16 3-D sound reproduction and virtual reality systems

Computer simulation of acoustic scenes is an important prerequisite of rendering. Technology for 3-D sound reproduction, the so-called “audio front end” or the “acoustic human-machine interface” is an essential component of VR systems, which must be capable of fulfilling high quality standards involving psychoacoustically relevant cues. These cues may differ from one VR application to the next. Some applications require an exact localization, while for others, monaural spectral features such as reproduction with exact loudness and timbre are more important.

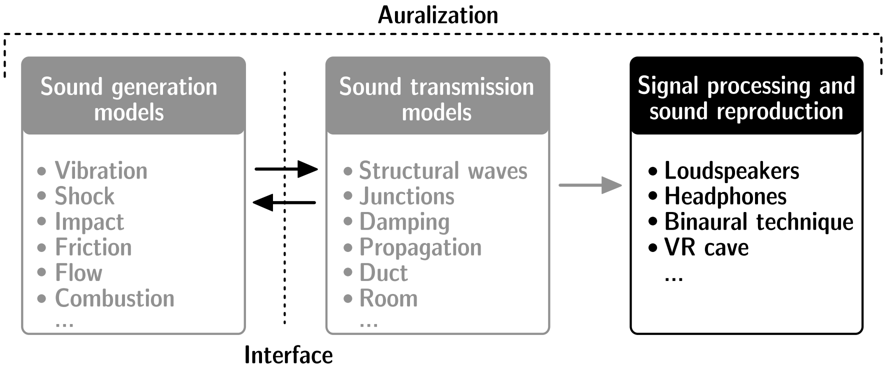

Fig. 16.1. The components of sound reproduction

In visual analogy, modern shutter glasses based on polarization filters or green-red filters in connection with high-definition video displays provide excellent stereoscopic reproduction. A 3-D audio reproduction system for VR applications, however, should not be confused with surround sound systems in consumer electronics. The main difference is that VR applications are related to physical models and a high degree of realism in the components of sound and vibration generation, transmission and reproduction. Recording engineers for classical music have a different goal. Even for live recordings, they will use recording techniques and strategies of microphone placement and mixing to obtain the best result in the final sound for

Fig. 16.2. Concert hall model in a CAVE-like environment (CAVE™ Automatic Virtual Environemnt)

home environments, replayed by stereo or 5.1 equipment. Thus, the balance of instruments or instrument groups is manipulated for aesthetic effects.

First of all, for acoustic virtual reality, we accept the strict and exclusive goal of as much authenticity as possible related to three-dimensional perception. A "neutral" reproduction related to linear distortions is required anyway.

16.1 Headphone systems

Headphones or other audio systems integrated into head-mounted displays are well qualified to serve as reproduction transducers and, thus, they are widely in use. Unfortunately, some disadvantages must be discussed which are caused by physical effects in the sound field between the active element of the headphone and the listener's ear canal. Wearing comfort and unnatural ear occlusion are additional factors that affect the quality of the hearing sensation. The so-called "in-head localization" is one example of such unwanted effects. Externalization of sound sources is one of the main issues in a discussion of headphones. When externalization is insufficient, immersion in a VR system is drastically reduced. With proper equalization and special attention to high-frequency radiation into the ear canal, this problem can be partly solved. Adaptive filtering (head tracking; see Sect. 15.1.1) for taking head movements into account is also a very important tool for creating realistic localization and externalization.

Fig. 16.3. Headphones of various types: open, closed, semiopen

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 such as a "piston in free half space." Instead, the radiation impedance into the ear canal is relevant, which brings us to the first difficulty. The properties ear canals of listeners have a large variance 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 dependent on the individual anatomy.

Artificial ears were developed as kind of "average ear," but their applications are restricted to special headphone types.[83] Even in ideal measurement conditions for digital equalization and calibration, there remains the uncertainty introduced by mounting at the real ear. Particularly in closed headphones, uncertainties are observed due to leakage. Interindividual differences are also important there.

[83] typically audiometric headphones.

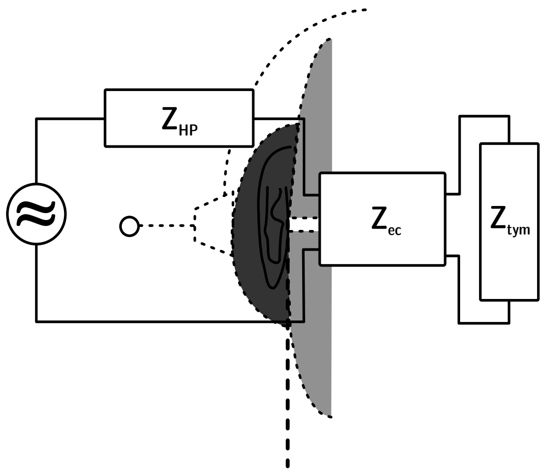

Fig. 16.4. Model of a headphone mounted at the ear

In Fig. 16.4, the headphone source impedance, $\underline{Z}_{HP}$, is coupled with the ear canal impedance, $\underline{Z}_{ec}$, and finally with the termination impedance of the eardrum. Headphones connected to ear canal and eardrum impedances

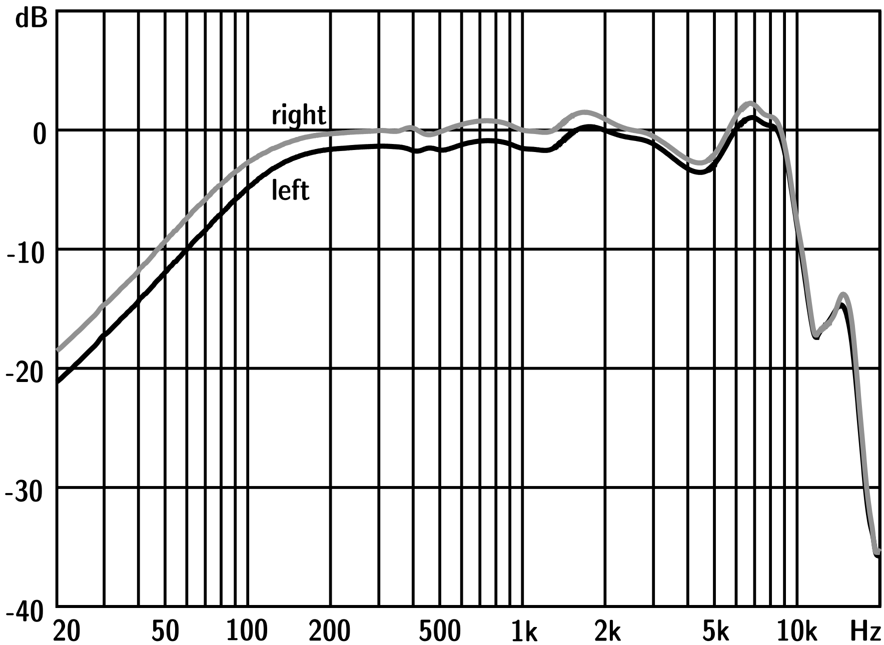

Fig. 16.5. Example of an in-ear headphone frequency response measured with an ear simulator (IEC 60711)

