VIRTUAL ACOUSTICS: OPPORTUNITIES AND LIMITS OF SPATIAL SOUND REPRODUCTION

Michael VORLÄNDER

RWTH Aachen University
Institute of Technical Acoustics
52056 Aachen, Germany
e-mail: mvo@akustik.rwth-aachen.de

Virtual Acoustics is part of the emerging field of "Virtual Reality". The technology for creating a Virtual Reality, VR, for wide variety of applications in university and industry has been developed in the last decade. Mostly VR is understood as a tool for 3D visualization, rather than for spatial audio or room acoustics. Nevertheless an important requirement of VR is the multimodal approach which includes vision, sound, tactile and haptic stimuli. The process of creating a physical stimulus based on computer data is called "rendering". The development of rendering and reproduction of acoustic stimuli in VR is now at a stage where integration of 3D sound is feasible by using PCs. This applies to multi-channel binaural synthesis as well as to full room-acoustic simulation algorithms and to various applications of 3D sound stimuli for audiology, neuropsychology or any other application in acoustics and noise control.

Keywords: virtual acoustics: simulation and auralization, spatial audio.

1. Introduction into Virtual Reality

Virtual Reality (VR) is a system approach which provides the user's immersion and presence in computer-generated virtual environments [5]. An important characteristic of VR is a three-dimensional and multimodal interface between computer and human. Besides vision and acoustics, more senses covering haptics (force feedback), tactiles and eventually others are added. Acoustics in VR, however, was so far mostly included just as effect and without plausible or authentic reference to the virtual scene. This situation is changed when a physically consistent computer simulation of the acoustic signal with regard to sound and vibration sources and transmission is implemented.

The process of generating the cues for the respective senses (3D image, 3D audio, . . .) is called "rendering". Simple scenes of interaction, for instance when a person is leaving a room and closes a door, require complex models of room acoustics and sound insulation. Otherwise the colouration, loudness and timbre of sound in and

Immersion: Intuitive three-
dimensional user interface

VR

Multimodality: Interaction real-time:
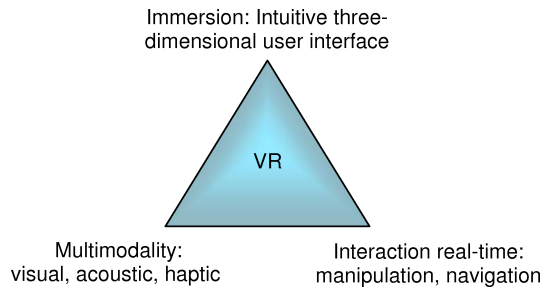visual, acoustic, haptic   manipulation, navigation

Fig. 1. Dimensions of virtual reality.

between rooms are not represented sufficiently. Another example is the movement of a sound radiating object behind a barrier or inside an opening of a structure, so that the object is no longer visible but can still be touched and heard. Sound also propagates by diffraction, one of the most difficult phenomena in general linear acoustics. The task of representing a realistic acoustic perception, localization and identification is therefore a big challenge.

Personal computers have just recently become capable of solving acoustic field problems in real time. Still numerous approximations must be made, but in the end the resulting sound needs not be physically absolutely correct, but only perceptually correct. Knowledge about human sound perception is, therefore, a very important pre-requisite to evaluate auralized sounds and to set targets on the algorithmic performance. One of the basic tasks in creating a virtual acoustic scene is to place a sound source into a real-life environment, in order to create a natural spatial impression.

In audiology, this technique is of interest for testing the performance of binaural human hearing in complex environments of spatial sound, speech and noise configurations. Furthermore, the performance of hearing aids and cochlear implants (CI) can be tested in laboratory conditions of more realistic real-world situations. This opportunity includes sound and noise source rendering with in principle all degrees of freedom. Also full room-acoustic situations of direct and diffuse sound can be created. These approaches of 3D sound reproduction are in fact state of the art in research in audiology and neuropsychology. The key feature of virtual reality, however, is not yet implemented, which is multimodality and interactivity. In future it can be expected that more tests with 3D sound and vision, including also multimodal scenarios and interaction, will be used in audiological testing and hearing aid fitting. The VR technology must be more user-friendly and more flexible concerning individual filters. The necessary modifications of standard VR technology for audiology are discussed in more detail in this contribution.

2. Spatial sound systems

A relatively dry room and artificial reverberation introduced by a number of loud-speakers is a first approach of a multi-purpose test environment. The control parameters

are the direct-to-reverberation sound level and the reverberation time. The temporal structure of early reflections and the corresponding perception of room size is represented not exactly, but at least with a certain plausibility. This system approach is sufficient for many applications in test laboratories, but it does not comply with the specific requirements and control parameters of spatial sound field synthesis.

