Auralization

Fundamentals of Acoustics, Modelling, Simulation, Algorithms and Acoustic Virtual Reality

Bearbeitet von
Michael Vorländer

1st Edition. Softcover version of original hardcover edition 2008 2011. Taschenbuch. xv, 335 S. Paperback
ISBN 978 3 642 08023 4
Format (B x L): 15,5 x 23,5 cm
Gewicht: 539 g

8 Characterization of sources

The so-called “dry” source signal is generally defined as a source signal free of reverberation and of any other cues introduced by sound transmission. As soon as the relevant transfer functions are known by simulation or measurement and the input signal is recorded (or simulated) properly, the signal transmission path is identified and the output signal can be obtained.

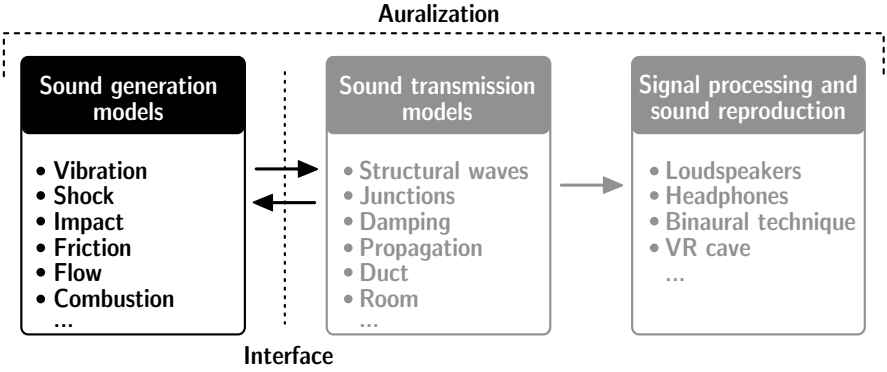

Fig. 8.1. Sound source characterization

In this section we consider how sound generation and sound sources can be recorded and processed.

8.1 Airborne sound sources

Input signals, $s(t)$, are related to specific sources. Typical examples are musical instruments, the human voice and a noise-generating machine. At this point, the question of feedback will be shortly discussed or rather the neglect of feedback will be justified. The radiated sound power of an instrument or voice should be independent of the environment. In airborne auralization problems this prerequisite is mostly fulfilled easily, whereas in structure-borne sound auralization, feedback might occur as a severe problem. The reason is the relation between the inner impedance of the source

and the radiation impedance (Sect. 2.3). As long as the radiation impedance is small compared with the inner impedance of the source mechanism, the radiated volume velocity is invariant to changes in the environment (room, duct, etc.).

Modelling and auralization of a musical instrument, the human voice or loudspeaker is usually based on the sound signal recorded at a specific point (in the direction of "main" radiation or in a symmetry axis). The directional radiation pattern is accounted for by using a directivity database. For musical instruments, the human voice or loudspeakers such data can be found. For noise sources, however, such generally applicable directivity data are typically not available. The directivity must be modelled by assuming specific elementary radiation conditions, or it must be measured in an anechoic situation or in the near field. Microphone arrays on a measurement surface such as arrays used in sound power measurements, are an option for collecting data for source directivity encoding. In this respect, it is also of interest to reconstruct, at least in approximation, the surface vibration of the noise source. Methods of multipole analysis (Ochmann 1990) and acoustic holography (Williams 1999) may yield equivalent source parameters which represent the vibration pattern and the far field radiation directivity.

Source encoding and reconstruction on the basis of several mathematical models is described in detail in (Magalhães and Tenenbaum 2004). Here we focus on some specific examples. As an example of a detailed mathematical model, multipole synthesis is introduced. Another option explained in this book already is the functional basis of spherical harmonics.

8.1.1 Multipole synthesis

The prerequisite of multipole synthesis is that sound fields can be developed into a series of spherical harmonics (Sect. 2.5) or of multipoles (Sect. 2.4). The aim of this procedure is coding of measured or calculated directivity patterns. In simulations. the sets of multipole coefficients are used to reconstruct the original source directivity. This is achieved by linear superposition.

Spherical harmonics strictly are related to polar coordinates and a source origin in one point, whereas multipoles can be distributed in space, so that more degrees of freedom in sound field approximation of arbitrary sources can be used as synthesis parameters.

Due to the fact that multipoles are special solutions of spherical harmonics, the principle will yield identical results for sources of spherical geometry. The decomposition of the sound radiation problem into spherical functions allows the a posteriori reconstruction of the directivity. The

reconstructed results are best matching the original source directivity for sources of spherical geometry. Furthermore, it can be used as a general basis for other series of elementary radiators such as monopole and multipole sources and combinations of these.

For radial symmetry, the spherical Hankel function of zero order is equivalent to the concept of a monopole. Now, higher orders of m include higher orders of sources, for instance, monopoles and dipoles in the three Cartesian orientations, x, y, z, are given by

$$\begin{aligned} \underline{p}_{\mathrm{o}}(k,\vec{r}) &= \frac{\rho_0 c k^2 \hat{Q}}{4\pi}\psi_{00}^{1}(k,\vec{r}) \qquad \underline{p}_{8,\mathrm{x}}(k,\vec{r}) = \frac{\rho_0 c k^3 d\hat{Q}}{4\pi}\psi_{11}^{1}(k,\vec{r}) \\ \underline{p}_{8,\mathrm{y}}(k,\vec{r}) &= \frac{\rho_0 c k^3 d\hat{Q}}{4\pi}\psi_{01}^{1}(k,\vec{r}) \quad \underline{p}_{8,\mathrm{z}}(k,\vec{r}) = \frac{\rho_0 c k^3 d\hat{Q}}{4\pi}\psi_{10}^{1}(k,\vec{r}) \end{aligned}, \tag{8.1}$$

in direction $\vec{r} = (r,\vartheta,\varphi)^T$. Quadrupoles exist as lines or square arrays. The general expansion into spherical harmonics is