have been under investigation in basic research for many years, until today (Hammershøi 1995). In a purely technical sense, the eardrum sound pressure represents the complete excitation signal of the auditory system. For this assumption, other paths of sound transmission to the cochlea, such as bone conduction, are neglected. With a good model of the ear canal impedance and definition of the sound pressure at the ear canal entrance as the driving signal, the eardrum pressure is given unambiguously. The problem, however, suffers from many uncertainfactors, as described reviously. Also, occlusion by a headphone affects the standing wave pattern and, thus, the transfer impedance of the ear canal. Only in a specific test scenario such as a blocked ear canal is the driving pressure measured with a clear and accurate reference.

16.1.1 Headphone equalization for binaural signals

Headphones used for binaural reproduction should reproduce the binaural signals implemented by binaural synthesis or binaural (dummy head) recording without affecting binaural cues.

Correction (equalization) filters have twofold purposes. First, they must ensure that the sound pressure at the eardrum is identical in the recording and replay situation. If the sound pressure at the eardrum was perfectly recorded, at replay, the sound path between headphone and eardrum is included twice. Thus, a correction filter must be used to extract the ear canal (see Fig. 16.4). Other filters are also used, but these are defined for compatibility between headphone and loudspeaker reproduction of binaural signals.

A plane wave irradiating a human head at frontal incidence is usually the reference. Perfectly equalized loudspeakers create a quasi-plane wave (p and v in phase). When this wave hits a head, however, the eardrum sound pressure is distorted. In fact, the HRTF for frontal incidence is multiplied by the loudspeaker (plane wave) flat spectrum. Binaural recording replayed

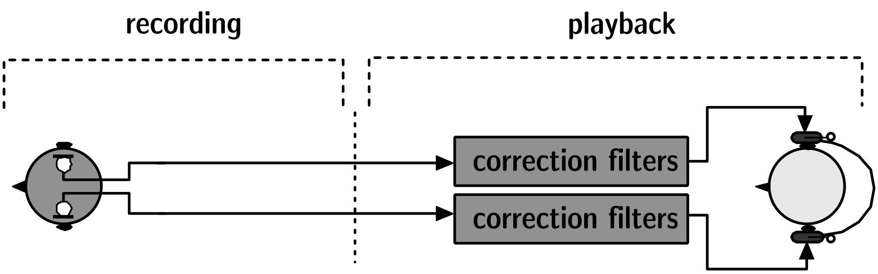

Fig. 16.6. Arrangement for reproducing binaural signals (after (Blauert 1996))

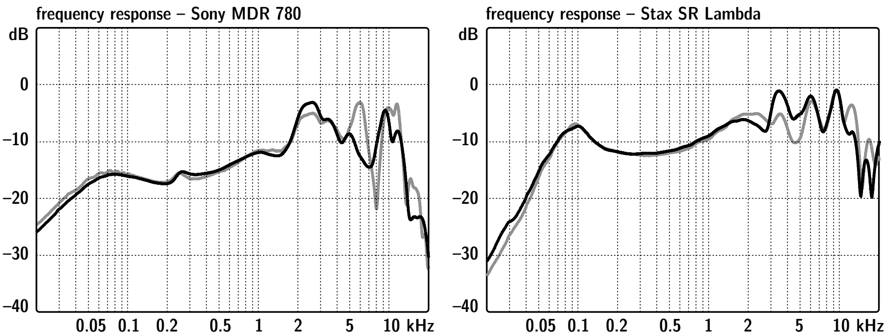

Fig. 16.7. Examples of typical free-field sensitivities of headphones

by loudspeakers, accordingly, carry the HRTF twice, one from the dummy head in the recording situation and the other from the listener's head. Correction filters for compatibility refer to plane wave at frontal incidence, diffuse field incidence or independence of direction (Genuit 1984). When headphones are used, these filters are obsolete for binaural signals.

But due to the need for compatibility with loudspeakers, commercial headphones are equalized with respect to reference sound fields. Free-field headphones deliver, by definition, the same hearing impression as a sound event incident in a plane wave at frontal incidence; diffuse-field headphones accordingly, are used for random incidence of incoherent sound waves. A reproduction of a recorded diffuse field[84] by a diffuse-field headphone yields the same hearing sensation as that in the real (recorded) situation.

16.1.2 Individual filters

All definitions given above hold for standard HRTF and accordingly for average listeners. It was shown by several authors that localization errors can be significantly reduced by using individual filters. The main reason can be identified in individual features in the HRTF, particularly in the frequency range above 6 kHz.

When individualized HRTFs are available by measurement at the listener using probe microphones, all filters can also be obtained by measurement, data processing and feeding into digital filters. The reference situation in this case may be the quotient of the electric voltage fed to a loudspeaker in a certain direction and at a large distance (> 2 m) and the sound pressure measured in the ear canal of a test subject.

[84] for instance, created in a reverberant room far outside the reverberation distance.

