# AUDITORY VIRTUAL ENVIRONMENTS

Pedro Novo
Institut für Kommunikationsakustik
Ruhr-Universität Bochum
Universitätsstr. 150
44787 Bochum, Germany
Tel.: ++49-234-3222490
Fax.: ++49-234-3214165
Email: novo@ika.ruhr-uni-bochum.de

## ABSTRACT

The aim of an Auditory Virtual Environment (AVE) is to create situations in which humans have auditory perceptions that do not correspond to their real environment but to a virtual one. Applications such as man-machine interaction, entertainment or telepresence frequently require plausible interactive auditory virtual environments. By plausible it is meant that the user perceives the auditory events as having occurred in a real environment although no details of this environment are specified to him. In this presentation the most widely employed source, room and listener models will be reviewed and their simplifications discussed.

## INTRODUCTION

An Auditory Virtual Environment (AVE) is, like a Real Auditory Environment (RAE), composed of sound sources, a medium and a receiver. Besides, in an AVE there is a signal-processing unit. While in the real world a sound, once generated, propagates through the environment until eventually arriving at the listener, in a virtual environment the signal processing performs the equivalent task.

As illustrated in figure 1, the signal-processing unit accesses the audio signal at the sound-source module (sound generation and directivity), filters it with the delays and the effects of boundary reflections calculated in the medium module (geometrical and acoustical data of the environment, reflection-filter database and sound-field model) and subsequently filters the result according to the chosen reproduction format in the receiver module (head-related transfer-function database and reproduction formats).

Besides the references given on particular aspects of AVEs throughout the text, the paper by Lehnert and Blauert [1], the paper by Savioja et. al. [2] and the book by Begault [3] constitute a good complement to the short overview presented here.

## SOUND SOURCES MODELS

### Sound Generation

The generation of a digital audio signal can be achieved either by recording or by synthesis. In the former case the audio signal should be recorded under anechoic conditions to make it suitable for reflections to be added. The recorded signal should exhibit a high Signal-to-Noise Ratio to avoid undesirable artefacts in the virtual environment. It should also be recorded as a mono signal, because the sound sources are usually treated as point sources by the room model.

Instead of being recorded, the audio signals can also be synthesised. The physical modelling of the sound source is a powerful approach as it allows adjusting the model parameters in a physically predictable way [4]. When compared to recorded sound, synthesization requires less data to be transferred, but at the cost of increased computation. An area of present intense research is in audio-haptic interaction [5], where aspects such as asynchronicity between audio and haptic events are studied.

