary” connection mechanism between sound and image is of a more sophisticated nature. Since the visual objects which are created and evolve in the screen have an organic character, the sound qualities used have been produced by physical modelling synthesis methods, which, whilst remaining abstract enough so as not to cause a too clear-cut identification, point in the direction of bodies, of physical objects, of real matter. Thirdly, moments of clear change in the structure of the image (triggered, for example, by the user causing a mutation to take place) are co-ordinated with simultaneously noticeable changes in sound quality. These three more direct links create a work where sound and image are loosely but efficiently and noticeably connected.

5. Sound and Interaction

Interaction, interactive media, interactive art-works, seem to be the key-words for the future. However, at the time being, these only apply to a limited, though important, set of multimedia environments. It would have been natural, therefore, to expect only a relatively small presence of it at this particular workshop, in contrast to the more universal problems of sound in itself, of the relationship between sound and image, or even of the requirements to obtain an all media encompassing immersive space. It was, on the contrary, addressed directly by all lecturers, again in a most rich and varied fashion, and generously touched by all participants, in one or other aspect related to it, in the general discussions.

5.1 Examples of interaction models

A significant aspect to be underlined from the very beginning is that, even in the cases were the role of interaction might be of a secondary importance to the final object of research or creation, be it installation, broadcast or any other, its weight of influence in the design process seems to be paramount. This is not surprising, since the technology of interaction addresses the avant-garde of what the different media may offer in terms of technical requirements. The technical constraints it poses, the cunning in solving problems it demands, must, only naturally, sensibly colour the process of design and research; independently, one could almost add, of the degree of interaction or of the sophistication of the handles that the user may be offered. This, for instance, could be the case of Pierre Dutilleux's museum installation *AML*, where, as I already mentioned, the deep level of research on spatial acoustics and its further simulation, which apparently constitutes the real core of interest, was highly affected by all the big number of constraints posed by the fact that the final goal of his work was to have the form of an interactive installation. In his paper (see footnote 5) he describes in all detail the methodological steps he had to take care of at the time of planning the research.

Our report starts with the lightest degree of interaction, one that we could almost depict as virtual interaction. In Gerhard Eckel's *Stele*, the sound sculpture mentioned above, we could distinguish two levels of interaction. First, of extraordinary importance for the piece, between the object itself and its surroundings: the room where it is placed and its exact location within the room. Asymmetries and geographic accidents serving to enhance or hinder the mutual effect between the object and the room. The second level, and more precisely that of real interaction, human interaction, happens through the displacement of the audience around the room, approaching and distancing from the sound-sculpture, surrounding it, looking for the spots where the better interaction with the surrounding room could be perceived at a given moment. It must be added that the installation does really invite the visitor to roam around, to stop here or there and then move again. In his lecture, G. Eckel was further showing a video of an amusing interactive demo, *The Spheres*, particularly interesting in this context, because of the relevance of the relationship between sound and image. This interactive demo runs on the *CyberStage*, introducing the user, equipped with glasses a joy-stick and a pointer, in a darkened and silent environment, where small coloured spheres, sort of balloons, can be pointed at, that is selected, and inflated by means of the joy-stick, which would cause both a growing process and an accompanying sound of coloured noise, which acted as a symbol, not really a simulation, of the noise of air insuflated into a plastic sphere. The combination of these two elements, increase of size plus air-being-blown-like noise, managed to give a vivid sensation of real interaction by means of very simple paradigms, arriving to the point where the balloon if further inflated would blow out with a cymbal-like explosion.

In Torsten Belschner's and Bernd Lintermann's *SonoMorphis*, the user is offered a small console of faders and switches, with which on the one hand he can modify the camera perspective from which the object is seen, and therefore the spatial aural processing of the sounds, and second to interact in the structural configuration and evolution of the object itself, and with it that of the sounds produced. With this interface the amount of possibilities to "play" around is enormous and challenging. One limitation can be noted, and that is that there is no real interaction to the sound structures themselves other than indirectly. It could have been interesting perhaps to have had the possibility to modify or influence the interfacing mechanisms between visual and sound structures, which of course would have not been simple, since as we know, both "surface" structures are representations of an underlying abstract structure, which is the only path of relationship between them.

In the lecture of Michael Saup, interaction was touched in very many different ways. Most noteworthy in this respect, because it somehow summarises a certain number of his approaches, but also because of its images full of expressive power and interest, was his installation *plasma/architexture* of 1994, of which a video was shown.