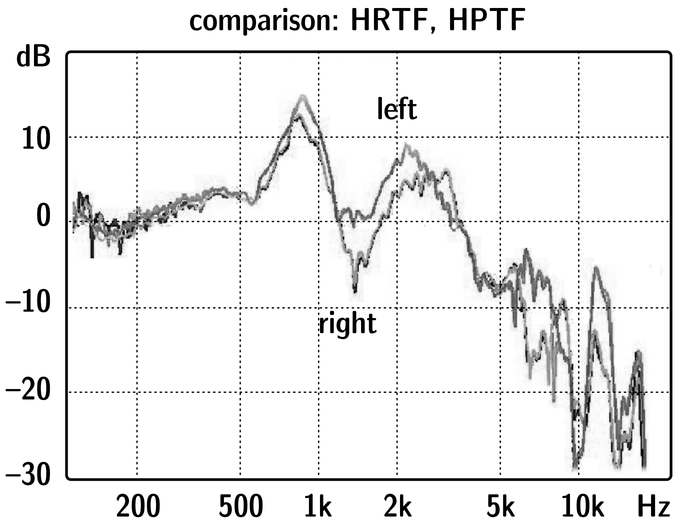

Fig. 16.8. Four equalized binaural transfer functions, two for a loudspeaker and two for headphones (HPTF: headphone transfer function) into the ear canal (individual test subject). The two curves corresponding to left and right ears are identical

This electroacoustic transfer function can be interpreted as an individual HRTF.[85] The same transfer function should be present when the headphone is applied. Figure 16.8 shows a comparison of this experiment. In this figure, four curves are shown, two of each for the right and the left ears, respectively. They are identical with respect to the resolution of the plot. The deviations are smaller than 0.5 dB. Accordingly, no audible effects should be noticed in a listening comparison. In fact, practically no significant differences were found. And those small differences found are more related to the test conditions of placing and replacing the headphone than to technical matters. Furthermore, the psychological component in such tests which aim at very small perceptual effects, must be considered. After all, it has been found that with exact equalization of input signals into the hearing system, the auditory sensation is exactly the same (Blauert 1996).

After individual equalization, the remaining differences between loudspeaker and headphone binaural reproduction systems are less than the just noticeable differences; see Fig. 16.9. More detailed results were published by Møller (1992), also including considerations of the accuracy of binaural cues and the final performance of headphone reproduction in listening tests. Finally, one can state that headphones can be used for 3-D sound reproduction with good results, but their equalization requires much attention.

[85] With exact calibration of the loudspeaker, it is actually the HRTF.

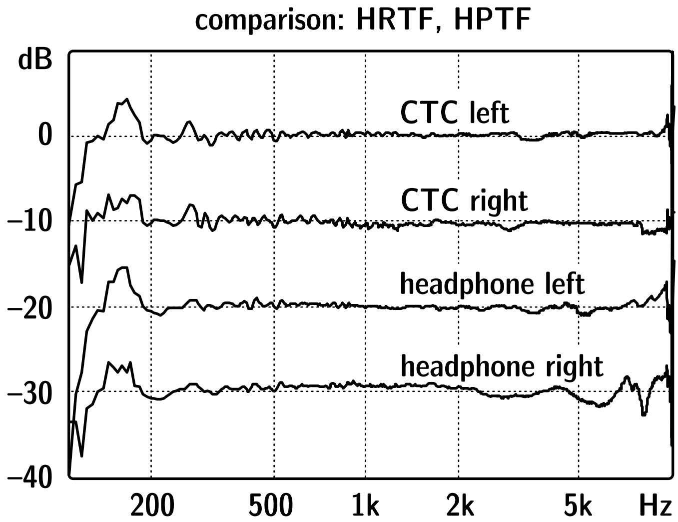

Fig. 16.9. Individually equalized transfer functions of loudspeakers (top, HRTF CTC) and headphones (HPTF bottom) from the free field to the reference point in the ear canal, divided by the individual transfer function between an electric voltage fed into a loudspeaker in a free field and the ear canal sound pressure. Curves are shifted in steps of 10 dB for better visibility

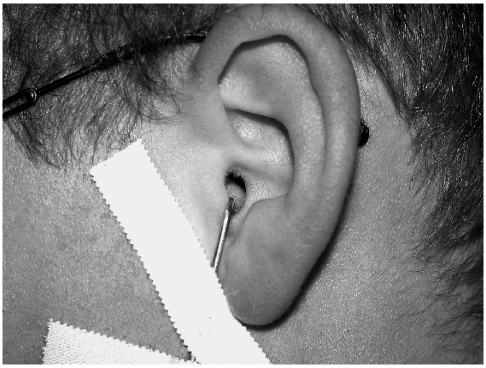

Fig. 16.10. Probe microphone in the ear canal of a test subject

Head-mounted displays

Head-mounted displays are equipped with small LCD or OLED[86] (or other) displays which create stereoscopic images for the two eyes. The resolution is rather small compared with full screen CAVE-like solutions, and the main disadvantage is the limited view angle, so that even one's

[86] liquid crystal device; organic light emitting diode

own hands and fingers must be rendered. The acoustic path is usually covered by integrated headphones. Due to the discomfort involved in wearing the device, immersion is affected. Advantages, however, can be seen in the feature of easily added components for haptic or tactile stimuli which do not interfere with vision and sound. All aspects of headphone technology hold for sound reproduction.

16.2 Loudspeaker systems

Spatial sound fields can be created with loudspeakers by using one of two general concepts. One can try to reproduce head-related signals, taking advantage of the fact that the hearing sensation depends only on the two input signals to the eardrums. Also, loudspeakers arranged around a listening point ("sweet spot") may serve as 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.

Binaural technology is described in Sect. 16.2.4 below. Sound field technology can be defined in various ways. Another basic form of sound source imaging in the horizontal plane is the well-known stereo setup or a surround sound system[87]. These approaches use the psychoacoustic effect of phantom sources. The basis is a multichannel microphone separating the incident field at one listener point (sweet spot) into spatial components.

More accurate for sound field reproduction is the method of wave field synthesis (WFS; see Sect. 16.2.4). Here, we create an approximation to the spatial sound field incident by using a microphone array too, but not at

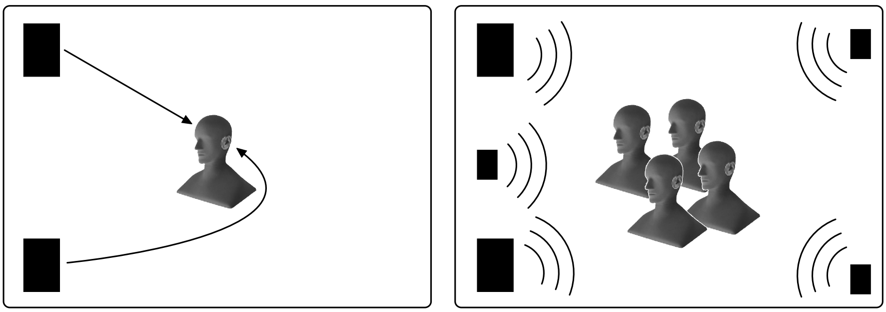

Fig. 16.11. Binaural technology (left) and sound field technology (right)

[87] mostly in use is "Dolby Surround®"

a listener point. Instead, the microphone arrangement is larger and it is located on elementary geometric figures such as straight lines or circles around the listening area. In this way, the artificial sound covers a larger area, and more than one listener can be served with spatial sound.

