11 Simulation of sound in rooms

11.1 General

Computer modelling of room acoustics was proposed during the 1960s by Schroeder et al. (1962) and used in practice by Krokstad et al. (1968) and Schroeder (1973). Although room-acoustic scale model experiments are still a powerful tool today, computer simulations are increasingly taking over the part of scale models in consulting. Commercial software became more user-friendly, more accurate and, last but not least, cheaper than scale models. As soon as the architectural plan is transferred into a computer file and the wall data, source and receiver locations are defined, the sound propagation in the room can be simulated quite fast, and modifications can be tested without large effort.

The algorithms of typical programs are based on geometrical acoustics (Sect. 10.1.3). In geometrical acoustics, the description of the sound field is reduced to energy, transition time and the direction of rays. This approach is correct as long as the dimensions of the room are large compared to wavelengths and if broadband signals are considered. These approximations are valid with sufficient accuracy in large rooms above the cutoff frequency f_c; see Eq. (4.13).

In intercomparison tests (so-called "round-robin tests") (Vorländer 1995; Bork 2000, 2005a, 2005b), the efficiency and the limits of room acoustics computer simulation (see Sect. 11.5) were evaluated. Finally, auralization has been included in room acoustics simulation since the beginning of the 1990s (see Sect. 11.6).

In this chapter, we will discuss the fundamental algorithms, the variations of their implementation in software and their efficiency. It will be shown that pure specular models (image models) are not capable of simulating room sound fields accurately enough, but combinations of models will allow simulation with acceptable plausibility. Hybrid models that can handle specular and diffuse reflections for estimating the late reverberation spectrum are the solution to obtaining impulse responses very near measurement results.

The methods are, at first, used to calculate the room acoustic criteria (T, EDT, D, C, T_S, LF, IACC …); see Sect. 6.4 to get a more specific result than given by the estimate (Sect. 6.4.6).

Two techniques of geometrical acoustics have to be distinguished: "ray tracing" and "image sources." Independent of their software implementation, they represent different physical approaches. To achieve a clear understanding, this point must be stressed, since ray tracing algorithms can well be used to calculate and handle image sources. The main difference between ray tracing (in the classical definition) and image sources is the way energy detection and the internal nature of physical energy propagation are implemented.

Table 11.1. Basic algorithms of room acoustics computer simulations

Algorithm	Category	Energy spreading by distance	Energy detectors
Ray tracing	Stochastic	Stochastic by counting	Volumes
Image sources	Deterministic	Deterministic by distance	Points

11.1.1 CAD room model

To implement the algorithm described in the software, the room geometry, the sound sources and the receivers must be defined as mathematical objects. We start the discussion with the question how to create a CAD model of the test room. The complexity of the model may be very high, particularly when it was exported from software for architectural design. These models include small details such as staircase steps or even smaller geometric features.

From an acoustic point of view, however, this kind of model would not only be too large in memory size and polygonal complexity, it would also be wrong. As will be seen below, the acoustic characterization of surfaces is based on absorption and scattering, and the physical principles of wave reflection and scattering are clearly defined (see Sect. 3.3). Objects or surface corrugations that are not large compared with wavelengths, have to be taken out the CAD model and replaced by flat surfaces with adequate acoustic properties. This holds also for chairs and audience seats, for lamps and for doorknobs. For the purpose of visualization, these elements are essential for a realistic impression. For the "acoustic view," they are invisible or at most visible with some diffuse halo.

As a rule of thumb, the guideline might be used to draw the room surfaces and the interior with a resolution of 0.5 m, representing a wavelength scale corresponding to a frequency of about 700 Hz. Below that frequency, the surfaces should be modelled as flat, specularly reflecting polygons and