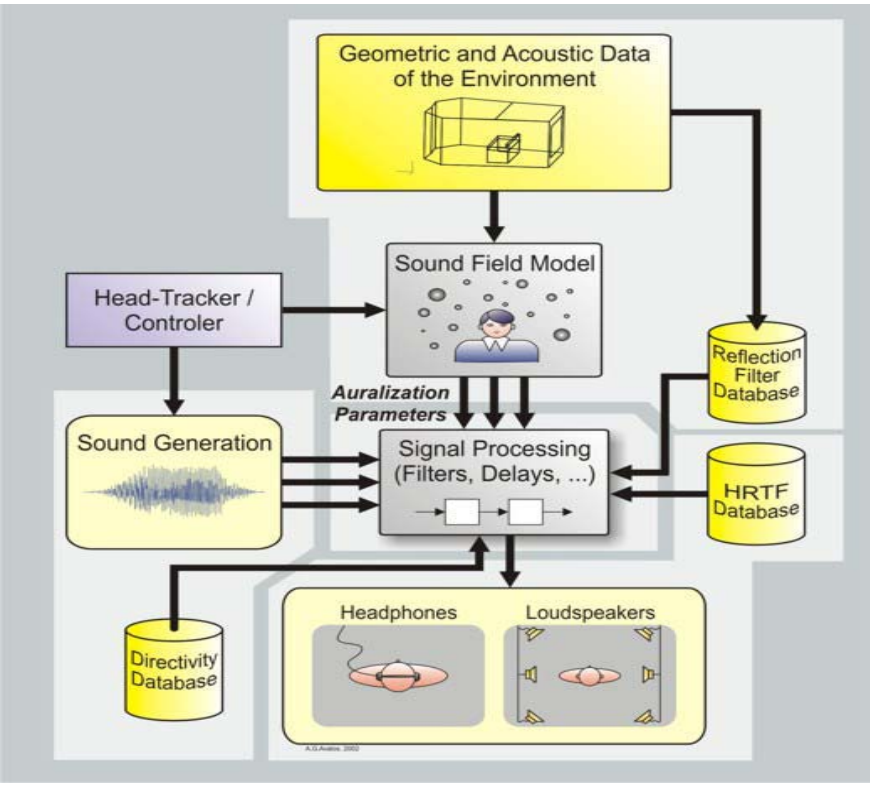


*Fig.1 Block diagram of a typical Auditory Virtual Environment*

Directivity Models

Directivity models can be implemented using two basic approaches: directional filtering and a set of elementary sources. Directivity models can be used, for example, to model the human head directivity [6] or to model the directivity of musical instruments or loudspeakers [7]. When the radiation pattern of a sound source is not suitable for a point-source approximation as is, for example, the case of a clarinet, several point sources can be used simultaneously. If the filters are to be used in real time applications, simplifications have to be introduced in order to match real-time constraints [8].

## ROOM MODELS

Room Models

In a bounded space the listener receives both direct sound and the sound reflected from the boundaries. In order to calculate the boundary reflections, an acoustic model of the environment needs to be built. This involves the definition of both the geometry and the acoustic properties of the boundaries. The methods available to model sound propagation in rooms can be divided in three classes: wave-based, statistics-based and ray-based methods [9, 10].

Wave-based methods are the most accurate as the wave nature of sound is preserved in this approach. However, as the wave equation has analytical solutions only for the simplest geometries and boundary conditions, numerical methods have to be employed to solve most problems of practical interest. Among them are the Difference Methods (DM), of which Finite-Difference Time-Domain (FDTD) is a widely used method [11], and Element Methods (EM), of which the Boundary-Element Methods (BEM) and the Finite-Elements Methods (FEM) [12] are the most employed. FEM and BEM are only appropriate for the calculation of low frequencies, due rapid increase in computational requirements as the frequency increases. But even in the low frequency range neither the DM or the EM methods are suitable to be used in real time systems due to their high computational requirements.

Statistical Methods (SM) such as the Statistical-Energy-Analysis (SEA) method [13] are usually used for noise-level prediction in coupled systems in which sound transmission by structures plays an important factor. However,

the output of the SM is not adequate for subsequent auralisation, as it does not provide information on individual reflections.

In Ray-based Methods rays substitute waves and geometrical-acoustics rules the propagation of sound. Therefore, phenomena such as interference or diffraction are not easily taken into account. The geometrical approximation is valid when the sound wavelength is small compared with the global dimensions of the surfaces and large compared with their roughness.

The most employed Ray-based methods are the Ray-Tracing (RT) [14] and the Image-Source (IS) method [15]. In the RT algorithm the source emit rays, which are reflected at the domain boundaries. The rays are followed throughout the domain until they became attenuated below a specified threshold, leave the domain or reach the listener. The listener is modelled as a detection object and a sphere, due to its isotropic characteristics, is commonly used. The most used reflection rule is specular reflection, although diffusion can be added at the cost of extra computation [1].The IS method performs, like the RT, a geometrical approximation to the sound propagation. The reflection paths from the source to the listener are calculated by sequentially mirroring the sound source against the room boundaries. With this methodology, reflections of any order can be obtained. However, it becomes highly expensive as the order of reflections increases.

A hybrid IS/RT method, [16], has been developed to improve efficiency. With this approach the early reflections are calculated by the IS method due to its high accuracy and efficiency in finding early reflections and the late reflections are calculated by the RT method due to its better efficiency in finding higher order reflection paths.