16.2.1 VBAP surround sound

Starting from a home stereo setup, where the balance control between right and left channels is used for image shifting, additional pairs of loudspeakers can be placed around the sweet spot. The stereo balance controls the relative amplitude between right and left. Thus, the same concept is used for other directions. This perceptual effect stems from the interaural level differences competing with interaural time differences. The overall perception is created from the influences of both cues, which can well be proven by using models of binaural signal processing (Braasch 2005).

This technique was applied in home entertainment in the early 1970s with the name "quadrophony" (with no commercial success, however). Today, some virtual reality systems include surround sound systems called vector-base amplitude panning (VBAP). For ambitious integration of acoustics into VR, however, this method involves too many shortcomings such as inherent implementation of binaural processing mechanisms. Accordingly, VBAP is fine as an effect device, but not as a aplatform for unbiased studies of the human hearing and neural systems.

16.2.2 Ambisonics

Based on work by Gerzon (1976), the so-called B-format or technique of ambisonics was developed. The B-format is a four-channel recording standard that uses a sound field microphone.

Usually the channels represent the front-back (Y), up-down (Z), left-right (X) and mono (W) signals. X, Y and Z signals stem from figure-of-eight microphones in each specific orientation. The W channel is fed from an omnidirectional microphone. The four channels represent a decomposition into spherical harmonics (see Sect. 2.5).

Thus it is possible to achieve an isotropic sound incidence. In a replay situation, a certain geometric approximation of spherical shape by polyhedra or similar arrangements is required. Mixing strategies of the multichannel recording can be adapted to match the recording format to the replay situation.[88] In spatial sound field processing, the decoding process is

[88] usually performed by decoding matrices.

Fig. 16.12. Sound field microphone for recordings in B-format (courtesy of Lab. of Acoustical Imaging and Sound Control, TU Delft)

nothing but a reconstruction by linear combination of zero (W) and first-order (X, Y, Z) spherical harmonics.

Apart from the strict physical approach of spherical harmonics, decoding schemes are in use, which take benefit from the localization and distance perception of human hearing. It must be kept in mind again, however, that virtual sound fields may be used for investigating the hearing system. A generally applicable reproduction system must not introduce any artificial auditory cue which is not part of the simulated sound.

16.2.3 Wave field synthesis

The technology of wave field synthesis, (WFS) is a rigorous approach to wave field reconstruction (Berkhout 1888). Its concept stems from seismic physics and source localization by array techniques. The basic mathematical algorithms are a kind of spatial Fourier transformation between the space and the wave number domains. Thus, complex wave fields are decomposed into elementary waves such as plane, spherical or cylindrical waves, by an appropriate transformation.

In plots of the position along a line versus propagation time, these wave fronts can be studied, either in a static condition or in an animated movie in motion, similar to results from FTD simulation.

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

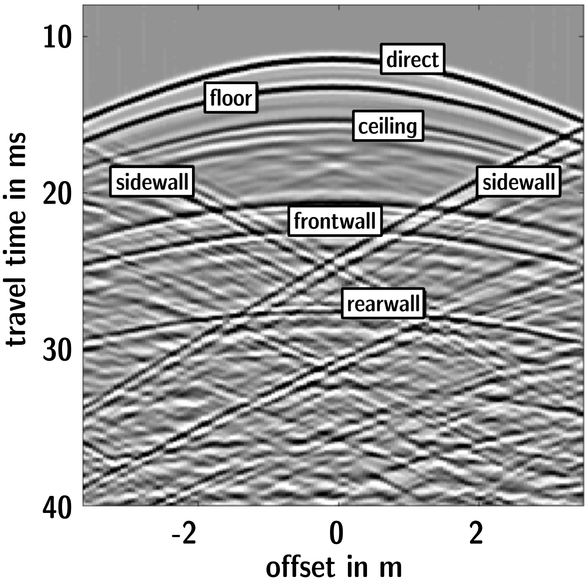

Fig. 16.13. Wave field analysis (courtesy of Lab. of Acoustical Imaging and Sound Control, TU Delft)

elementary sources. In a replay situation, the wave field is reconstructed by sending waves from these points,. This illustrates the step from wave field analysis to synthesis.

As long as the discrete spatial sampling is sufficiently high, any wave field can be reconstructed. Unfortunately, this prerequisite creates severe practical problems. A physically perfect sound field with frequency content up to 10 kHz within a space of $3 \times 3 \times 3 = 30\,\mathrm{m}^3$ required some

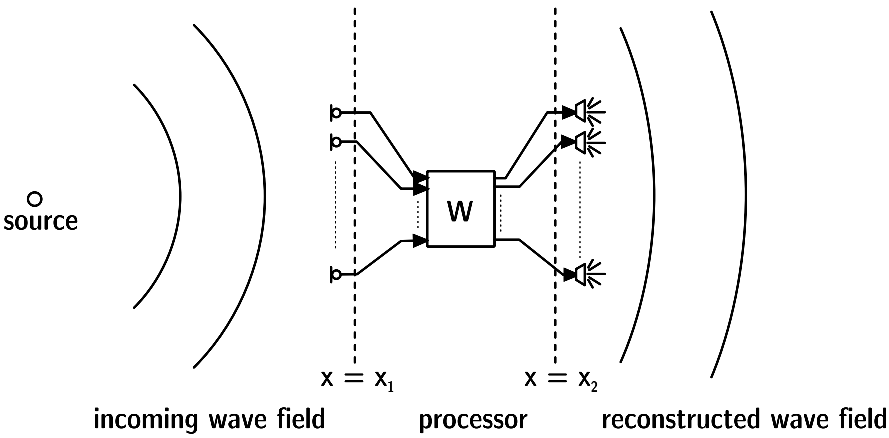

Fig. 16.14. Field reconstruction by loudspeaker arrays (after (Berkhout 1988))

50,000 loudspeakers of 3 cm size. However, in the same way as discussed in several sections of this book, physical constraints are not as relevant as psychoacoustic constraints. The existence of broadband signals and masking in higher frequency bands allows more relaxed conditions for loudspeaker size and distance. And the spatial resolution of human hearing is more sensitive in the horizontal plane. Hence, a two-dimensional array serves well for the most crucial spatial information. With these simplifications, practically installed WFS systems are in use with some 500 loudspeakers, thus, 100 times less effort than theoretically required.

