# THE SPATIAL SOUND LAB AT FRAUNHOFER IMK.VE

*Joachim Goßmann*
*Dr. Florian Dombois*

Fraunhofer Institute Media Communication
Schloss Birlinghoven
D-53754 St. Augustin, Germany
[gossmann, dombois]@imk.fraunhofer.de

## ABSTRACT

Sound in VR, generally a sidekick to the visual presentation, is a central issue in the ongoing VR opera production within the project “Digitales Beethoven-Haus” at Fraunhofer IMK. To be able to answer the technological and artistic questions brought up by such an endeavour, a lab room was established allowing for high quality immersive VR for both sound and image at the same time. Strategies were developed to successfully stage a piece of classical music within a virtual environment, allowing the audience to interact with the virtual world in an interesting way while being faithful to the musical performance that forms the basis of what is audible.

## 1. INTRODUCTION

In a virtual environment 3D computer graphics are used to provide an illusion of space, that surrounds an user, and of interaction with objects, that an user becomes to see. A new form of experience is intended: instead of having the results of programming in front on a screen, they form a whole little world where one can dive into, navigate through and try to become involved. In the context of virtual environment research the catchword "immersion" plays therefore an important role and two major strategies are followed to improve the persuasive power of that 3D projection: on the one hand rendering techniques are developed that make virtual objects look more real; on the other hand displays become multimodal: auditory, haptic, even olfactoric channels are added to make the user believe more in his experience of that virtual reality. For instance seeing a flower could trigger some fragrances to be blown into the room or a falling glas could be assisted by a characteristic noise.

Nevertheless the multimodal virtual environment is nearly always dominated by the visual. This has several reasons -our cultural preference for the eye, the historic development of hardware etc. - and becomes also very obvious in the utilization of sound in VR. In most virtual reality demos sound is only used either for illustrating visual effects or for creating a certain mood by background music. One may think of film productions: first comes the camera, later the sound artists for filling the gaps. The technical conditions of a video screen are also in most cases not too appropriate for audio: rear projections for instance enforce the speakers to be set up at the margins.

But every cloud has a silver lining. Some examples can be found where VR displays were used for acoustic research and audio-driven applications. We want to name here only two groups: Gerhard Eckel, who presented at ICAD '94 his approach for reproducing 3D sound in VR [2] and the group at the University of Helsinki [4], who joined ICAD for many years. Like our predecessors we want to bring here the audio side of virtual environments into focus and develop our application from that acoustic point of view. The project, that gives us the chance for that audio centered research, is called "Digitales Beethoven-Haus" and will be briefly characterized in the following.

### 1.1. Digitales Beethoven-Haus

The house, at which of Ludwig van Beethoven was born, still exists in Bonn, Germany, and is today a museum and archive of international reputation. The *Beethoven-Museum* and *Fraunhofer Institute Media Communication* brought together a major project "Digitales Beethoven-Haus" to update the memorial with digital technology which is funded by German Ministry. More than 26.000 scans of autographs and books have been already taken, a sophisticated web-site is building up [1] and - this is our part - a new permanent VR theatre will be installed within the museum, where virtual environments can be shown. This new set up will be used to show artistic music visualizations of pieces of Beethoven's œuvre.

Our first production is a mise en scène of Beethoven's only opera 'Fidelio', which means in our case 19 minutes of music from the second act (II 1-5). We use a multitrack recording from 1978 with Bernstein, Sotin, Kollo, Janowitz, Jungwirth and *Wiener Philharmoniker* from the archive of *Deutsche Grammophon*. Our visualization is abstract and we have our idols e.g. in Oskar Fischinger, Len Lye or the Whitney brothers. Unfortunately not much of their know-how in music visualization can be transposed from film to VR since it is not only a change from 2D towards 3D but also that interaction makes any prediction impossible. We have to face new artistic and technological challenges for the sound: How can we create a 3D interactive soundscape from classical music? What aspects in the sound can be changed, what rendering processes applied, without loosing the artistic and technological quality of the source material? What range of freedom is reasonable from an aesthetic point of view? And last not least, how can a set up be constructed that provides perfect audio reproduction but gives space for a coexisting 3D computer graphic screen at the same time? These questions have the same fundamental quality like the one we already mentioned: How can we provide an illusion of acoustic space, that surrounds a user, and of interaction with sonic objects, that a user gets to hear? The production will be finished in fall 2003.