Generally, spatial sound fields can be created with loudspeaker by using one of two general concepts. One can try to reproduce head-related signals, taking advantage of the fact that the hearing sensation only depends on the two input signals to the ears. Also, loudspeakers arranged around a listening point ("sweet spot") may serve for a spatially distributed incident sound field. Furthermore, one can try to create a complete wave field incident on the listening area. The potential to involve more than one listener in the second approach illustrates the conceptual difference between the two methods.

Surround sound technology can be defined in various ways. One basic form of sound source imaging in the horizontal plane is the well-known stereo setup or a surround sound system. These approaches make use of the psychoacoustic effect of phantom sources. However, the perception of phantom sources as such cannot be studied with a system which inherently includes this effect. Alternatively, the basis can be a multi-channel microphone separating the incident field into spherical harmonics. With proper reproduction setup, the 3D sound field at the listener's point is reconstructed exactly. This method does not inherently include the effects of phantom sources and, accordingly, these effects can be part of the test. The accuracy of the spherical harmonics approach ("Ambisonics" [8]) depends on the frequency range and the corresponding order of spherical harmonics functional basis.

More accurate for sound field reproduction is the method of Wave Field Synthesis, WFS [3]. Here, an approximation of the spatial sound field is created by using a microphone array, but not at a specific listener point. Instead, the microphone arrangement is larger and it is located on elementary geometric figures like straight lines or circles around the listening area. Using this approach, a large sweet spot can be created and more than one listener can be served with spatial sound.

$x = x_1$ $x = x_2$

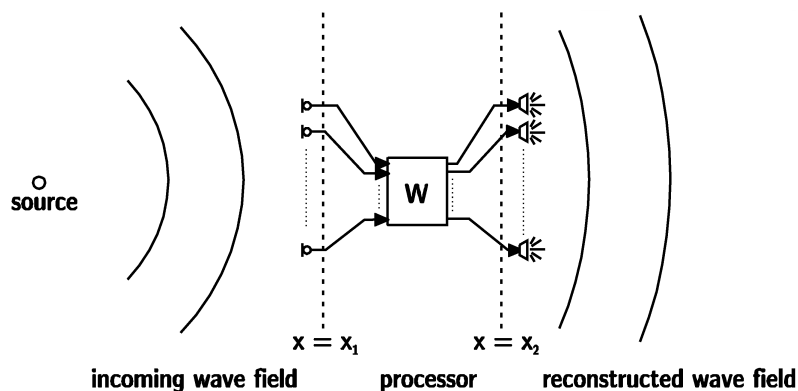incoming wave field   processor   reconstructed wave field

Fig. 2. Field reconstruction by loudspeaker arrays (after [2]).

The wave decomposition is achieved by analyzing the signals in microphone arrays. According to Huygens principle, the points where the sound pressures were recorded at the microphone positions can be interpreted as elementary sources. In replay situation, the wave field is reconstructed by sending waves from these points. This illustrates the step from wave field analysis to synthesis.

2D WFS is quite well feasible and successful. In case of approximately some hundred loudspeakers arranged in a surrounding line array in the horizontal plane, the wave field is reconstructed with good accuracy up to several kilohertz. Shortcomings for the time being are a rather big amount of hardware needed and the lack of easy-to-use synthesis and authoring tools. The same, by the way, holds for binaural synthesis which will be explained in the next sections.

3. Binaural technology

A mono source signal, properly characterized and calibrated according to well-defined specifications (see below), can be processed in such a way that its perceptual cues are amended by a spatial component. A stereo or surround setup is capable of creating an effect of virtual sources which can produce an appropriate spatial effect [1]. A binaural mixing console can be used for processing headphone or loudspeaker signals by using head-related transfer functions, HRTF [4]. With a database of HRTF or the corresponding head-related impulse responses, HRIR, any direction of sound incidence can be simulated, when a mono source $s(t)$ is convolved with a pair of head-related impulse responses.
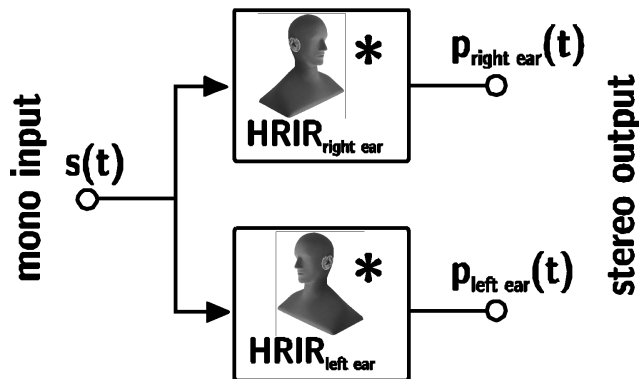