The results of the path-finding process (the room impulse response) calculated either by the RT method, the IS method or the hybrid RT/IS method is exemplified in figure 2. The impulse response shown in this figure was calculated using a specular reflection rule, which explains the gaps, observed between reflections.

An alternative to the physical approach is the perceptual approach. In this case the reflections (early and late) are not related to a particular geometry or boundary properties, but are adjusted to convey specific auditory perceptions (e.g. room size or source distance) [17, 18].

Air and Wall Filters

The interaction of a wave front with a boundary is a very complex phenomenon, which depends both on the frequency and the angle between the incoming wave and the boundary [19]. For real-time applications simplifications such as an angle independent treatment and neglecting the wave phenomena (scattering and diffraction) are often adopted. The absorption of sound by air depends mainly on distance, temperature and humidity. Tables with analytical expressions for the absorption values can be found, for example, in [20].

Reverberation Modelling

Real-time calculation of a complete room impulse response (as in figure 2) is beyond present computational capabilities. However, if the late reverberant field is assumed as nearly diffuse and the corresponding impulse response exponentially decaying random noise it is possible to calculate the late reverberation without having to calculate it's individual reflections. Reverberation models should exhibit an exponentially decaying impulse response with a dense pattern of reflections to avoid fluttering in the reverberation, a reverberation which decreases as a function of frequency in order to simulate the air-absorption and low-pass-filtering characteristics of the materials and the production of partly incoherent signals at the listeners' ears in order to produce a good spatial impression. Two good reviews of reverberation models can be found in [21, 22].

## REPRODUCTION FORMATS

The formats available for audio reproduction can be divided according to their use or not of Head-Related Transfer Functions (HRTF). An HRTF represents a transfer function from a fixed point in space to a point in the test-person's ear canal [23, 24].

In a headphone system a monophonic time-domain signal is filtered with the left and right HRTFs to create in the listener the perception of a virtual source. The inverse headphones transfer function is also used to deconvolve the signal from the headphones' own filtering. In order to include dynamic cues, such as head movement, a head tracker should be employed.

In a loudspeaker-based system, the monophonic time-domain signal is filtered with the left and