$$\begin{aligned} \underline{p}_{\mathrm{xy}}(k,\vec{r}) &= \frac{\rho_0 c k^4 d^2 \hat{Q}}{24\pi}\psi_{22}^{-1}(k,\vec{r}) \\ \underline{p}_{\mathrm{xz}}(k,\vec{r}) &= \frac{\rho_0 c k^4 d^2 \hat{Q}}{24\pi}\psi_{21}^{1}(k,\vec{r}) \\ \underline{p}_{\mathrm{yz}}(k,\vec{r}) &= \frac{\rho_0 c k^4 d^2 \hat{Q}}{24\pi}\psi_{21}^{-1}(k,\vec{r}) \\ \underline{p}_{\mathrm{xx}}(k,\vec{r}) &= \frac{\rho_0 c k^4 d^2 \hat{Q}}{24\pi}\left(\psi_{22}^{1}(k,\vec{r}) - 2\psi_{20}^{1}(k,\vec{r}) - 2\psi_{00}^{1}(k,\vec{r})\right) \\ \underline{p}_{\mathrm{yy}}(k,\vec{r}) &= \frac{\rho_0 c k^4 d^2 \hat{Q}}{24\pi}\left(-\psi_{22}^{1}(k,\vec{r}) - 2\psi_{20}^{1}(k,\vec{r}) - 2\psi_{00}^{1}(k,\vec{r})\right) \\ \underline{p}_{\mathrm{zz}}(k,\vec{r}) &= \frac{\rho_0 c k^4 d^2 \hat{Q}}{24\pi}\left(4\psi_{20}^{1}(k,\vec{r}) - 2\psi_{00}^{1}(k,\vec{r})\right) \end{aligned} \tag{8.2}$$

The total pressure field is formed by superposition of spherical wave functions with proper relative amplitudes. For a complete set of field synthesis coefficients, the velocity is of interest, too. Thus not only the continuous pressure but also its gradient can be optimized for approximating and interpolating arbitrary source data. We obtain the velocity addressed to spherical wave functions by using the pressure gradient

$$\vec{v} = \frac{j}{\rho_0 \omega}\nabla p\,, \tag{8.3}$$

which yields

$$\nabla \psi_{mn}^{1} = \begin{pmatrix} k(2m+1)^{-1}\left[mh_{m-1}^{2}(kr) - (m+1)h_{m+1}^{2}(kr)\right]P_m^n(\cos\vartheta)\cos(n\varphi) \\ r^{-1}h_m^2(kr)\left[(m-n+1)P_{m+1}^n(\cos\vartheta) - (m+1)\cos\vartheta P_m^n(\cos\vartheta)\right]\cos(n\varphi) \\ -n(r\sin\vartheta)^{-1}h_m^2(kr)P_m^n(\cos\vartheta)\sin(n\varphi) \end{pmatrix} \quad 8.4)$$

and

$$\nabla \psi_{mn}^{-1} = \begin{pmatrix} k(2m+1)^{-1}\left[mh_{m-1}^{2}(kr) - (m+1)h_{m+1}^{2}(kr)\right]P_m^n(\cos\vartheta)\sin(n\varphi) \\ r^{-1}h_m^2(kr)\left[(m-n+1)P_{m+1}^n(\cos\vartheta) - (m+1)\cos\vartheta P_m^n(\cos\vartheta)\right]\sin(n\varphi) \\ -n(r\sin\vartheta)^{-1}h_m^2(kr)P_m^n(\cos\vartheta)\cos(n\varphi) \end{pmatrix} \quad (8.5)$$

As mentioned above, straightforward expansion of the sound field in spherical wave functions can produce severe problems of poor convergence, if the source is nonspherical symmetry. Multipoles can be placed at various positions. With an expansion of the sound pressure measured at some reference points into coefficients of a set of multipoles, the field can be reconstructed to achieve an approximation of the reference sound pressures and, furthermore, interpolated sound pressures at other positions. The same strategy, by the way, can also be applied with regard to the particle velocity. Guidelines for the number and spatial discretization of the reference points are available (Ochmann 1990).

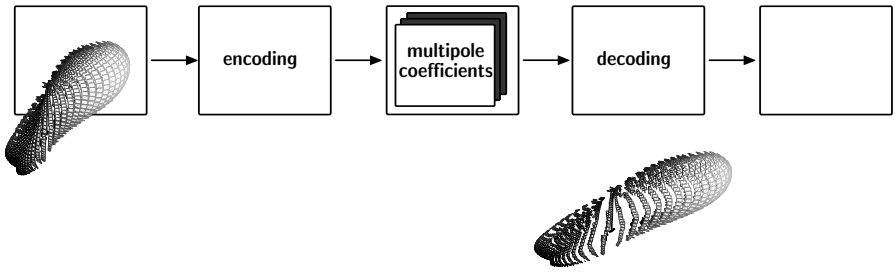

Fig. 8.2. Source directivity encoding

8.1.2 Musical instruments

Signals from musical instruments for auralization can be recorded in anechoic environment (anechoic chamber or, at least a highly absorbing room such as a recording studio) (Giron 1996). Recording must be done in the far field and sufficiently many microphone positions must be used to cover the directional characteristics properly. It must further be ensured that the directional characteristics are constant for all signal components. This fact seems to be no problem for brass instruments since the radiation is dominated from the horn opening (which remains unchanged while playing). In contrast, woodwind instruments have a fluctuating radiation pattern