Fig. 3. Binaural synthesis.

$$
\begin{aligned}
p_{\text{right ear}}(t) &= s(t) * \text{HRIR}_{\text{right ear}}, \\
p_{\text{left ear}}(t) &= s(t) * \text{HRIR}_{\text{left ear}}.
\end{aligned}
\tag{1}
$$

With the technique of binaural mixing we can create multi-channel binaural synthesis. This tool is adequate for auralization of free-field environments for a small number

of sources or reflections. Modelling of free-field propagation requires a source recording and analytic calculation of the complex amplitude of the sound pressure signal at the receiving point. For spherical waves the free field spectrum corresponding to one source reads

$$\underline{H}|_{\text{left,right}} = \frac{\mathrm{e}^{-jkr}}{r} \cdot \underline{H}_{\text{source}}(\theta, \phi) \cdot \underline{H}_{\text{air}} \cdot \mathrm{HRTF}(\vartheta, \varphi)|_{\text{left,right}}, \tag{2}$$

with τ denoting the delay, jkr the phase lag due to retardation, $1/r$ the distance law of spherical waves, H_{source} the source directivity in source coordinates, H_{air} the low pass of air attenuation, and HRTF the head-related transfer function of the sound incidence in listener coordinates at a specified orientation. For N sources the resulting signal $p(t)$ is created by superposition. In time domain formulation, this procedure leads to

$$p(t)|_{\text{left,right}} = \sum_{i=1}^{N} s_i(t) * \mathrm{IFT}(\underline{H}_i|_{\text{left,right}}), \tag{3}$$

with H_i denoting the spectra of the propagation function of source i. "Sources" in this respect are speech, sounds and noise.

With the possibility to place and move 3D sounds in space, the most important components of virtual acoustics are introduced already. Moving sources or listeners in the scene is done by block-wise processing of Eq. (3) with filter updates according to the actual positions, as will be discussed below. But still the problem of virtual room acoustics including full reverberation must be discussed.

4. Simulation of sound fields in rooms

Computer simulation of room acoustics first applied in [11]. Many articles about development of algorithms were published since then. The algorithms of typical programs used today are based on geometrical acoustics. The description of the sound field is reduced to energy, transition time and direction of rays. The methods were, at first, used to calculate the room acoustic criteria (T, EDT, D, C, TS, LF, IACC...). Finally, auralization was introduced for room acoustics at the beginning of the 1990's [10]. Two techniques of geometrical acoustics have to be distinguished: "Ray Tracing" and "Image Sources". Independent of their software implementation, they represent different physical approaches. After all, a binaural impulse response is created from direct sound, early reflections and scattered components by using the concept of binaural synthesis. All components are added. In formulation in the frequency domain, this process is described by the multiplication of the transfer and filter functions representing sound travelling from the source to the receiver, see Eq. (3). As explained above, also scattering plays an important role particularly in the late response. Apparently Eq. (3) deals only with image sources. But it can be generalized easily, if the contributions of the scattered and late part of the impulse response are represented by a set of equivalent reflections H_j, the arrangement of which must be constructed. The basis for this kind

of construction may be stochastic ray tracing, radiosity, free path statistics, or an artificial reverberation process. All methods mentioned yield estimates of the late impulse response envelope, a function of frequency and time. With adding an adequate fine structure which represents the actual reflections statistics, the binaural impulse responses can be created.

5. 3D sound reproduction in virtual reality systems: interaction

Headphones or other audio systems integrated in head-mounted displays are well qualified to serve as reproduction transducers and accordingly they are widely in use. Unfortunately, some disadvantages must be noted which are caused by effects in the sound field between the active element of the headphone and the ear canal of the listener. Wearing comfort, unnatural ear occlusion are additional factors affecting the quality of the hearing sensation. The so-called "in-head localization" is one example for such unwanted effects. Externalization of sound sources is one of the main issues in discussion of headphones. In case of insufficient externalization the immersion in VR system is drastically reduced. With a proper equalization and special attention to high-frequency radiation into the ear canal this problem can be partly solved. Adaptive filtering (head tracking, see above) for taking head movements into account is also a very important tool for creation of realistic localization and externalization.