## 2. DISPLAY

The concept of the planned virtual opera production features a chairless auditorium with a large projection screen in frontal position. Cohering with the projected virtual objects, an immersive three-dimensional soundscape covers the entire room. The projection serves as a window to the virtual world that has a consistent geometric relation to the physical auditorium, in which four interaction devices allow audience members to interact with the ongoing presentation.

### 2.1. The spatial sound-lab

In order to investigate the questions arising from this combination of virtual reality with high quality sound rendering and reproduction, a lab room was installed at Fraunhofer IMK in St.Augustin that allows to prototype installations before they are installed in the actual museum.

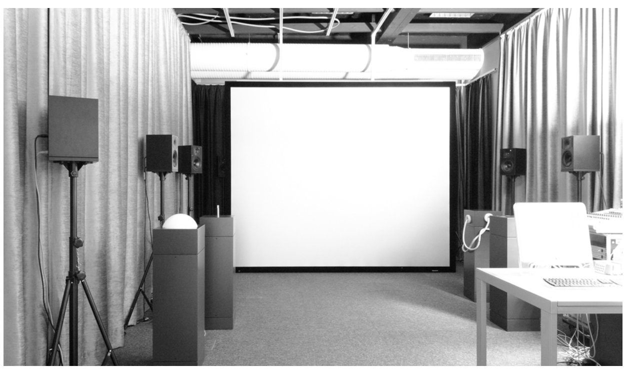

Figure 1: *The Spatial Sound Lab at Fraunhofer IMK*

### 2.2. Display technology: Video

One of the first questions to be solved was how to place a virtual sound source for the entire audience exactly in the place where a virtual object is visualized. Because of the screen dimensions, speakers placed at the margins of the screen could not reliably create the desired phantom sources. The only convincing solution is to have the sound physically emanate from the direction where the picture is seen. The screen we needed for our application has to be sound transparent in order to be able to place loudspeakers behind it, and at the same time silver coated to allow the use of polarized light for projection. The perforation needs to be subtle as the minimum distance of the audience members is only three meters in this scenario. Using perforated screens in conjunction with digital projectors also yields the danger of visible moiré patterns. This needs to be considered in the choice of projectors. In our case, D-ILA projectors will be used. The visual rendering is done on a HP Linux workstation and a FireGL4 card, and is based on our own virtual environment software infrastructure, Avango [3].

### 2.3. Display technology: Audio

Our choice of audio display technology was based on the evaluation of actual sensory, sounding results. In this respect, we took the perspective of an artist or audience, using our perception as tools of judgment to find the appropriate audio display type for the task at hand:
Three approaches of spatialising sound in a public installation were investigated. They can be roughly differentiated into headphone-based systems with binaural rendering, multi-loudspeaker systems using derivatives of VBAP [4], and arrays of loudspeakers using wavefield synthesis. For the planned installation, headphone approaches seemed not suitable or advantageous because the isolation inherent to headphone systems seemed inappropriate for the group experience we were aiming at. (We are investigating the use of tracked headphones with binaural rendering in other projects, such as LISTEN [8]).
Wavefield synthesis [5] is a powerful approach to spatialising audio. It solves many problems inherent to transmission with discrete loudspeakers. However, in our tests it had three shortcomings that kept us from making it our method of choice: At the time we needed to build the display, dynamically moving sources, vital to us, were unavailable, and there were considerable artifacts in the frequency response due to spatial aliassing, when the listener's head moves. The transmission of high frequency impulses is somewhat clouded due to the extensive convolution, so that the system still lacked the quality of sound transmission we were aiming at. We will however continue to investigate wavefield-synthesis within the VR context as it seems an interesting approach for many applications.
We decided to solve our problem using a specialized installation of loudspeakers that will be described below.
The sound-rendering is done on the Max/MSP platform on a Power Macintosh G4.

### 2.4. A first application for the VE spatial sound lab: Staging a classical music recording within an interactive virtual environment.