The theoretical background of WFS is the Helmholtz–Kirchhoff integral (see also Eq. (10.5)) which is derived from a general form of the Helmholtz–Huygens integral by excluding the source term:

$$p(r) = \iint \left(g(r|r_0) \frac{\partial p(r_0)}{\partial n} - p(r_0) \frac{\partial g(r|r_0)}{\partial n} \right) \mathrm{d}S_0 , \qquad (16.1)$$

where r denotes the position of the receiver and the set of r_0 denotes the source positions in the loudspeaker array.[89] The integral covers a line or area of sources with the amplitudes $p(r_0)$. The resulting sound pressure at the receiver is thus $p(r)$. The kernel of the integral includes Green's functions of monopole and dipole (pressure gradient) sources. In general, any wave field can be decomposed into the elementary waves from monopoles and dipoles. One consequence of the Helmholtz–Kirchhoff formulation is that sources within the surrounded volume cannot be modelled. As a compromise, solutions for near sources are in development, but these approaches require head-tracking, and this is valid only for one person. For sources from outside arrays, however, any arrangement is theoretically possible, and groups of listeners can be fed with 3-D sound (see Fig. 16.15).

For two-dimensional arrays and reproduction of the sound field in a plane instead of a volume, the functional basis is cylindrical waves. The corresponding Green's functions do not represent the "correct" distance law in 3-D environments. Amplitude correction must, therefore, be applied.

The process of sound recording and mixing is different from usual techniques applied in audio engineering and music production. The spatial information must be integrated in a flexible way that sets no limitations on the way of listening such as position, orientation or movement. Spatial decoding is accordingly transferred in the WFS reproduction system, while each source[90] is recorded separately from other sources in the best possible anechoic situation. Furthermore some control parameters such as position and relative level of the source are encoded.

[89] The source notation by a running index is omitted here.

[90] typically a musical instrument or a voice.

Also the performance environment such as a room is not included in the recording, but is created at the end in WFS by mapping the parameters and signals into an actual, synthetic or augmented spatial situation.

Fig. 16.15. Installation of WFS (courtesy of the Fraunhofer Institute of Digital Media Technology (IDMT), Ilmenau, Germany)

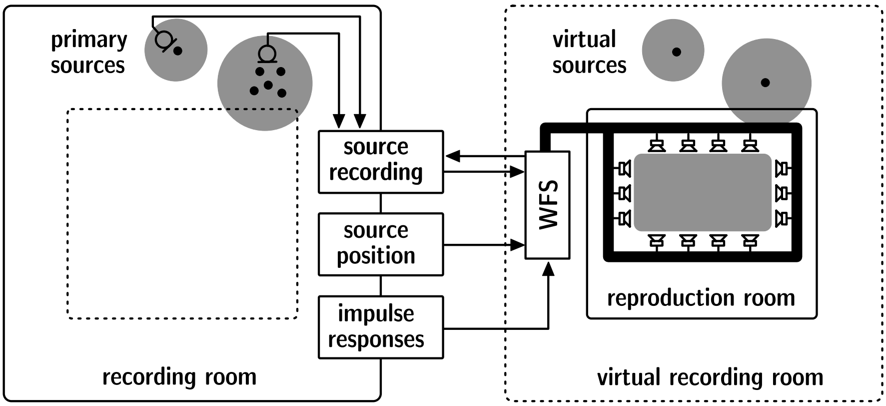

Fig. 16.16. Recording formats for WFS (after (Spors et al. 2004))

Fig. 16.17. CAVE-like environment with WFS (courtesy of the Fraunhofer Institute of Digital Media Technology (IDMT) and Technical University Ilmenau, Engineering Design Group)

For a room acoustical simulation, the room impulse responses related to all sources must be calculated for all reference signals (microphone array) from which the spatial information can be transformed into the wave-functional basis. Since the impulse responses contain direct sound(s) and reflections, the complete spatial field is reproduced in the complete listening area. The set of room impulse responses can be obtained by any of the methods described in Sect. 11.6, but excluding binaural filters.

16.2.4 Binaural loudspeaker technology

We now go from approximate creation of wave fields in large audience areas to the one-listener solution. The concept is part of binaural technology. When we use loudspeakers, the binaural loudspeaker arrangement should act as a "virtual headphone." This condition is already given in the expression "binaural." The two ears must be fed with binaural signals, as created by binaural synthesis or dummy head recordings. Hence, a left/right stereo setup should serve well. However, as has been common knowledge for many decades, cross talk takes place which interferes with the wanted ear-related signals. The virtual headphone in its function is disturbed by insufficient channel separation.

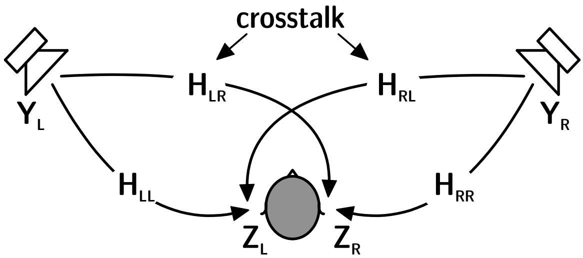

Fig. 16.18. The loudspeaker signals Y_{LL} and Y_{RR} should represent the ear signals Z_{LL} and Z_{RR}. Unwanted cross talk happens over paths H_{LR} and H_{RL}

Without a compensating filter, the aim of localization by binaural synthesis is clearly not successful. In an experiment, test subjects are asked to mark the perceived direction of sound incidence. The diameter of the circles represents the number of responses. Expecting all responses on the diagonal, we observe that the uncertainty of judgment is large, and also we identify front–back confusions near the 90° direction, as indicated by perpendicular groupings of circles.