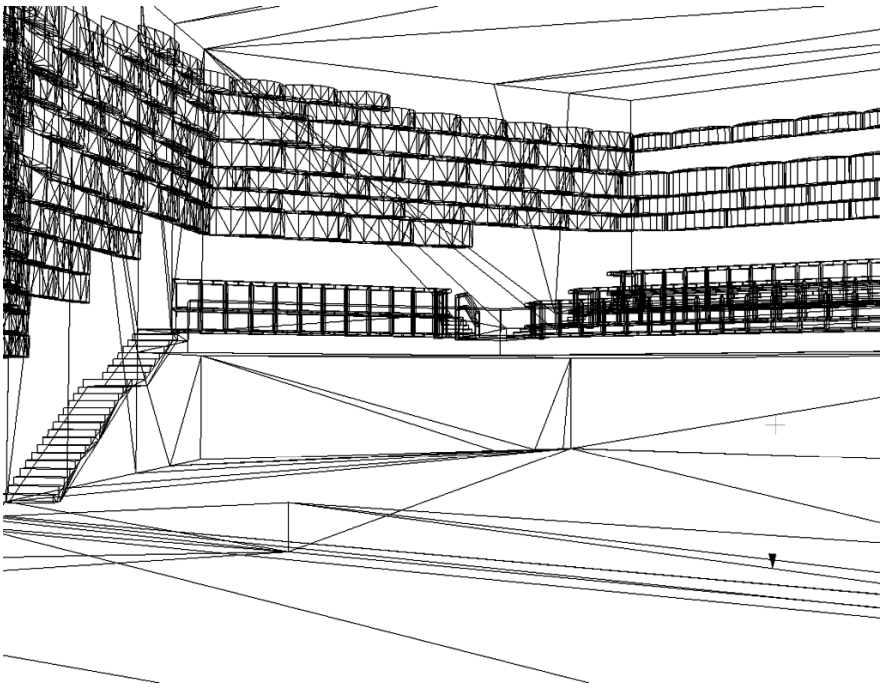

Fig. 11.1. Visual wire-frame model of a multipurpose hall

above that frequency, as partly scattering surfaces or objects. Sound scattering dominates at high frequencies anyway, except for very large walls or the ceiling (> 10 m). The only reason for including a small resolution of details (< 0.5 m) would be a study of a frequency range with wavelengths small compared with 0.5 m, say, $\lambda = 5$ cm. Then, frequencies around 7 kHz and above are discussed. For several reasons, this would be of interest only in very special cases. The main arguments for choosing a range of lower frequencies are as follows:

1. The spectral content excited by natural sources such as the voice of instruments is small above 7 kHz.
2. In broadband signal situations, masking will not allow humans to identify details of low levels at high frequencies.
3. The accuracy of the simulation model is not sufficiently high to guarantee results within an acceptable confidence limit.

Not much research has yet been done on automatic simplification of CAD models from details toward a specific acoustically relevant resolution. Polygonal smoothing by using spatial low-pass filters with a cutoff wave number of 125 m^{-1} (equivalent to a minimum length resolution of 50 cm) could be an interesting option for future automated CAD user interfaces.

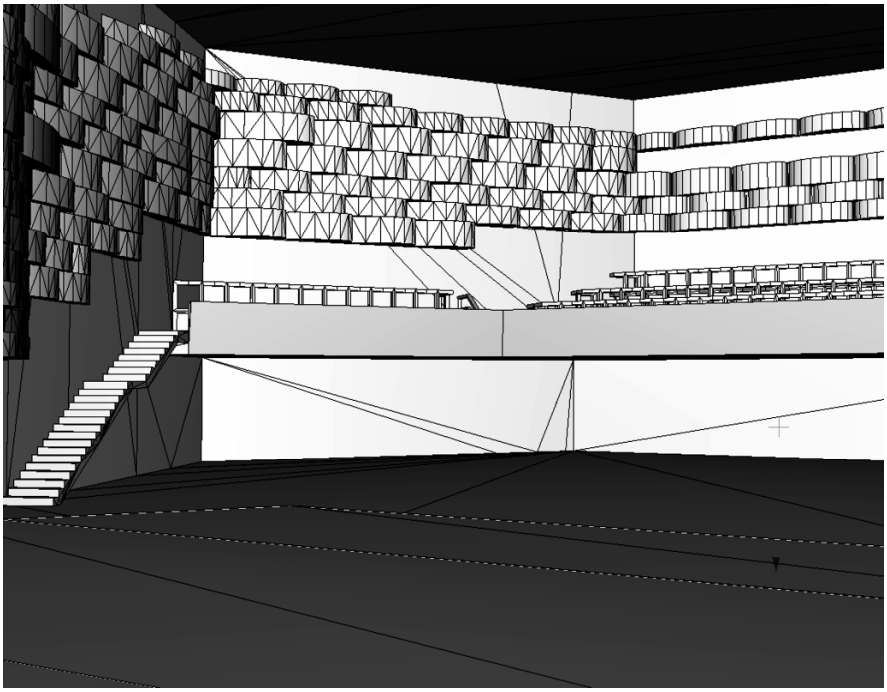

Fig. 11.2. Visual surface model of a multipurpose hall

In the end, the room must be approximated by analytic surfaces, typically by planes. Surfaces of higher order (cylindrical, spherical, parabolic, …) can be used in principle (see below). All surfaces lying in a common plane form a "wall." A wall, thus, need not be simply connected. Walls can also be made up of more than one surface material. One very effective simplification, however, is the requirement for convex wall shapes, as it simplifies and thus accelerates time-consuming tests, such as the point-in-polygon test which is a frequently called function in all kinds of room acoustics simulation algorithms. This rule is no practical constraint at all in modelling a room shape.

All plane surfaces are defined by three points in a previously fixed coordinate system. Preferably points (called "vertices"[41]) of the wall polygon are used. To define the polygon completely, the other vertices of the polygon must be added, too. In practice, geometric uncertainties in adding the fourth, fifth, etc. vertices in the polgon's plane can be allowed, if the new corners are automatically adjusted to match the plane (at least numerically) exactly.