Headphone equalization is by far more difficult than loudspeaker equalization. The radiation impedance acting on the transducer cannot be approximated by using elementary field conditions like "piston in free half space". Instead, the radiation impedance into the ear canal is relevant, which brings us to the first difficulty. Properties of the listeners' ear canals vary significantly among a population of test subjects. As concerns the input impedance, resonances which can be related to individual physiological features are known only in principle. They can be modelled rather easily, but the model parameters are depending on the individual anatomy. Artificial ears were developed as kind of "average ear", but their applications are restricted to special headphone types. Even in ideal measurement conditions for digital equalization and calibration, there remains the uncertainty introduced by mounting at the real ear. And it is very difficult, maybe impossible to achieve a good performance in particular situations such as investigations of test subjects fitted with hearing aids or for individuals not matching the headphone dimensions (children).

Binaural sound can be pre-processed or recorded for some locations and orientation of the listener. Those situations can be created in virtual environments or they correspond to real rooms. In both cases the binaural impulse responses are available in a certain grid in lateral and spherical coordinates. In replay situation, dry sound is convolved with the valid binaural impulse response for the actual position and orientation of the listener (so-called "walkthrough" [6]). Listener movement is tracked. Accordingly the best matching binaural filter for the position and orientation is chosen for convolution. Headphone reproduction without head-tracking would be head-related and not room-

related. For obtaining sufficient presence in the virtual environment, only a coupled system of head tracking and binaural synthesis of room-related sources, image sources and reverberation is appropriate. In fact, an extension towards room-related coordinates creates a very big effect of enhancing plausibility and immersion, even for usage of non-individualized HRTF. This finding illustrates the importance of dynamic cues of localization and the necessity to implement this feature into binaural systems such as the stereo dipole [9] and the virtual headphone [7, 13].

6. Outlook: opportunities and limits

Virtual Reality is a computer-generated environment for interaction in real time. One important specification of VR is multimodality of the human-computer interface. Most VR systems were developed initially for 3D vision. In order to obtain presence and immersion of the user, VR is not complete without the modalities of acoustics and haptics (and more). The driving forces for establishing VR applications are task-specific interaction scenarios and their acceptance by the user and user feedback.
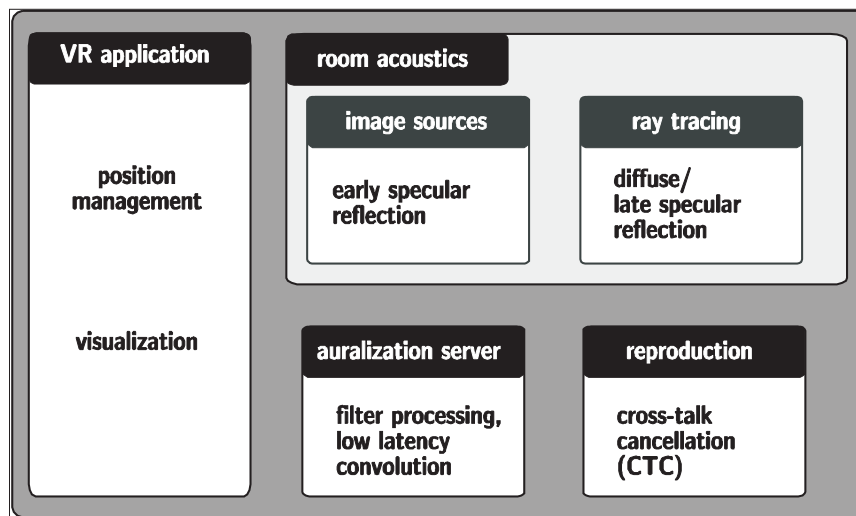

Fig. 4. VR System with integration of binaural room acoustics, after [13].

In the discipline of visual VR the recent development of graphic processors enabled simulation and reproduction of quite complex scenarios with high degree of realism. Furthermore, the state of technology for VR subsystems like motion trackers and projections units has gained a very high quality at reasonable cost. As soon as application software is available, it is not out of range that VR technology as such is transported from highly sophisticated laboratories arrangements to solutions of daily use and for the consumer market. Particularly the visual dimension of VR is used today by many

groups dealing with visualization of complex numerical or experimental data, like in fluid dynamics or molecular physics.

The field of research discussed in this contribution, virtual acoustics, is just at a starting point. The concept of auralization and real-time processing will not remain being focused on room acoustics, as illustrated in Fig. 5, but it will be extended towards application in general acoustics and noise control engineering [14]. It will require much effort in establishing virtual reality systems with dynamic interaction of the user with the virtual world at same auditory quality as today's offline auralization. The big advantage, however, is the multimodal concept (audio-visual and more).