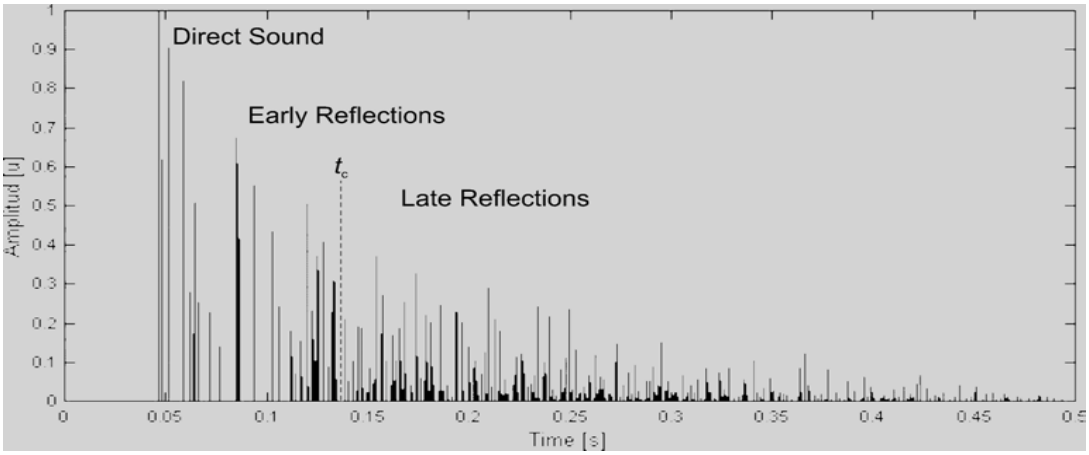


*Figure 2 Simulated room impulse response*

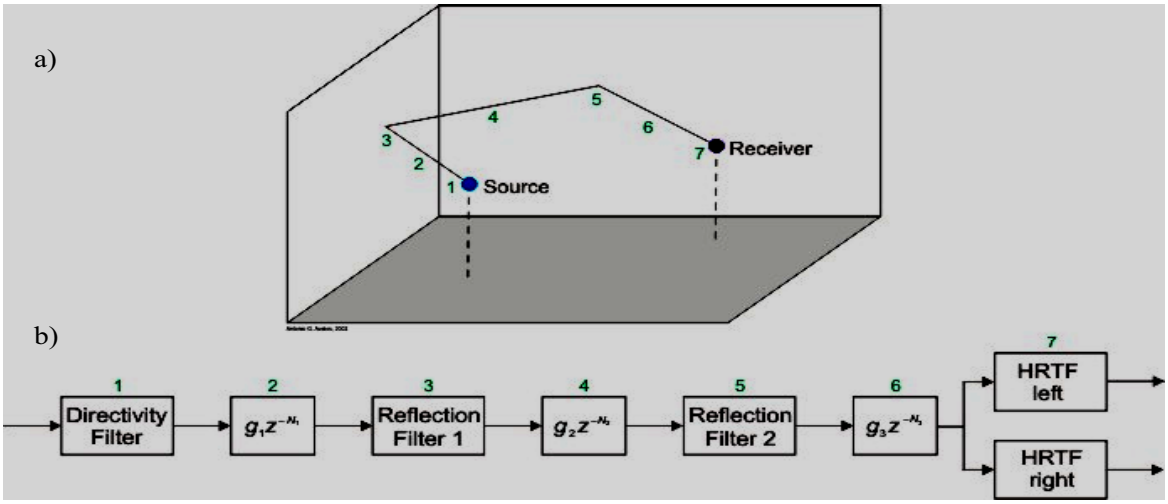


*Fig. 3 a) Spatial and b) Block diagram of a second order auralization unit*
(g – gains ; $z^{-n}$ – delays)

right HRTFs and the inverse loudspeaker transfer function is used to deconvolve the signal from the loudspeaker's filtering. Besides, it is also necessary to consider the loudspeaker's ipsilateral and contralateral transfer functions to design the cross-talk cancelling filters. These filters are necessary to cancel the left-loudspeaker signal in the listener's right ear and vice-versa. This system limits the listener position to a very reduced area unless a headtracker is used [25]. Also, the listening room should be appropriately damped in order to avoid wall reflections to disturb the signal arriving at the listener's ears.

Within the reproduction formats not using HRTFs the Vector-Based Amplitude Panning (VBAP) [26], Ambisonics [27], Wave-Field

Synthesis [28] and ITU 5.1 [29] are the most widely used. The VBAP reproduction method allows arbitrary loudspeakers placement and it uses the principle employed in standard stereo. Ambisonics performs sound field synthesis using spherical harmonics for the decomposition and composition of the sound field. This format can reproduce sounds situated over the 360 degrees of the horizontal plane (pantophonic systems) or over the full sphere (periphonic systems). Wave-field synthesis aims at reproducing the wavefield over the entire space. Unlike the previous methods, Wave-Field Synthesis does not limit the listener to a particular listening spot, although its set up requires a considerable number of loudspeakers. This format is presently subject to intensive research and it promises to be the next breakthrough in audio reproduction[30].

## SIGNAL PROCESSING

Having described the main components of an AVE the focus turns now to the signal-processing module. Figure 3 illustrates the transformations undergone by a sound-source signal, which goes through two wall reflections before arriving at the listener (g represents a gain and $z^{-n}$ represents a delay). Stage 7 of the block diagram is dependent on the chosen reproduction format (in this example HRTFs filters are employed).