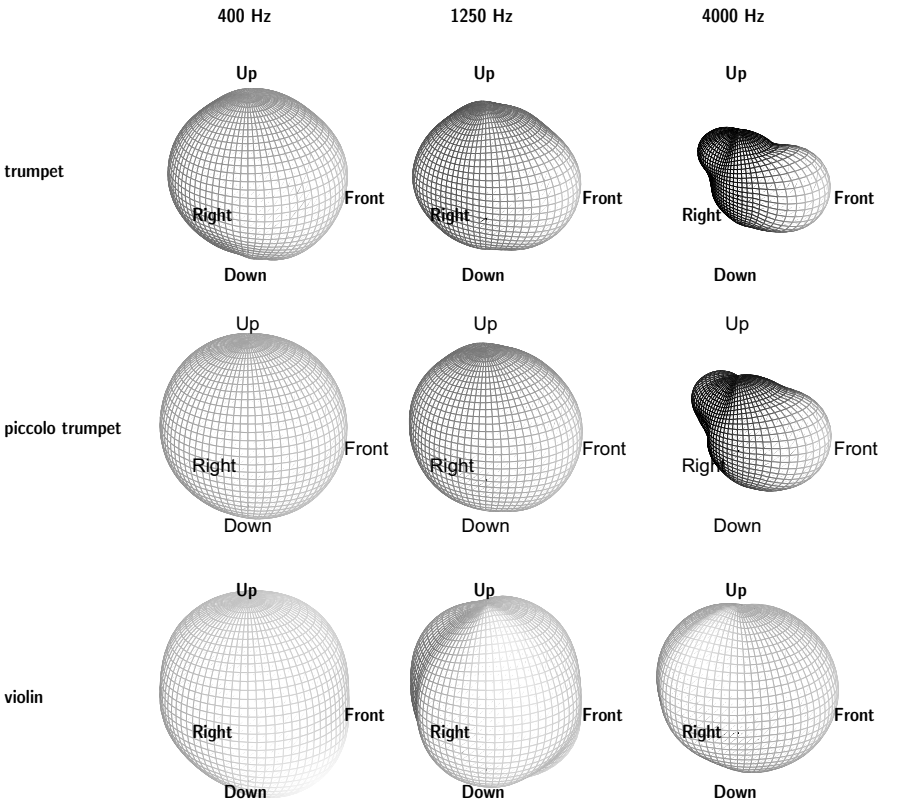

Fig. 8.3. Examples of directional characteristics ("balloons") of musical instruments

because the valves are opened and closed while playing the instrument.[23] Thus, we have to face the problem that not just one directional pattern is valid but the pattern depends on the signal frequencies (tones played). Multichannel recording is one way to overcome this problem (Otondo and Rindel 2005).

Another aspect is the floor reflection. In hemianechoic rooms, the floor reflection will be included in the recording, such as in the actual performance on a stage. In this case, the floor reflection must not be included in the simulation as well.

The most comprehensive collection of directional characteristics of musical instruments was published by (Meyer 1995).

[23] Note that the valves form a line array of volume sources.

8.1.3 Singing voice

The radiation pattern of the human voice in talking or singing is not constant either since the mouth opening depends on the text spoken or sung. A method for recording of singers' directivities was developed by (Kob 2002) who aimed at an artificial singer representing a human singer. For this purpose, singers were recorded in an anechoic environment with a two-channel technique. One channel was recording the signal sung (glissando over at least one octave) in the far field at 2 m distance while the other channel was serving as reference near the mouth. With proper equalization and normalization to the frontal incidence the directional pattern is

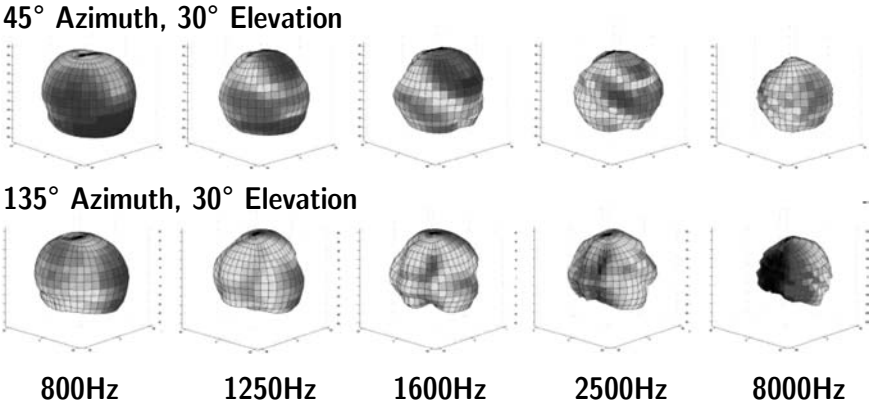

Fig. 8.4. Directivity of the singing voice at two view angles (after (Kob 2002))

Fig. 8.5. Artificial head for the singing voice (after (Behler and Müller 2000; Kob 2002))

obtained after placing the singer on a large turntable and repeating the procedure in angular steps.

8.1.4 Speaking voice

Speech sources have to be separated from singing voice sources since the mouth opening is different. For speech auralization (and speech excitation in measurements), data from an artificial head can be used. These data are found in telecommunication standards (ITU p. 58), and artificial "talking" heads are available.

8.1.5 Anechoic recordings

Single instruments

Anechoic recordings, as described above, are made by using a set of microphones around the source. The reference condition is typically the frontal direction. More or less radiation in the specific directions is taken into account by the source directivity. This approach is appropriate when the source radiation pattern is independent of the signal (music) played.

For some instruments, however, such as woodwind and string instruments, the directivity changes dynamically. This created the necessity for multichannel recording of single instruments (Rindel et al. 2004).

These recorded sounds contain the source signal as well as the directivity pattern, provided the channels are calibrated. In the reproduction situation, the channels must be treated separately and independently for the

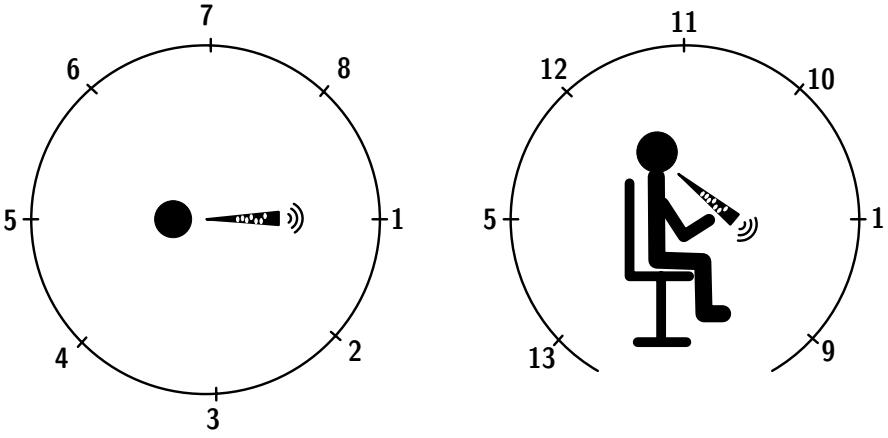

Fig. 8.6. Multichannel source recording (after (Otondo and Rindel 2005))

specific direction. This is easy for room simulation (Sect. 11.1) where the simulated source can be driven in independent angular segments and the channel-specific results superposed at the receiver point.