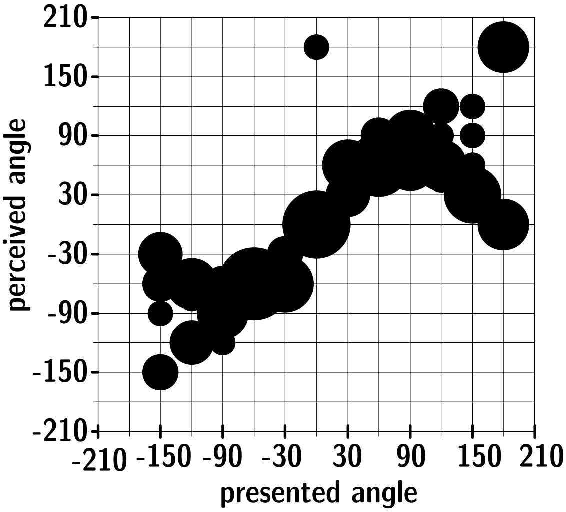

Fig. 16.19. Results of a localization experiment in an anechoic chamber. The diameters of the circles represent the numbers of the perceived direction as responses to the synthesized direction of sound incidence

Cross talk cancellation (CTC)

The required channel separations can be achieved by signal processing. Based on a proposal by Bauer (1963), the first filter structure was formulated by Atal and Schroeder (1963) as a kind of subtraction filter. To discuss the process implemented in this filter, we define the so-called ipsilateral ear oriented toward the direction of sound incidence and the contralateral ear in the shadow region. Cross talk happens to the contralateral ear. For example, the right ear for the right loudspeaker is the ipsilateral ear, and the left ear is contralateral. The cross talk signal component from the right loudspeaker into the left ear can be compensated for by radiating a respective signal from the left loudspeaker and vice versa.

The compensation signal, however, also creates cross talk. We must continue with the same strategy for some interactions. These iterations are stable because over the iterations, the contralateral signals have less and less energy. Five iterative steps are usually sufficient to achieve channel separation of 20 to 25 dB. Simple solutions (low-order iteration), by the way, are implemented in portable radios to create a "wide-function," however suffering from loss in bass performance. The weak bass is an indicator of the quality of the iterations.

Møller (1992) proposed a closed solution for the iterative process. This is possible due to the inherent mathematical structure of a geometric series; see also (Schmitz 1994). The validity of the approach and the stability of the filter, however, depend on the level difference between ipsi- and contralateral transfer paths. When the denominators become small, the filter produces ringing effects due to large narrowband peaks in the transfer

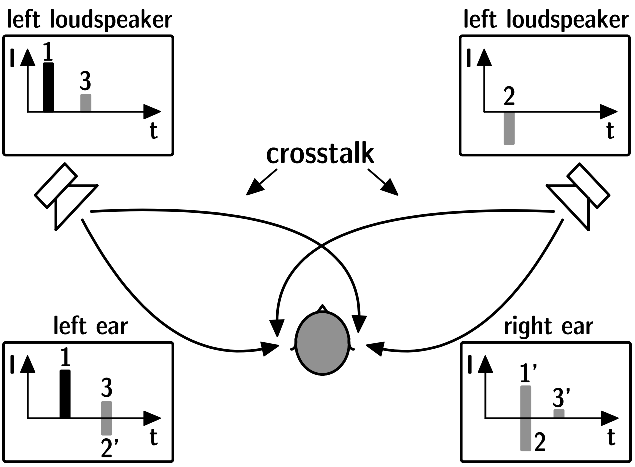

Fig. 16.20. Iterative cross talk compensation

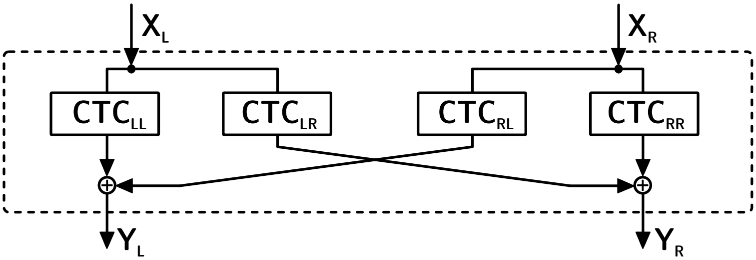

Fig. 16.21. Cross talk filter design

function. The angular zone which provides a natural damping of the contralateral paths is thus limited and it depends on the geometric situation between the loudspeaker pair and the listener.

Notation according to Fig. 16.18 yields

$$Y_L = \frac{1}{L} \cdot \left[\frac{H_{RR}}{H_{LL} \cdot H_{RR} - H_{LR} \cdot H_{RL}} \cdot X_L - \frac{H_{RL}}{H_{LL} \cdot H_{RR} - H_{LR} \cdot H_{RL}} \cdot X_R \right] \quad (16.2)$$

and Y_R accordingly.

The ear signals X_L and X_R are constructed from four filters as represented in the quotients in Eq. (16.2) and the corresponding equation for the right channel.

Stereo dipole

This technique is similar to the cross talk cancellation system. The loudspeaker arrangement, however, is a left/right dipole source placed in the front of the listener with at most a 10° angle span. Enhanced cross talk damping is achieved by adjusting the null axis of the dipole into the front plane. Both ears receive signals from only one lobe of the dipole. The result of the stereo dipole source is similar to a theoretical monopole and dipole combination.

The stereo dipole system performs very well in symmetric geometry between loudspeaker and listener, and it performs best at close distances, such as in PC displays. More information can be found in (Kirkeby et al. 1998).