There is also the possibility to use a pre-recorded or a pre-computed impulse response. This approach has the advantage of allowing the use of real impulse responses, which are interpolated in real time for positions and orientations not present in the database [31]. It has, however, the disadvantages of requiring a huge storage capability and not allowing real-time modifications of the source, room or listener parameters.

Whatever approach is chosen, the "smoothness" and the "responsiveness" of the system are critical parameters of a real time AVE. Responsiveness is related to delay with which the system responds to an action of the subject. Smoothness is related with the refresh rate with which the auralization unit takes account of a changing auditory scenario [23, 32].

Parallelizing of the signal processing is a possible strategy to speed-up the calculations. However, care should be taken to prevent the potential gains obtained by the parallelization to be overthrown by the communications overhead.

## PERCEPTION OF ONE'S OWN VOICE

Among recent developments in auditory virtual environments the perception of one's own voice (POV) has a particular interest due to its impact in applications such as teleconference systems, systems for training singers or professional speakers and also due to its potential to increase the acceptance of headphones. Pörschmann [33] has extensively investigated in this area and his main conclusions are described next.

The sound of our own voices contributes significantly to our perception of real and virtual environments and has strong implications on the way we speak. It allows our speech to be controlled, which is of particular relevance in acoustically "difficult" situations, for example when speaking in a noisy environment. As early as 1911 Lombard recognized that people speak louder when they are not able to hear their own voices [34]. Since then, several investigations on the influence of a changed perception on the way we speak have been performed. However until now it has remained unclear how the perception of one's own voice affects the perception of a complete auditory scene.

According to Lehnert and Giron [35], Békésy's model of the relevant pathways for the perception of one's own voice can be diversified by splitting the air-conducted sound in two components. Hence three components influence the perception of one's own voice: Sound transmission through the air and around the speaker's head to the ear drums (direct air-conducted transmission); Internal sound transmission inside the head through bones and the skull to the colchea (bone-conducted transmission); Reflection of ones' own voice on acoustically relevant surfaces in the surrounding environment (indirect air-conducted transmission). The absence of one or more of these components commonly leads to an unnatural perception of one's own voice. For example the absence of reflections can be perceived while speaking in an anechoic chamber, which can make people who are not used to such an environment very uneasy.

Two psychoacoustic experiments were performed. The first one aimed at one hand to evaluate the naturalness of the compensation of the headphones' insertion loss by comparing the sound of one's own voice with and without

headphones. On the one hand it was shown that compensating the headphone's insertion loss enhances the naturalness of the perception of a speaker's own voice. On the other hand it was investigated which parameters of the compensation of the insertion loss filter have to be treated carefully and which ones affect the perception of one's own voice less. It has been shown, for example, that the feedback filter extremely susceptible to delays.

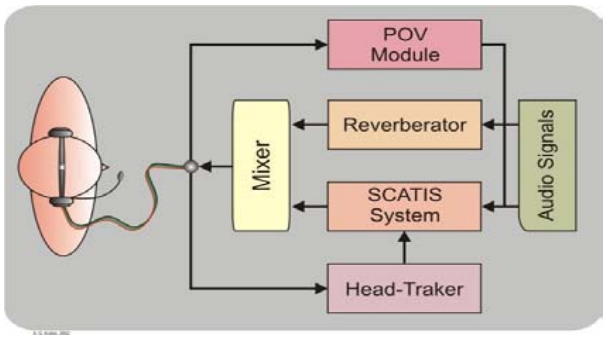


*Figure 4 Structure of the complete auditory virtual environment system (Scatis System - IKA's dsp based AVE; POV- Perception of One's Own Voice)*

In the second experiment the implications of an adequate presentation of one's own voice in the sense of presence were investigated. The results of the psychoacoustic experiments indicate that both the presentation of reflections and the compensation of the insertion loss contribute significantly to an increase in presence.

Generally speaking the adequate perception of a speaker's own voice in a virtual environment can be regarded as being one step to enhance virtual environment generating systems which require vocal communication.