For a general source characterization applicable to loudspeaker reproduction, the situation is more difficult. The channels now must be mapped to a kind of omnidirectional loudspeaker with an adjustable directional characteristic. If the characteristic is not constant but depends on the music played, such as in the example of woodwind instruments, the directional pattern must be controlled by an adaptive process. This process requires a set of parameters. One option in this respect is a set of multipole coefficients.

Orchestra recordings

At present, some recordings of anechoic music and speech are available (Denon 1995; Hansen and Munch 1991; Freiheit 2005; Vigeant et al. 2007). These recordings were made with orchestras and choirs in anechoic chambers or in near field conditions with the least possible cross talk. Other signals must be created in each case specifically. The "cleanest" solution (but the most tedious one) is surely to record an ensemble by the instruments one-by-one, while replaying the other voices by insert earphone (Vigeant et al. 2007).

The recordings, mostly available on CD, contain the music material listed in the tables below. The first project to be mentioned here goes back to the 1970s. These first anechoic recordings were used for studies on psychoacoustic evaluation of room acoustic field in several European concert halls (Gottlob 1973; Siebrasse 1973).[24]

Later, about the mid-1980s, it was aimed at anechoic recordings for more general use, such as those produced in a concert hall in Osaka, Japan. The concert hall stage was modified to obtain approximate anechoic conditions. Due to relatively near microphone positions to pick up the instruments and instrument groups (clearly within reverberation distance), the direct field could be recorded with only very little influence of the hall's reverberation at a very low level.

[24] For these studies, Mozart's Jupiter symphony was played by the BBC orchestra in an anechoic chamber. Originally it was not intended for use in computer auralization. The recording was broadcast in European concert halls, recorded binaurally by using dummy heads and then replayed for listening tests on the psychoacoustic dimensions of hearing in rooms.

Table 8.1. Denon "Orchestral Music Recording", Osaka Symphonic Orchestra (Denon 1995), 1987

Excerpts from	
Mozart	Overture Le Nozze di Figaro
Mendelssohn	Symphony no. 3 "Scottish"
Bizet	Menuet L'Arlèsienne
J. and J. Strauss	Pizzicato-Polka
Glinka	Overture Ruslan and Lyudmila
Verdi	Prelude La Traviata
Bruckner	Symphony no. 4 "Romantic"
Debussy	Prélude à l' Après-Midi d'un Faune

In the years 1987–1992 in Denmark, the "Archimedes" project was focused on subjective effects of loudspeakers in room acoustics. Listening rooms were also simulated so that studies could be concentrated on the balance between direct loudspeaker sound and room reflections. For this purpose, anechoic recordings were produced:

Table 8.2. B&O "Music for Archimedes" (Hansen and Munch 1991)

Samples	
Speech	English and Danish
Guitar	Tárrega, Bach, Sor, Villa Lobos
Violoncello	Weber, Martini
Drums and percussion	Various
Brass	Haydn, Nielsen, Mason, Purcell

Since choral source material was rarely available until the year 2000, Wenger Corporation initiated a project for recording the St. Olaf College Choir in Northfield, Minnesota in an anechoic chamber.

Table 8.3. Wenger anechoic choral recordings, October 2003 (Freiheit 2005)

Song list	
C.V. Stanford	Beati Quorum Via
R. Thompson	Alleluja
J. Ferguson	Who is this
E. Aguiar	Psalm 150
H. Willam	Kyrie from Missa Brevis no. 4

Fig. 8.7. Multichannel source recording session (after (Vigeant et al. 2007), courtesy of J.H. Rindel, DTU Lyngby)

Finally, a study by (Vigeant et al. 2007) examined which recording technique is superior in realism and source width. Multichannel auralizations compared to single channel auralizations for both solo instruments and a full orchestra were compared in listering tests. The recordings made were based on the multichannel technique proposed by (Otondo and Rindel 2005), but extended for larger ensembles (each instrument playing one-by-one). This set of anechoic sound examples representing music played by solo instruments and orchestra is currently the most advanced approach to channel separation between instruments and directions or radiation.

Table 8.4. Five-channel anechoic recordings of an orchestra of solo instruments using the technique developed in the DoReMi project (Otondo and Rindel 2005), June 2005

First bars (about 1 min 30') from	
Mozart	Symphony No. 40 in G minor, 1st movement
Brahms	Symphony No. 4, 3rd movement

8.2 Structure-borne sound sources

Recordings of structure-borne sources must be made in each specific case under specific conditions. The variety of sources is very large. Numerous recording details have to be considered. At present, there is no standard set of structure-borne sources known. The variety is given by

- dimensionality of motion (velocity, force vectors)
- dependence of force and velocity output on parameters of the transmitting system
- point, line or area contact

However, one standard force source is well known: the tapping machine in building acoustics. The tapping machine should represent the force injected into a floor by a walking person. This force will be transmitted completely into the contact surface, provided the input impedance is very high compared with the impedance of the hammer. This kind of force source represents the ideal case with contact conditions:

- point force,
- high contact impedance of the transmitting system.

8.2.1 General approach

For a more general treatment of sources in combination with structures, the kind of contact and the impedances involved must be considered. In a first (one-dimensional) approach, the open-circuit force and inner impedance of the source must be determined. One solution is to substitute this mechanical problem by its electrical equivalent. The force source, thus, is given by an ideal source with blocked force F_0, and a serial impedance, Z_i or mobility Y_i. Real force sources can be characterized in a first approximation with this simple model. Any source can be described by an open-circuit (blocked) force, F_0, a short-circuit (free) velocity, v_s, and a mechanical impedance, Z_i. or mobility Y_i. The crucial point is that the source parameters can be determined rather easily by using different (known) load impedances, Z_{a1} and Z_{a2}, to determine force and velocity.

For a point contact, the interaction of the source with the structure is then obtained, and the actual force and velocity generated by the source acting on any structure can be calculated. The problem, however, is that point contacts are not a sufficient model in the majority of cases in structure-borne vibration sources; see Sect. 8.2.2.