Dynamic cross talk cancellation

The compensating method of cross talk cancellation and of the stereo dipole are based on interferences and shading effects. They are thus sensitive to phase errors caused by geometric variations, particularly in translational or rotational right/left movement. If the compensating filters are calibrated

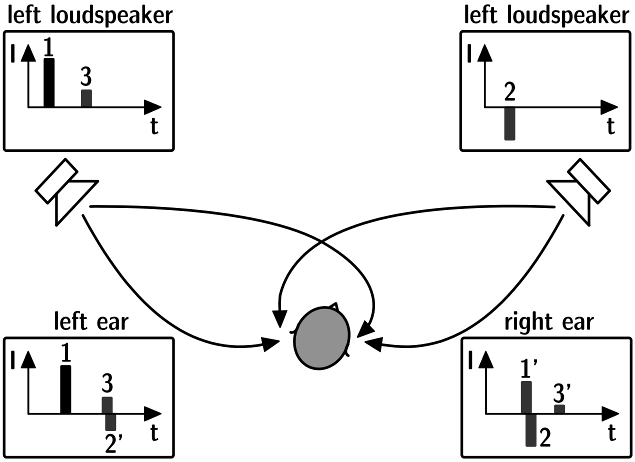

Fig. 16.22. Mismatch by head movement

for one specific position (sweet spot), offside positions of the order of magnitude of 2 cm may lead to audible effects of distorted timbre and localization.

The possible range for these movements is larger for the stereo dipole. For interactivity and free movement by the listener of the order of magnitude of metres, however, all static techniques fail.

Cross talk compensation need not be installed in a symmetric arrangement of loudspeakers and listener. Hence, there is no need to fix the sweet spot at one arbitrary location. The sweet spot can also be fixed at the listener's position, even when the listener is moving. The solution for getting this performance is adaptive filtering (Gardner 1997).

First, the position of the head must be determined by using a head tracker.

Four-speaker system

Dynamic cross talk cancellation requires a valid HRTF filter set for each position. Further research by (Gardner 1997) and (Lentz and Behler 2004) has shown that a dynamic CTC using two speakers is stable only within the angle spanned by the loudspeakers. To provide free rotation for the listener, the two-speaker solution is expanded to four channels, two of which run at the same time.

Between the areas, a cross-fade algorithm is implemented using two parallel working CTC systems, which are then partly superimposed. In this way, the current binaural audio signal is filtered with the correct cross talk

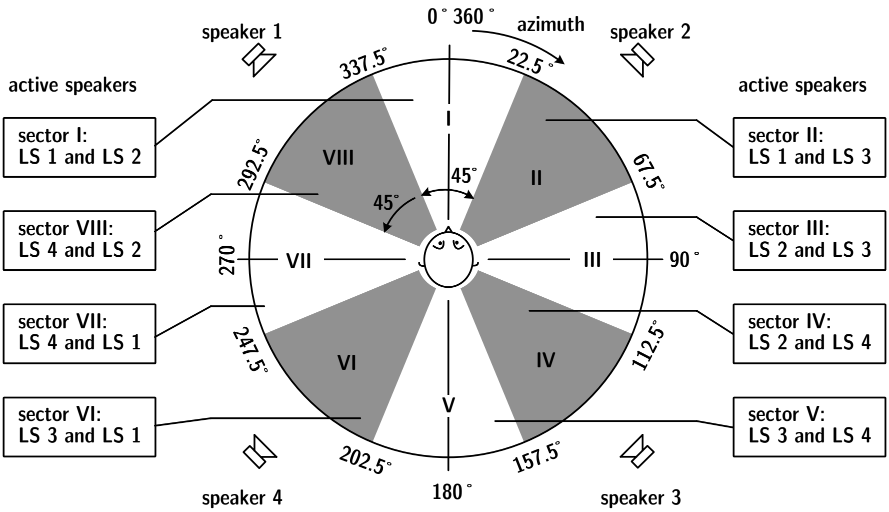

Fig. 16.23. Getting rid of the limitations in listener rotation: Sectors of activation of loudspeaker pairs in a four-speaker dynamic CTC system

cancelling filter for the listener's present position. More details on the implementation and evaluation are found in Lentz et al. (2007).

One example of an integrated system, including room acoustic rendering based on physically consistent models and headphone-free 3-D sound reproduction is illustrated in Fig. 16.25. It consists of auralization software with high precision in the early part (accurate on scales in microseconds) of the impulse response and smooth updates of the late decay (accurateo in scales of milliseconds) which take place only after the listener has moved by some metres.

16.3 VR technology and integrated VR systems

Virtual reality (VR) 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 3-D vision. 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, their acceptance by the user and user feedback.

User interfaces are well-established for several kinds of applications, in computer operating systems, in specific application software or in computer games. Any kind of control display for machinery, vehicles or any