## CONCLUSIONS

Auditory virtual environments replicate real auditory environments by creating models for sound sources, propagation medium, and listener characteristics. In a virtual environment the signal processing module performs what in a real environment is performed by the laws of physics: the propagation of the audio signal through the environment. Different approaches are available for source, medium and listener models. However, for most applications real-time processing constraints the models selection and the simplifications imposed.

Among recent developments, studies in the perception of one's own voice have a particular relevance due to its impact in applications such as immersive teleconferencing. Areas with a great research potential include multimodal interactions and the further development of methods to judge the quality of a simulation.

## ACKNOWLEGMENTS

I would like to thank Prof. Jens Blauert for the invitation to present this overview within the EAA-SYMPOSIUM on Communication Acoustics. Parts of this paper will be published at the EAA meeting to be held in September 2002 in Seville, Spain.

## LITERATURE

[1] H. Lehnert and J. Blauert, "Principles of Binaural Room Simulation", *Appl. Acoust.*, vol. 36, pp. 259-291, 1992.

[2] L. Savioja, J. Huopaniemi, T. Lokki, R. Väänänen, "Creating Interactive Virtual Acoustic Environments", *J. Audio Eng. Soc.*, vol. 47, No.9 pp. 675-705, Sept.,1999.

[3] D. Begault, "*3-D Sound for Virtual Reality and Multimedia",* Academic Press, Cambridge, MA, 1994.

[4] J. O. Smith, "Physical Modeling Synthesis Update", *Comput. Music J.*, vol.20, pp.44-56, Summer,1996.

[5] R. P. Wildes and W. A. Richards. "Recovering Material Properties from Sound", In Whitman Richards, editor, *Natural Computation*, Cambridge,Massachusetts, The MIT Press, 1998.

[6] J. Huopaniemi, K. Kettunen, and J. Rahkonen, "Measurement and Modeling Techniques for Directional Sound Radiation from the Mouth", *in Proc. IEEE Workshop on Applications of Signal Processing to Audio and Acoustics* New Paltz, NY, Oct. 1999.

[7] J. Meyer, "*Acoustics and the Performance of Music",* Verlag das Musiknstrument, Frankfurt/Main, Germany, 1978.

[8] J. Huopaniemi, M. Karjalainen, V. Välimäki and T. Huotilainen, "Virtual Instruments in Virtual Rooms – A Real Time Binaural Room Simulation Environment for Physical Models of Musical Instruments", in *Proc. Int. Computer Music Conf. (ICMC' 94),* pp. 455-462, Aarhus, Denmark, Sept. 1994.

[9] H. Kuttruf, "Sound Field Prediction in Rooms", in *Proc. 15th Int. Congr. on Acoustics (ICA '95)* pp. 545-552 Trondheim, Norway, June 1995.

[10] A. Pietrzyk, "Computer Modeling of the Sound Field in Small Rooms", in *Proc. AES15th Int. Conf. on Audio and Acoustics on Small Spaces,* pp. 24-31, Copenhagen, Denmark, Oct 1998.

[11] D. Botteldooren, "Finite Difference Time Domain Simulation of Low-Frequency Room Acoustic Problems", *J. Acoust. Soc. Am.*, vol.98, pp. 3302-3308, 1995.

[12] M. Kleiner, B.-I. Dalenbäck, and P. Svensson, "Auralization- An Overview", *J. Audio Eng. Soc.*, vol. 41, pp. 861-875, Nov.1993.

[13] R. Lyon and R. Dejong, "*Theory and Applications of Statistical Energy Analysis",* 2nd ed. Butterworth-Heinemann, Newton, MA, 1995.

[14] A. Kulowski, "Algorithmic Representation of the Ray Tracing Technique", *Appl. Acoust.*, vol. 18, pp. 449-469,1985.

[15] J.Borish, "Extension of the Image Model to Arbitrary Polyhedra", *J. Acoust. Soc. Am.*, vol. 75, pp. 1827-1836, 1984.