The reproduction of classical music has a long standing tradition in radio and CD production.
Traditional western music, like the music of Beethoven, was to be experienced with the performers and the audience being present in the same space at the same time: The audience can see the actions of the performers. The sounds, complex fields of oscillation in air pressure, are produced by instruments or singers and are propagated through the air to the ears of the listener. Along the way, they are reflected and absorbed by walls and objects within the room. The room is thus an integral part of the experience of the music. Classical music is, with few exceptions, composed with the acoustic responses of certain performance spaces in mind [6][7].
The room that was involved in the production of the multi-track tape forming the basis of our endeavor, making itself heard on all channels, is the auditorium of the Vienna State opera.
The tape features two main tracks for an overall image of the orchestra, two tracks with a "close-up" image of the strings, and four tracks devoted to the four singers appearing in the excerpt used. When the singers have no part, the free tracks are used for spot-microphones for the woodwinds and the timpani.

We are very grateful to the Deutsche Grammophon that they put this piece of recording art into our hands.
Originally, these elements were meant to be experienced after being carefully blended together by a balance engineer on a stereo-triangle loudspeaker display into a two-channel mixdown published on a record or CD. Our task is now to re-interpret this original tape as a spatial sound environment that allows the visitors of the Beethoven house, within responsible limits, to interact with the composition of the sound, namely its spatial aspects, while making the acoustic and artistic reality captured on the source tape audible in the best possible way.

### 2.4.1. *The multichannel tape and the models of VR-sound*

Sound for VR is generally modeled as sound-sources (point sources), emanating "direct" sound devoid of room information, and a room acoustic model representing the space the sound-source is placed in. Thus, the spatial imaging in terms of localization, distance and room-information of the sound is completely in the hands of the rendering, and the actual sounding result is the product of a sound model oriented towards emulating a "physical reality".
However, certain aspects of recording an orchestra or large sounding body in general contradicts this way of modeling the sound: For the pickup of a good orchestra sound, main microphones are needed that are carefully positioned to pick up a beautifully sounding "wet" blend of all instrument groups contributing. If done right, this yields a three-dimensional image of the orchestra in itself if played back through a defined speaker setting. Moreover, the room in which the performers were at the time of the recording is heard on all channels, so that a virtual room model can only be added and not used as a general replacement. To our favor, the bleed of the singers into the orchestra microphones and tracks is quite low (<-14 dB), with a natural time-delay of around 25 ms, allowing us a to get a good grip on the singer localisation while keeping the orchestra untouched.

### 2.4.2. *Spatial setting of the sound elements*

Since the tape does not contain enough information to make the orchestra sound "navigable", the orchestra will be left in a fixed position most of the time, emphasizing the consistency between the physical space and the frontal window to the virtual world. This forms the first layer of the sound.
The second layer, the ensemble of singers, is represented by moveable virtual sound-sources, dynamically rendered using VBAP and early reflection rendering.
The two layers are made coherent by the room signal produced by the singers that bleeds into the orchestra track, and by a third layer: an added virtual room model enveloping the audience. Each of these layers has its own requirements for rendering and sound projection, and is treated specifically in order to allow a maximum degree of control while limiting the artifacts of rendering to a minimum.

### 2.4.3. *Loudspeaker setup*

In order to stage the given multi-channel tape, the approach taken is a combination between a VBAP installation and an orchestration of discrete loudspeakers set into the auditorium space:
VBAP, while able to create stable virtual sound sources, also faces the problem of phase cancellation when the signal is present on two or more loudspeakers during panning. Therefore, it is only applied to sound elements that need to be moved dynamically. Spatially static elements of the sound such as the orchestra, the strings and the woodwinds are played back through groups of discrete speakers that also allow possible interchannel signals to be experienced.
The loudspeakers used for VBAP are located in a ring along the walls of the audience at an angular resolution that varies from 7 degrees (front) to 30 degrees in the back. This ring is used to spatialize the ensemble of singers. A set of discrete loudspeakers is used to play back the static elements of the orchestra. A selection of the speakers from the VBAP installation is used to envelop the audience with an expanded replicated version of the room heard on the original recording.