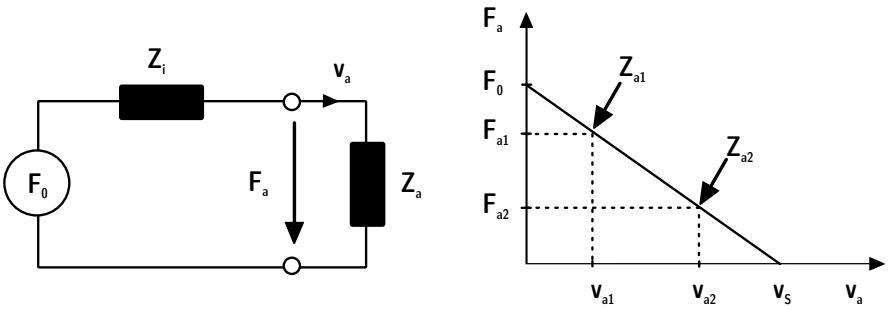

Fig. 8.8. Characterization of real force sources

$$Z_{\mathrm{i}} = \frac{F_{a2} - F_{a1}}{v_{a2} - v_{a1}}, \tag{8.6}$$

$$F_0 = F_{a1} - v_{a1} Z_{\mathrm{i}} \text{ or } F_0 = F_{a2} - v_{a2} Z_{\mathrm{i}}. \tag{8.7}$$

To discuss the interface problem further, we choose the model of power flow. The power of a source injected into the structure is

$$P = \frac{1}{2} |v_s|^2 \frac{\mathrm{Re}\{Y\}}{|Y_i + Y|^2}. \tag{8.8}$$

Two extreme cases are easily defined. As mentioned above and valid for the tapping machine used on heavy construction (force source)

$$P \approx \frac{1}{2} |F_0|^2 \mathrm{Re}\{Y\}, \tag{8.9}$$

and for heavy sources working on lightweight construction,

$$P \approx \frac{1}{2} |v_s|^2 \mathrm{Re}\left\{\frac{1}{Y^*}\right\}. \tag{8.10}$$

If the source is acting on a plate, its mobility is of interest. Plate mobilities depend strongly on the modal distribution. In a very large area, the point mobility of a plate with bending stiffness B and mass surface m'' can be approximated by

$$Y_{\mathrm{plate}} = \frac{1}{8\sqrt{B' m''}}. \tag{8.11}$$

For finite-size plates, the modal pattern must be calculated first (Cremer and Heckl 1973) to obtain the point mobility (see also Sect. 5.2.1). One important aspect is that the modal pattern is not independent of the point of contact.

8.2.2 3-D force sources

The principle described in the section above can also be extended to more complex structures. The complexity, however, may become very much greater in dynamic interaction of distributed sources or line area contacts. Furthermore, various wave types will interact in a complicated way. Structure-borne energy will be transmitted through various paths and via various contacts. These contacts may involve several degrees of freedom which illustrates the complexity of interaction. For these problems of multiple contact points, the formulation of the total injected power is still similar to Eq. (8.10), but the force/mobility coupling is expressed in a matrix. Petersson and Plunt (1982) and Gibbs et al. (2007) extend this approach toward practical cases in building structures. The main extension is that a multiple contact is considered (see also (Petersson and Gibbs 2000)):

$$P = \frac{1}{2}\left(\left(v_s\right)^T \left[\left(Y_i\right)^T\right]^{-1} \left(v_s^*\right)\right). \tag{8.12}$$

For multiple point contact, the relative force amplitude, F_j/F_i, and the coupling transfer mobility, Y_{ij}, between the contacts i and j yields the effective point mobility of the ith contact point (Gibbs et al. 2007):

$$Y_i^{\Sigma} = Y_i + \sum_j \frac{F_j}{F_i} Y_{ij} \,. \tag{8.13}$$

Simplified equations may be used for forces of same magnitude and random phases between the point contacts, for example, on plate fields with statistically many modes. In the latter case,

$$\left|Y_i^{\Sigma}\right|^2 = \left|Y_i\right|^2 + \sum \left|Y_{ij}\right|^2 . \tag{8.14}$$

9 Convolution and sound synthesis

If the source signal and the system's transfer function or impulse response are obtained separately, the resulting output signal can be calculated by convolution (Eq. (7.3)). The convolution can be processed in various ways, either directly in the time domain by using FIR filters or by using FFT convolution. In the latter case, however, it should be kept in mind that FFT requires fixed block lengths and is related to periodic signals. Time windows might be required for reducing artefacts from discontinuities. The same holds for framewise convolution in slowly time-variant systems. Also, the technique of convolution or "filtering" (IIR, FIR) is valid for LTI systems exclusively. For time-varying systems, the excitation signal must be processed in frames representing pieces of approximate time invariance. In this case, filters might be adapted while processing and fading must be used to move from frame to frame.

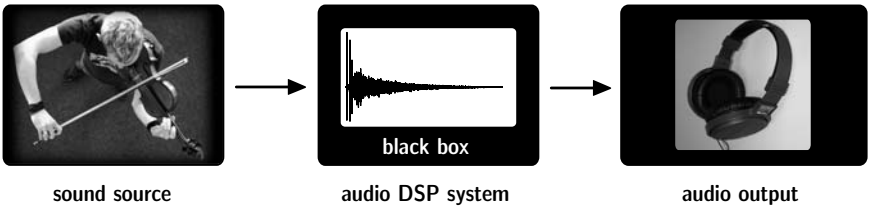

Fig. 9.1. Block signal processing for auralization

9.1 Discrete convolution

Signal processing requires a certain number of calculations steps. Floating point multiplications are the most time-consuming operations, while adding and storing is normally negligible in single-processor programming. Higher computational cost may take place when parallel processing and network communication are involved (see also Chap. 15).

Another special aspect is given by the usage of DSP processors in integer data formats. In this case, the signals must be normalized before processing to avoid data full scale overflow. Although being faster in convolution, they might show disadvantages due to a smaller dynamic range available.