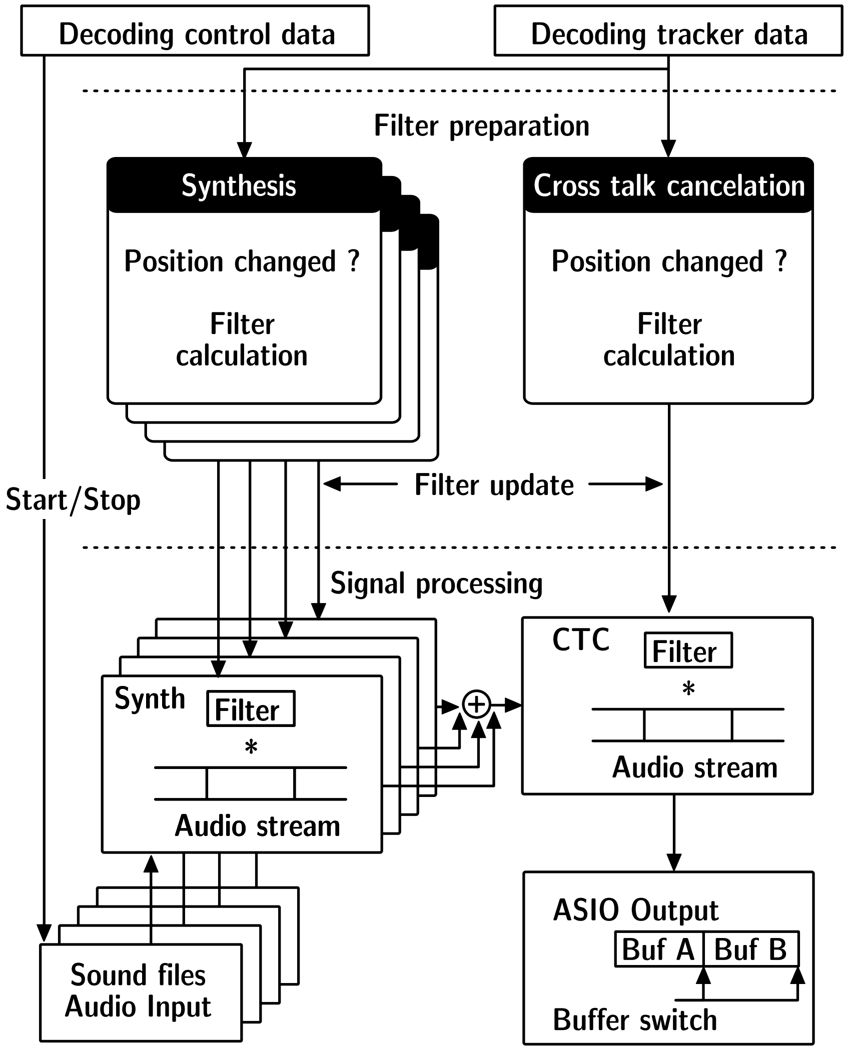

Fig. 16.24. Block diagram of dynamic CTC processing with head-tracker data

other technical object or system can be interpreted as a human–machine interface. Acoustic information may replace visual information displayed for system control or monitoring (Begault 1994).

A so-called "auditory display" such as an "icon" has the purpose of creating attention and information for the visual sense. Therefore acoustic icons are also called "earcons." When presented spatially, a large variety of situations and applications are available, either related to the actual location and direction of an event, or augmented to the specific spatial situation. Examples can be found in cockpits (see, for example, (Bronkhorst

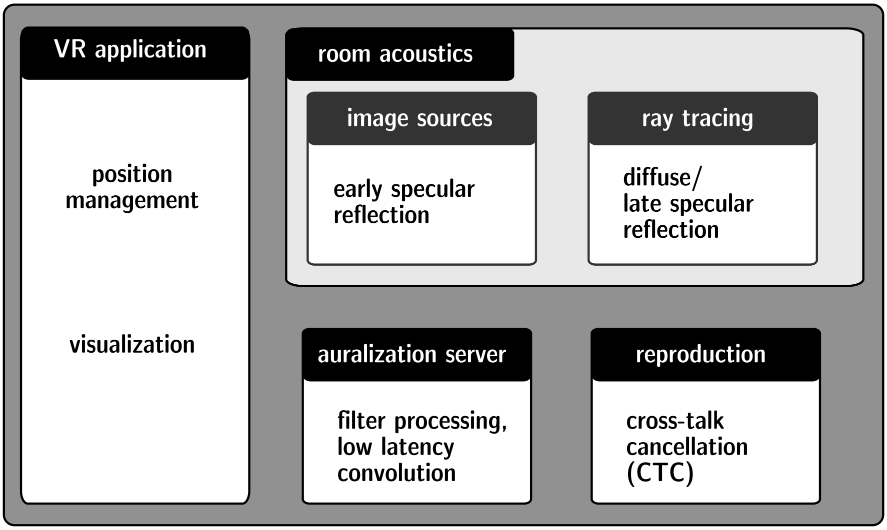

Fig. 16.25. VR System with integrated binaural room acoustics

et al. 1996)) or systems for orientation and navigation for blind persons. Earcons can hence represent information of function.

In the discipline of visual VR, the recent development of graphic processors enabled simulation and reproduction of quite complex scenarios with a high degree of realism. Furthermore, the state of technology for VR subsystems such as motion trackers and projections units has gained very high quality at reasonable cost. As soon as application software is available, it is not unexpected that VR technology will be transported from highly sophisticated laboratory setups 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, such as in fluid dynamics or molecular physics.

The rapid development of CPU and memory technology will also influence VR. PCs are being equipped with multicore technology and parallel concepts. Ray tracing, image source audibility check or other software components can well be implemented on parallel CPU architecture. Apart from constraints due to data traffic, the gain in simulation speed is almost given by the number of processors involved.

This computational power is required for physical interactive models of haptic, acoustic and visual feedback. In such an environment, the scenes are not simple effects,[91] but are calculated physical reactions in real time.

[91] such as in a movie, created once and for one specific intention of effect or feedback.

The main conceptual difference is that no database of a number of fixed reactions is used, such as in computer games, but all reactions are created based on physical model equations. The variety of scenes, the plausibility of the virtual environment is thus much larger, so that there is no practical limit on memorized action and reaction. The user can operate more intuitively with the virtual scene. Interesting studies can also be performed with modified physical parameters such as objects falling and bouncing in gravitation on the Moon or Mars or alternative behaviour of the physical properties of objects.

To obtain the best possible immersion and presence, the human senses must be addressed in virtual environments in a most realistic way. The user of a VR system should act as freely and naturally as possible. For better immersion, spatial acoustics have necessarily been included, and this is required more and more in large projection systems such as the L-bench or CAVE-like environment. Head-mounted displays were quite common in the 1990s, but today there seems to be a tendency toward releasing the user from wearing too much hardware.

The same holds for training in driving and flight simulators, where the projection unit is combined with a large mechanical platform that includes all human–machine interfaces required for the task. The aircraft cabin, for instance, in this respect must be equipped with all components of the virtual environment, and it must reproduce all cues of vision, sound, forces, etc., which represent the actual training situation as much as possible.

From the acoustic point of view, implementations of loudspeaker systems are difficult in combination with rigid video screens. Acoustically transparent screens, on the other hand, do not guarantee sufficient sharpness, contrast and polarization information of the video images. Therefore, compromises are unavoidable. For an acoustic rendering, several methods have been published (systems integrated in WFS or CTC are examples which were already mentioned), for instance, (Huopaniemi et al. 1996; Gröhn et al. 2007). Systems on special hardware are controlled by software plug-ins (\). Spatial acoustic parameters such as in (OpenAL 2006) are useful for programming acoustic sources.[92] This approach is typically used for creating simple acoustic images for computer games or other fields of consumer electronics which don’t have the demand for physical correctness. Other software solutions are given by (Storms 1995; Naef et al. 2002), while (Savioja 1999; Savioja et al. 1999; Hahn et al. 1995) focus on the component of sound reproduction.

[92] but not for realistic spatial scenes in rooms

Fig. 16.26. The CAVE-like environment at RWTH Aachen University

Observing these activities, we can expect further progress in research and development of auralization and virtual reality systems, including physical acoustic models and 3-D sound field reproduction. From high-quality technology for acoustic laboratories, some hardware and software solutions will also inspire inventions of products for the consumer market.