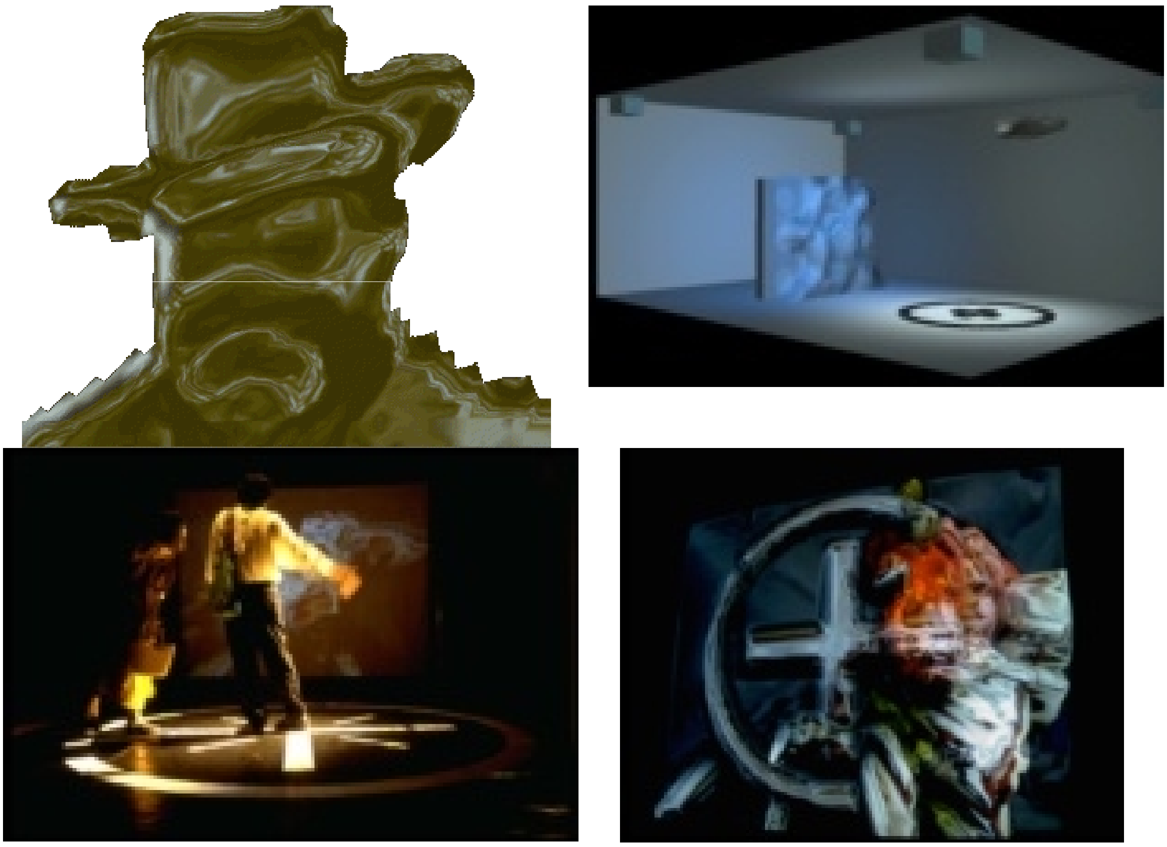

Figure 5: Images from the installation *plasma/architexture* by Michael Saup

It might be interesting to quote in extenso his own words about the installation: "As the viewer steps into the room, the screen shows the image of the nightvision camera: a wideangle-shot of the place. Like in a mirror, that illuminates hidden structures, the viewer can see himself crossing the room. By stepping into the middle of the circular pattern, he activates the two b/w-cameras, and two shots of each person are made, from two different angles with two different focal lengths. These images are transformed into 3-d models and interpolated in real-time by the especially designed software *dwarphmorph*. Liquid, organic images are created. The nightvision image is put as a texture over these objects and - through further software-based manipulation - is turned into plasma. The software-mirror starts a life of its own: the 'reflection' expands, bulges and transforms continuously. At the same time, the pneumatic projection screen will transform his physical flatness into 3d-space to support the 'plasmatic' effect.