Here, we use the convolution integral (Eq. (7.3)) in discrete form. The input signal, $s(n)$, and the impulse response filter, $h(k)$ are stored in vectors as temporal sequences. Thus the discrete convolution for calculation of the output signal, $g(n)$, reads:

$$s(n) * h(n) = \sum_{k=0}^{N-1} s(k)h(n-k) = g(n) . \tag{9.1}$$

MATLAB® code for discrete convolution

Explicit code:

```
function g = convolution(s,h)

%Initialize signal vectors
s = s(:); %input signal
h = h(:); %impulse response

%Determine length of vector
N = length(s);
L = length(h);
M = N + L - 1;

%Generate matrix for convolution
S = zeros(M,L);
for idx = 1:L
  %copy shifted s vector in matrix
  S( (1:N)+(idx-1), idx ) = s;
end

%Process multiplication for convolution
g = S * h; %output signal
```

or in short, using the built-in Matlab function

```
g = conv(s,h)
```

This process requires $N \cdot L$ floating-point multiplications, with N denoting the length of the input signal and L the length of the impulse response (FIR filter coefficients).

9.2 FFT convolution

Here, we use the convolution integral in the frequency domain. The input data are stored in a vector as temporal sequence and the impulse response and the output signal are considered as vectors of same length. Then the FFT convolution reads as follows:

MATLAB® code for cyclic (FFT) convolution

```
function g = FFTconvolution(s,h)

%Initialize signal vectors
s = s(:); %input signal
h = h(:); %impulse response

%Determine length of vector
N = length(s);
L = length(h);

%append zeros in the end for same length
s = [s; zeros(L-1,1)];
h = [h; zeros(N-1,1)];

%Fourier Transformations
s_spk = fft(s);
h_spk = fft(h);

%element-wise multiplication
g_spk = s_spk .* h_spk;

%Inverse Fourier Transformation
g = ifft(g_spk);
```

9.2.1 Segmented convolution

For continuous signals, however, processing in one block is not possible, unless we have extremely large memory space available. But even if it was possible to store all data, we could not wait for the complete input signal being passed and stored in the output vector.[25] Furthermore, with adaptive filters, the latency would become unacceptably large. Therefore, we have to cut the signal into temporal segments, process them frame by frame, and transfer the results to the output unit sequentially. From frame to frame,

[25] This is obviously not possible in real-time applications, where the output signal must be replayed instantaneously.

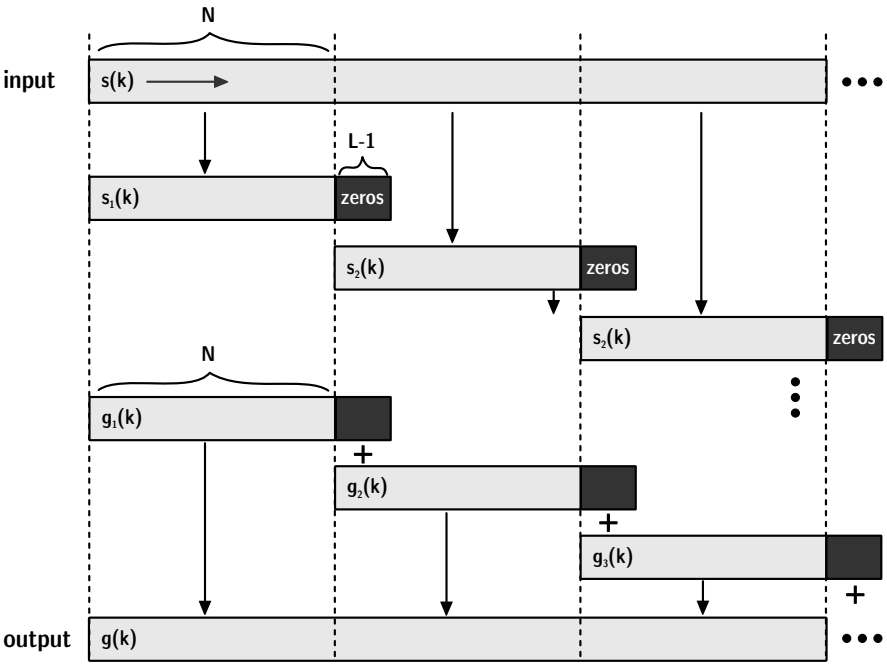

Fig. 9.2. Overlap-add convolution

the filter can also be changed. Fast segmented convolution typically is used for convolving short impulse responses with quasi-infinite signals. The overlap-add algorithm is one example out of various alternatives (Papoulis 1981).

In temporal sequences of Δt, parts of the signal are copied with a length $\Delta t + t_{\text{fade}}$. Δt denotes the reciprocal update rate of the filter (or any chosen filter length) and t_{fade} the duration of a fading interval from frame to frame.

A segment of the input signal is now discussed using one segment between t_1 and $t_2 + t_{\text{fade}} = t_1 + \Delta t + t_{\text{fade}}$. It is extended to the filter length (zero padding). Then the signal segment and the (actual) filter are transformed by FFT. The resulting spectra are multiplied and the product transformed back to time domain by IFFT. The convolution product is then multiplied by a fading function with exponentially weighted slopes to fade in and out.

Finally, the signal is fed into the data buffer of the output signal at the corresponding time index. The next frame is added into the buffer at time t_2, just overlapping with the previous segment in the fading zone. It must be ensured that the signal power in the overlap zone remains constant.