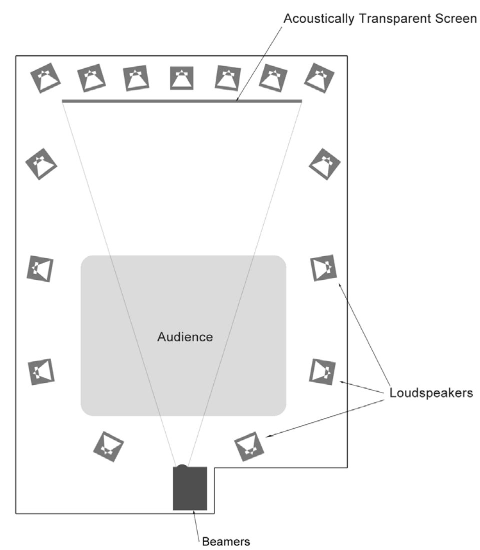


Figure 2: *The loudspeaker setup for the Fidelio production*

The loudspeakers behind the screen are equalized in order to compensate the loss in high frequencies caused by the perforated screen. Delays are applied to the loudspeakers in
to compensate the different latencies resulting from the varying distances from the audience, which is important for both the projection of the early reflections but also the VBAP-layer.

### 2.4.4. *Physical modeling versus controlled authoring of sound processes*

One of the central questions in our production is the generation of an adaptive soundscape that can be randomly influenced by the audience, while assuring all that can possibly happen sounds "good". Here, we get into another conflict of the physical modeling approach used in traditional VR sound: The sounding result was sometimes counterproductive in terms of overall balance, the blending of orchestra and soloists, and proved to be cumbersome to control in order to deliver a rich sound in all cases.

The strategy we use to regain control is to discretely map out the critical parameters that determine the processes for the rendering of the sound-sources onto a topological 2D map of the virtual space: A 2D GUI control realized in Max/MSP allows for a detailed definition not only of the position of the source within the virtual audio space, but of all aspects of the sound processing requiring detailed control. Many of these parameters change along a radial pattern around an imagined center of the audience, but it is also possible to give special attributes to locations within the virtual space. With this control, we can determine the sound rendering in high detail and avoid "dead corners" in the space the sound-source moves through. An effective staging for the virtual sound-sources can be authored, and dramatic relationships of different source-locations within the virtual audio space are possible.

## 3. CONCLUSIONS

The reproduction of sound is traditionally a place where technical and theoretical models meet the sensual and artistic application. Sometimes, the theories and models developed for the sound rendering can "blind" the ear for the actual sounding result that is achieved. Music reproductions have always claimed to try to be as close to the "original" as possible ("His master's voice"), however, it is more than doubtful if that is what the audience, the producer or the artist would indeed wish for. Representations within a medium are always media-specific, and in the best case, the medium becomes intuitive. In that case, the people creating the multichannel tape, the interactive soundscape, the sound-rendering and its spatial parameter curves become unheard, creating an inseparable multi-modal experience together with the visualisation and the interaction. If that is achieved, we have reached our goal.

## 4. ACKNOWLEDGEMENTS

We would like to thank Aeldrik Pander for the audio programming, Johanna Dombois for the artistic concept, Jeremy Eccles for the visual support and our group leader Martin Goebel. Special thanks go to Deutsche Grammophon for the multichannel recording.

## 5. REFERENCES

[1] cf. http://www.beethoven-haus-bonn.de

[2] Eckel, G., "A Spatial Auditory Display for the CyberStage", Proc. 5th International Conference on Auditory Display, electronic Workshops in Computing (eWiC) series, British Computer Society and Springer, Glasgow, 1998.

[3] Tramberend, H. (1999), Avango: A Distributed Virtual Reality Framework. In: Proceedings of the IEEE Virtual Reality '99 Conference, Houston,Texas, USA, 1999.

[4] Ville Pulkki: "Virtual Sound Source Positioning Using Vector Base Amplitude Panning", Journal of the Audio Engineering Society, vol 45, no 6 1997 June

[5] Diemer de Vries, "Variable acoustics by wave field synthesis: new developments and applications", proc. 5th meeting of the FWO Research Society on foundations of music research: Authentic music performance and variable acoustics, Gent, 1998.

[6] Weinzierl, Stefan: Beethovens Konzerträume. Raumakustik und symphonische Aufführungspraxis an der Schwelle zum modernen Konzertwesen, Bochinsky, Frankfurt 2002

[7] Meyer, Jürgen: Akustik und musikalische Aufführungspraxis, Leitfaden für Akustiker, Tonmeister, Musiker, Instrumentenbauer und Architekten, 3.Auflage, Bochinsky, Frankfurt 1995

[8] http://listen.imk.fraunhofer.de