“Confronted with his fragmented image, the spectator wants to recognise himself in the mirror, and starts to act. As he moves, the plasma is starting to react. Not only visual patterns, but also interpolated speech is being created: a microphone collects all language and noises in the room, storing them in small units in the computer's memory where they are being 'soundmorphed' and interpolated into an artificial language that consists not of concrete messages but has a strong emotional value, at least as long as the viewer is moving in a hectic, excited fashion. Should he change his behaviour to slower, smoother movements, the plasma, too, changes. It starts giving off pieces of comprehensible language that were stored earlier, pieces of information given by other viewers, whilst storing, at the same time, the sound-information given by the present viewer. What is more, the artificial being is no longer reduced to a passive imitation of the viewer, it starts moving after its own laws. As the viewer’s movement becomes more co-ordinated and smoother, the reaction of plasma is no longer mirrorlike but livelike, it starts playing the active part, ‘talking’ to the viewer using its ‘brain’ to construct 2way-communicaton.”[12]

5.2 Algorithmically layered interaction

The last lecture to be mentioned concerning the problem of interaction, that of John Bowers and Sten-Olov Hellström, was the most focused contributions to this topic in the workshop. Their lecture started off with a theoretical discussion, making some reflections on the state of the art, questioning the needs of different multimedia applications, defining and discussing a number of concepts which could be useful for interaction purposes, and ended with some practical examples of application of the previously defined concepts, as used by them in their activity as performers of real-time improvised music[13].

As they stated, conventional desk-top interface design of human-computer interaction is dominated by so-called *Direct Manipulation* (DM). In this approach, information is represented by objects which allow the user some way of direct manipulation on them, while giving the phenomenological impression that what she does is exactly what it produces. Is this kind of approach still appropriate for participation in large scale CVEs? They propose an alternative, not without risks, but which may add extra power and potentiality of usage to interfaces. This is *algorithmically mediated interaction*, wherein an algorithmic layer captures the user's gestures to be used for whatever purpose. The important thing is that the user's gestures are transformed into parameter values of some algorithm, therefore interacting on the algorithm itself. There is a risk, because there is a gap, both technically and phenomenologically, between the gesture at the interface and the effect. The challenge therefore being the design of an interface which still involves the body, the physical gesture in a satisfactory way, while having a clear recognition of an algorithmic layer.

To approach these problems they defined a number of concepts, namely, those of *loose coupling*, *expressive latitude* and *interactional affordance*. It should be noted that their interest on the type of interactive relationship to a mixed reality environment shifts from some kind of ideology of cyborg fusion with the technical, to some kind of inhabitation in the environment, wherein the user manipulates devices which she shares the space with. In this sense they mentioned the possibility of using devices which do not only allow this interaction between the user and the device, but also that they might enable the interaction to take place in such a way that its character can be perceived by third parties, be it audience or co-performers. The importance then is not only the possible engagement of the user, but that her actions might express or inform others. At this point the concepts of loose coupling and expressive latitude gain all their power: “By cutting down the dimensionality of the degrees of freedom in a control surface which might have some representation in the environment”, therefore, by allowing a 'loose coupling' between the body and some device, “there opens a space for other gestures and other activity”, which is not directly influential on the interaction itself but has a communicative and 'expressive latitude' in the context. As they stated there is an increase in the feeling of control when there is not a close coupling between technology and body.
Further details of this approach, and of the specific interface technologies developed to demonstrate it, is reserved for Deliverable D6.4.

6. Conclusions

We have to be aware of the underlying ideas in order to assess the achievements of the workshop. The focus was mainly on the exchange between experts, and more on “broadening the horizon” of the involved eRENA collaborators and other participants than on precise practical results. Nevertheless, there were fruitful discussions on the relationships between different works. For example, the work of Pierre Dutilleux, concerning the awareness and the construction of characteristic spaces, was related to the design of Inhabited TV, and was discussed in relation to the extent to which one could apply this access to the audio structure of the mixed reality environment.

It was greatly appreciated that most of the lectures were accompanied by demonstrations, often with real installation set-ups. The workshop participants took advantage of these unique opportunities and asked many questions concerning the particular works, pieces or set-ups. Moreover, the group of participants was a mixed one in terms of qualification and areas of professional expertise and interest, which meant that many questions that were obvious to the one group of participants were not at all clear to another group. These kind of “translation actions” between different worlds are time consuming activities. In any case, we have arrived at the crux of all interdisciplinary work: mixed reality environments can only be developed with a certain amount of knowledge of all different fields shared by the persons co-operating.