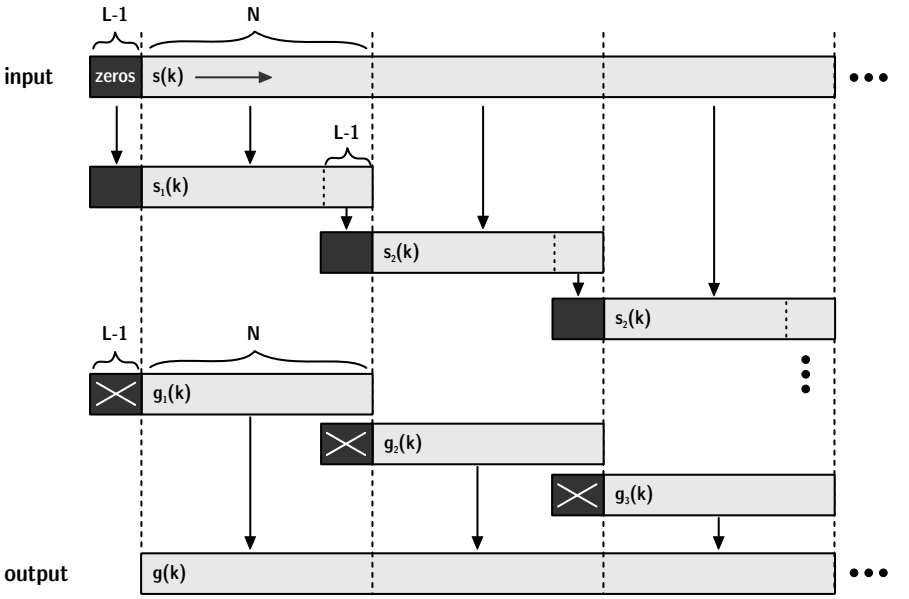

Fig. 9.3. Overlap-save convolution

9.3 Binaural synthesis

With the tools described in the previous chapter, it is a very small step toward the most important component of virtual reality systems: Filtering is used to connect a sound signal to a spatial cue. To achieve user immersion in VR systems, it is indispensable to create spatial sounds that match the visual spatial impression and other multimodal dimensions. Interesting applications of acoustic virtual environments are found already in free-field situations. Examples are outdoor or near-field scenes, indoor scenes with small relevance of reverberation or artificial scenes of augmented reality, where an additional sound is mixed into an existing acoustic stimulus.

The basic task in creating an auralization is to place a sound source into 3D space. Any mono source signal properly characterized and calibrated according to Sect. 8.1, can be processed so that its perceptional cues are amended by a spatial component. A stereo or surround setup is capable of creating an effect of phantom sources (see Sect. 16.2.1) which can produce an appropriate spatial effect. A binaural mixing console can be used for processing headphone signals by using HRTF.

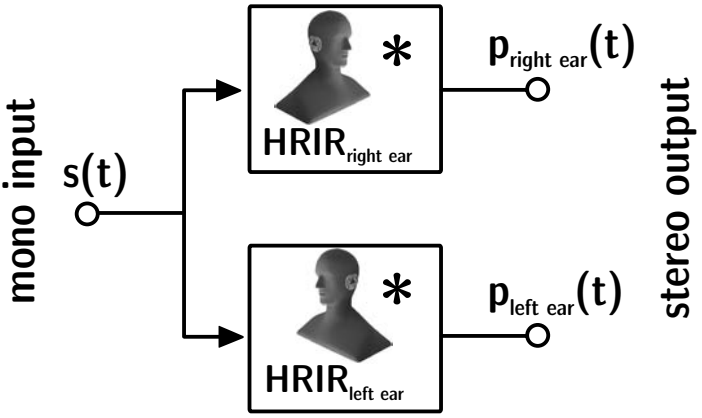

Fig. 9.4. Binaural synthesis

As mentioned in Sect. 6.3, sound localization and spatial hearing can be understood as an effect of the transfer function of the outer ear, the head-related transfer function (HRTF). With a database of the HRTF,[26] any direction of sound incidence can be simulated, when a mono source $s(t)$ is convolved with a pair of head-related impulse responses:

$$\begin{aligned} p_{\text{right ear}}(t) &= s(t) * HRTF_{\text{right ear}} \\ p_{\text{left ear}}(t) &= s(t) * HRTF_{\text{left ear}} \end{aligned}, \tag{9.2}$$

Instead of convolution (FIR filter, segmented convolution etc.), other filter techniques can be used, provided, the binaural cues (ILD, ITD) of HRTF are represented properly. IIR filters, for instance, were also created to simulate the poles and nodes of the HRTF with success (Kistler and Wightman 1992; Huopaniemi et al. 1997; Hammershøi and Møller 2002), see also the excellent overview in (Hammershøi and Møller 2005).

The method explained on this single page seems to be a minor detail, but it is not. Binaural synthesis is the key to many techniques for auralization and virtual reality systems. More information will be given later in Chaps. 14 and 15.

[26] Databases of the HRTF are available on the Internet, see Footnote 14 on page 90.

9.4 Binaural mixing console

Binaural mixing consoles are devices for creating multichannel binaural synthesis. This tool is ideal for auralization of free field environments where a small number of sources or reflections are to be modelled.[27] Modelling of free field propagation requires a source model or recording and analytic calculation of the complex amplitude of the sound pressure signal at the receiving point.

The methods introduced in Chaps. 2 and 3 are well suited. In a most elementary case, this is solved by sound recording at a certain distance and correcting for the propagation law (spherical wave, for instance) to account for the distance of the receiver. Binaural synthesis will take care of the direction of sound incidence. Of course, multiple sources can be simulated and auralized at the same time.

For spherical waves, the free field spectrum corresponding to one source reads:

$$\underline{H}\big|_{\text{left,right}} = \frac{\mathrm{e}^{-j\omega t}}{ct} \cdot \underline{H}_{\text{source}}(\theta,\phi) \cdot \underline{H}_{\text{air}} \cdot HRTF(\vartheta,\varphi)\big|_{\text{left,right}}, \tag{9.3}$$

where $\underline{p}$ denotes the sound pressure spectrum normalized to 1 m distance, t its delay, $j\omega t$ the phase lag due to retardation, $1/(ct)$ the distance law of spherical waves, $\underline{H}_{\text{source}}$ the source directivity (Sect. 2.4) in source coordinates (θ, ϕ), $\underline{H}_{\text{air}}$ the low pass of air attenuation (Sect. 3.6) and HRTF the head-related transfer function (Sect. 6.3.1) of the sound incidence in listener coordinates at a specified orientation (ϑ, φ).

The total free-field spectrum is used as a convolution filter for processing source signals $s(t)$. In case of N sources,[28] the resulting signals are superposed. In the time domain, this procedure leads to

$$p(t)\big|_{\text{left,right}} = \sum_{i=1}^{N} s_i(t) * IFT(\underline{H}_i\big|_{\text{left,right}}), \tag{9.4}$$

where $\underline{H}_i$ denotes the spectra of the propagation function of source i.

Sound transmission or diffraction is modelled by adding attenuation filters (see Sects. 3.4 and 12.2) into Eq. (9.3), whose parameters are related to geometric data or material properties. In the same way, transfer paths of BTPS can be integrated easily, as long as they do not affect spatial cues.

---

[27] This will be no problem up to 32 or 64 channels, but it is not applicable for simulating binaural reflections in room acoustics.