Fig. 5. Concert hall model in a CAVE-like environment (CAVE®Automatic Virtual Environment).

For audiology, special attention must be paid to uncertainties introduced by any of the spatial techniques. For WFS, the limit is given by spatial aliasing and bulky hardware. For binaural technology, the limits are given by non-individual HRTF, equivalent HRTF for hearing aid and CI microphones. With research and implementation of easy-to-use tools for practical application, such as software for spatial mixing, numerical methods for spatial filter generation (BEM), filter switching without audible artefacts, and specification of standards for real-life test scenarios (home, office, school, traffic, etc.) with specification of audio-visual interaction tests, virtual acoustics can be of great benefit for audiology and related fields of medicine.

Scientific room acoustics and room simulation is not a finished problem. When the publications of the last symposia on room acoustics (RADS 2004 Awaji, Copenhagen Symposium 2006, ISRA 2007 Sevilla) are scanned, we see manifold problems in practical room acoustics, due to the multi-dimensionality of factors influencing the listening impression. We also find activities in solving problems of the physical aspect of sound

fields near surfaces. Scattering plays an important role because the techniques for modelling are not sufficient yet. Calculation and measurement of surface scattering and absorption, too, is a tedious job. Edge diffraction, too, is still a big problem, particularly in implementation in room acoustic computer modelling.

New parameters should extend the scope of ISO 3382, including parameters to better characterize stage acoustics and specific acoustics of opera houses. Parameters for objectively characterizing echoes and coloration are not yet agreed upon unanimously.

Another expanding field in room acoustics is computer simulation and auralization. Here we have to solve two problems: a) source recording and b) real-time processing. The first problem, source recording, is by far not solved properly by the existing CD recordings of anechoic music. When convolved with room impulse responses, this audio event is related to an orchestra sitting in one point (source location). Directional characteristics are not sufficient to make the auralization realistic.

The field of research discussed in this contribution, virtual room acoustics, is just at a starting point. The concept of auralization and real-time processing will not remain in room acoustics, but it will be extended towards application in general acoustics and noise control engineering [14]. It will require much effort in establishing virtual reality systems with dynamic interaction of the user with the virtual room at same auditory quality as today's offline auralization. The big advantage, however, is the multimodal concept (audio-visual and more).

References

[1] BEGAULT D., *3-D sound for virtual reality and multimedia*, Academic Press Professional, Cambridge, MA, 1995.

[2] BERKHOUT A.J., *A holographic approach to acoustic control*, J. Audio Eng. Soc., **36**, 977 (1988).

[3] BERKHOUT A.J., DE VRIES D., VOGEL P., *Acoustics Control by Wave Fid Synthesis*, J. Acoust. Soc. Am., **93**, 5, 2764–2778 (1993).

[4] BLAUERT J., *Spatial Hearing: the psychophysics of human sound localization*, 2nd edition, MIT Press, Cambridge, MA 1996.

[5] BLAUERT J., [Ed.] *Communication acoustics*, Springer, Berlin, Heidelberg, New York 2005.

[6] DALENBÄCK B.-I., STRÖMBERG M., *Real time walk through auralization – the first year*, Proc. IoA spring conference, Copenhagen 2006.

[7] GARDNER W.G., *3-D audio using loudspeakers*, Ph.D thesis, Massachusetts, Institute of Technology, 1997.

[8] GERZON M.A., *Surround Sound Psychoacoustics*, Wireless World, **80**, 483 (1974).

[9] KIRKEBY O., NELSON P.A., HAMADA H., *The "Stereo Dipole" – A Virtual Source Imaging System Using Two Closely Spaced Loudspeakers*, J. Audio Eng. Soc., **46**, 387 (1998).

[10] KLEINER M., DALENBÄCK B.-I. , SVENSSON P., *Auralization – an overview*, J. Audio Eng. Soc., **41**, 861 (1993).

[11] KROKSTAD A., STRØM S., SØRSDAL S., *Calculating the acoustical room response by the use of a ray-tracing technique*, J. Sound Vib., **8**, 118 (1968).

[12] LENTZ T., *Binaural technology for virtual reality*, Doctoral thesis, RWTH Aachen University, Germany, 2008.

[13] LENTZ T., SCHRÖDER D., VORLÄNDER M., ASSENMACHER I., *Virtual Reality System with Integrated Sound Field Simulation and Reproduction* [in:] EURASIP Journal on Applied Signal Processing, Special Issue on Spatial Sound and Virtual Acoustics, 2007.

[14] VORLÄNDER M., *Auralization*, Springer, Berlin 2007.