Figure 6: Workshop discussions

In this respect, the workshop was a successful undertaking. This can be seen in various comments by participants, proved by different participants' statements, who said they gained insight into such things as:

- the specific acoustical and technical problems in mixed or virtual reality environments (insofar as they differ from the typical image problems)
- creative ideas and experiments from electro-acoustic music (fruitful in the area of amore pragmatic use of sound)
- scope of sound and acoustic issues.

The workshop, though broad in its remit, has also had some valuable specific consequences in shaping the research done in eRENA in Year 3. We explicitly note three clear lines of influence:

- the work on algorithmically mediated interaction (5.2) was targeted for this workshop, in part evaluated at it, and further refined in response to participant's comments (see Deliverable D6.4)
- the sound management work for scalable CVEs (2.5) similarly benefited from presentation and appraisal by sound professionals of varying backgrounds (Deliverable D7a.3 describes the use of these techniques in *Avatar Farm*)
- presentation of a variety of techniques for the spatialisation of sound (Section 3 here) gave direct motivation to the virtual sound mixing application implemented on the Round Table (see Deliverable D4.5).

In this way, the workshop can be seen to have had a clearly formative influence on the work of all eRENA partners in attendance as well as on specific items of work from Workpackages 4, 6 and 7.

References

[1] See the scientific references of William Gaver on http://www.crd.rca.ac.uk/~bill/ref.htm.
[2] Gerhard Eckel: Applications of the Cyberstage Spatial Sound Server. Proceedings of the 16th AES International Conference on Spatial Sound Reproduction 1999; GMD Sankt Augustin
[3] Bernd Lintermann & Torsten Belschner : *SonoMorphis* – Interactive installation with genetic graphics and sound. *SonoMorphis* has been shown at the exhibition surrogate[1], Nov 1st - Dec 6th 1998 at the ZKM Institute for Visual Media. The image was projected onto the front wall of a room of 6m width and 7m depth. The projection size was 5 x 3m. The interface consists out of a mixing board plus chair which were installed 4,50m in front of the screen. The image was projected in stereo using Chrystal Eyes Shutter Glasses. The moving sound sources were projected quadrophonically with two speakers in front and two in the rear. *SonoMorphis* has also been shown during the 11th Stuttgarter Filmwinter, Jan 14th - 17th 1999, in the CAVE of the IAO, Fraunhofer Gesellschaft in Stuttgart. The user interface was located in front of the CAVE. http://i31www.ira.uka.de/~linter/SonoMorphis
[4] www.particles.de
[5] Pierre Dutilleux and Christian Mueller-Tomfelde: AML: Architecture and Music Laboratory. A Museum Installation. AES 16th International Conference on Spatial Sound Reproduction 1999)
[6] Espaces Nouveaux/IRCAM. Spatialisateur, the Spat Processor and its Library of Max Objects. Reference Manual. IRCAM 1995 & Jot J.M., Warusfel O. 1995. A Real-Time Spatial Sound Processor for Music and Virtual Reality Applications. In: Proc. of the 1995 International Computer Music Conference. International Computer Music Association, San Francisco, 1995.
[7] see for example: W. Gaver: How do we Hear in the World? Exploration in Ecological Acoustics. Ecological Psychology, 5(4): 285-313 1993
[8] M. Radenkovic, C.M. Greenhalgh, S.D. Benford: "A Scaleable Audio Service for Collaborative Virtual Environments", UKVRSIG'99, Salford, September, 1999, Springer Press.
[9] Milena Radenkovic, Chris Greenhalgh, Steve Benford: Inhabited TV: Multimedia Broadcasting from Large Scale Collaborative Virtual World, Telsiks 99, Vol1, p68-72, IEEE Catalog Number: 99EX365
[10] Gerhard Eckel, 1994. Camera Musica. In: Gisela Nauck (ed) Interaktive Musik, Positionen 21. Berlin, 1994. pp. 25-28.
[11] Gerhard Eckel, 1997. Exploring Musical Space by Means of Virtual Architecture. Proc. of the 8th International Symposium on Electronic Art. School of the Art Institute of Chicago, Chicago, 1997.
[12] www.particles.de (December 1999)
[13] see eRENA Deliverable D6.4.