[28] which may also include "image sources"; see Sect. 11.3.

Fig. 9.5. Free-field scene in a simulation of the market place of the City of Aachen (courtesy of the Virtual Reality group of RWTH Aachen University)

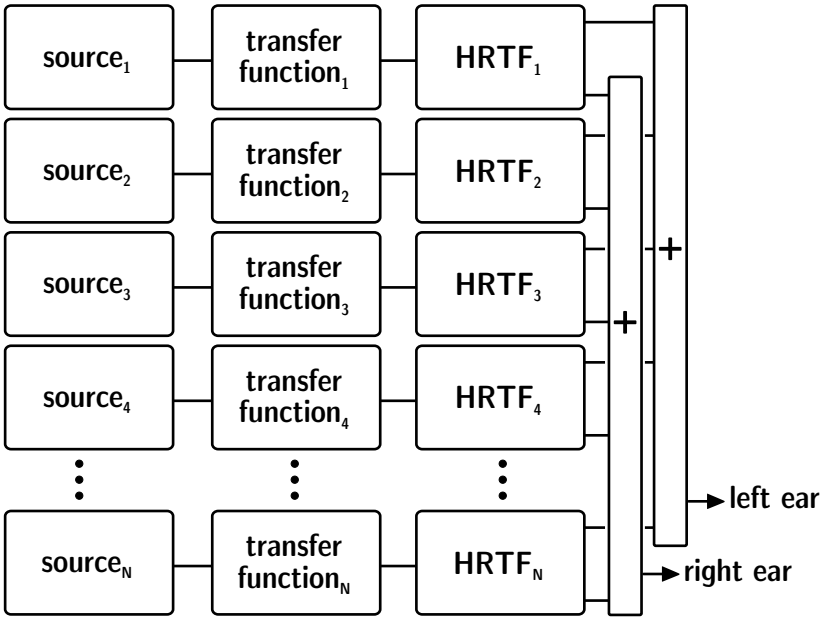

Fig. 9.6. Block diagram of free-field auralization

With the possibility of placing and moving 3-D sounds in space, the most important components of virtual acoustics are available. Typical block lengths of HRTF filters are in the range of 128 or 256 samples.[29] FIR or IIR filters (see Sect. 7.7) can be used, either on DSP platforms or by using PC environments with standardized audio interfaces. Investigations regarding the effects of reduced filter lengths on localization are available (Kulkarni and Colburn 1995, 1998). According to (Hammershøi and Møller 2005) filter lengths can be reduced to 72 taps without audible effects. Filter processing can also be performed in the frequency domain by using overlap-add or overlap-save FFT and multiplication by 64 complex data.

9.5 Spatial resolution of HRTF

The HRTF database should cover all psychoacoustic effects of localization and dynamic changes in localization. Subjective tests of localization showed that humans can discriminate between differences of 1° in azimuth in the frontal direction in the horizontal plane (see Sect. 6.3). In other directions, the angle intervals can be chosen larger. By using interpolation between the HRTF pairs available or by encoding the HRTF in a suitable functional basis (Torres et al. 2004), the size of the database can be reduced further (Minaar et al. 2005).

In applications of auralization with small distances between source and head, standard HRTF is not sufficient (Brungart et al. 1996). Standard HRTF is defined for plane wave incident from specified directions. The prerequisite of independence of the distance is valid for distances > 1 m. Most obviously, distances of sources can be modelled by changing levels, but in distance perception, not only is the relative level evaluated. Spectral changes and, thus, changes of ITD and ILD are also caused by sources approaching the head at constant angles, but from distances closer than 1 m. This kind of data set must be determined with special care and unambiguous definition. A reasonable definition for a reference sound field is the spherical wave. Accordingly, the near-field HRTF is defined in the classical way, but it corresponds to a point source at a near distance.[30]

[29] The sampling rate for audio processing is usually 44.1 kHz.

[30] Proper point sources must be constructed for this kind of measurement. Alternatively, near-field HRTF can be calculated by using BEM.

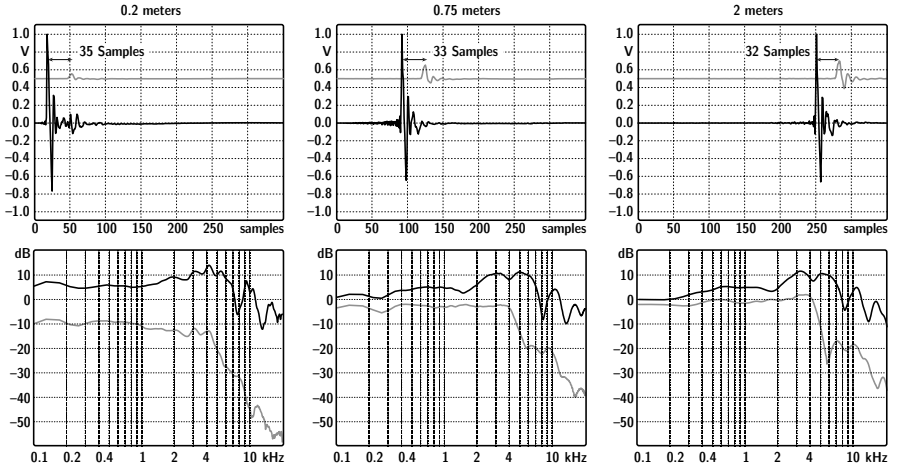

Fig. 9.7. Right to left: From standard to near-field HRTF (after (Lentz 2007))

One can easily calculate the memory space required for such a high spatial resolution of HRTF. Several approaches for data compression are possible. The complexity can be reduced either in space and distance or in FIR filter length (which corresponds to spectral resolution). Another possibility is to choose a rough amplitude scale (low bit scaling).

With the technique of binaural mixing and the opportunity to create any kind of spectral, temporal and spatial sound event, numerous applications where sound is used as carrier of information come into play. One tool in this respect is the so-called “auditory display.”

http://www.springer.com/978-3-540-48829-3

Auralization
Fundamentals of Acoustics, Modelling, Simulation, Algorithms
and Acoustic Virtual Reality
Vorländer, M.
2008, XVI, 335 p. 203 illus., Hardcover
ISBN: 978-3-540-48829-3