[16] M. Vorländer, "Simulation of the Transient and Steady State Sound Propagation in Rooms using a New Combined RayTracing/Image Source Algorithm", *J. Acoust. Soc. Am.*, vol. 86, pp. 172-178, 1989.

[17] R. Pellegrini "Perception-Based Room Rendering for Auditory Scenes", *109th Convention of the Audio Engineering Society,* Los Angeles, preprint 5229, 2000.

[18] J. Jot, "Spatialisateur", *Multimedia Systems*, vol. 1, 1999.

[19] J. Huopaniemi, L.Savioja, and M.Karjalainen, "Modelling of Reflections and Air Absorption in Acoustical Spaces - A Digital Filter Design Approach", in *Proc. IEEE Workshop on Applications of Signal Processing to Audio and Acoustics*, New Paltz, NY, Oct. 1997.

[20] H. Bass, and H.J. Bauer, "Atmospheric Absorption of Sound: Analytical Expressions", *J. Acoust. Soc. Am.*, vol. 52, pp. 821-825,1972.

[21] W. Gardner, "Reverberation Algorithms", in *Applications of Digital signal Processing Algorithms to Audio and Acoustics*, pp. 85-131, M.Karhs and K. Brandeburg, Eds., Kluwer Academic, Boston, MA, 1997.

[22] B. Blesser, "An Interdisciplinary Synthesis of Reverberation Viewpoints", *J. Audio Eng. Soc.* vol. 49, No.10 pp. 867-903, October, 2001.

[23] J. Blauert, *"Spatial Hearing, The Psychophysics of Human Sound Localization"*, MIT Press, 1996.

[24] H. Moller, M.F. Sorensen, D. Hammershoi, and C.B. Jensen, "Head Related Transfer Functions of Human Subjects", *J. Audio Eng. Soc.*, vol. 43, pp. 300-321, May, 1995.

[25] W.G. Gardner "*The virtual Acoustic Room"*, Master Science Thesis, MIT,1992.

[26] V. Pulkki "Virtual Sound Source Positioning Using Vector Based Amplitude Panning". *J. Audio Eng. Soc.*, vol.45, no.6, pp. 456-466, June, 1997

[27] D. Malham and A. Myaat, "3-D Sound Spatialization Using Ambisonic Techniques", *Comp. Music J.*, vol.19, no. 4, pp. 58-70, 1995.

[28] A.J. Berkhout, "A Holographic Approach to Acoustic Control", *J. Audio Eng. Soc.*, vol.36, Number 12 pp. 977-995, Dec 1988.

[29] Rec. ITU-R BS.775 "Multichannel Stereophonic Sound Systems With and Without Accompanying Picture", 1994.

[30] M. Boone, N. Verheijen, P. van Tol, "Spatial Sound Filed Reproduction by Wave Field Synthesis", *J. Audio Eng. Soc.*, vol. 43, Number 12, pp. 1003, 1012, Dec. 1995.

[31] R. Pellegrini, *"A Virtual Reference Listening Room as an Application of Auditory Virtual Environments"*, Ph.D. Thesis, dissertation.de – Verlag im Internet , Berlin, 2002.

[32] E. Wenzel, "Analysis of the Role of Update Rate and System Latency in Interactive Virtual Acoustic Environments", *J. Audio Eng. Soc. (Abstracts)*, vol. 45, pp. 1017, 1018, preprint 4633, Nov. 1997.

[33] C. Pörschmann, "One's Own Voice in Auditory Virtual Environments", *ACUSTICA united with acta acustica*, vol. 87, Number 3 pp. 378-388, May/June 2001.

[34] E. Lombard, "L Signe de l'Élévation de la Voix", *A. Maladies Oreille, Larynx, Nez, Pharynx*, pp. 101-119, 1911

[35] H. Lehnert, F. Giron, "Vocal Communications in Virtual Environments", *Virtual Reality World*, pp. 279-293, 1995.