[41] from "vertex"

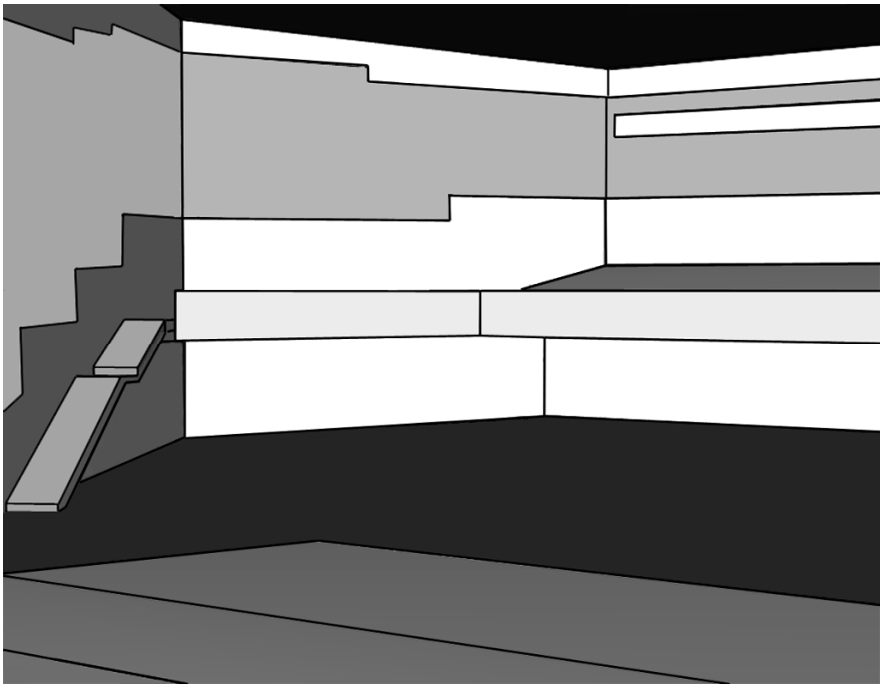

Fig. 11.3. From architectural to acoustic CAD models: Acoustic surface model of a multipurpose hall. Geometric details are replaced by smooth wall polygons

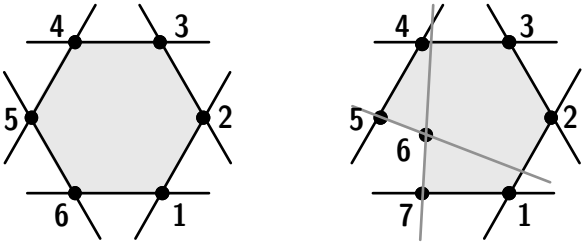

Fig. 11.4. Left: Convex polygon. Right: Concave polygon

Figure 11.5 illustrates the strategy for building a 3-D room from 2-D polygons. The polygon-to-vertex notation, the polgon orientation with reference to the normal vector and their address management must be specified clearly.

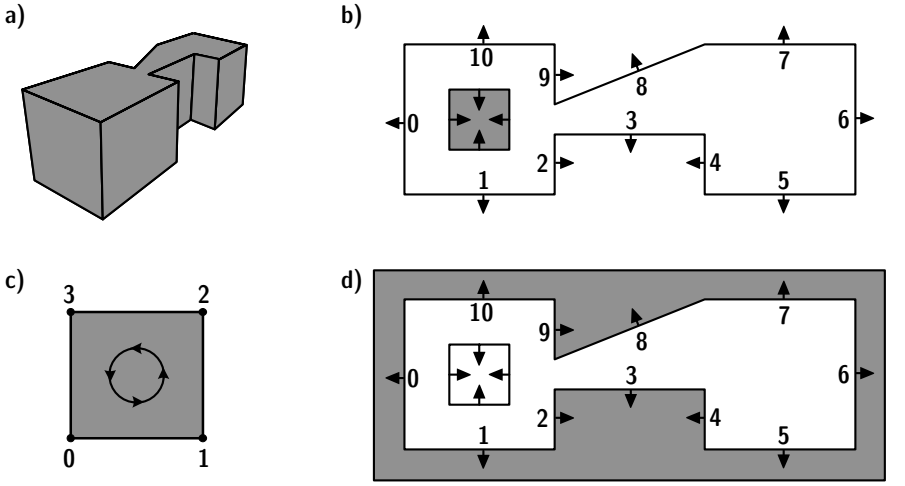

Fig. 11.5. a) Example: A simple room constructed by a set of convex polygons. b) The room's 2-D projection: the green area marks a solid object inside the room. c) Definition of the spanning direction for polygons. d) The room's 2-D projection: the green area marks the outside world (example after (Schröder 2004; Schröder and Lentz 2006)

11.1.2 Absorption coefficients

Absorption coefficients (see Annex) can be found in tables (see Annex), or they can be taken from specific test results. Most absorption coefficients correspond to diffuse sound incidence because they were obtained from measurements in reverberation chambers (ISO 354). In ray tracing, these data represent the average absorption at random-incidence. Thus, in rooms that provide an approximate diffuse field, they serve well, particularly in the late response where in the region of high-order reflections, the average absorption coefficient is an appropriate quantity. The energy losses of the first reflections at their specific angles of incidence, however, are not modelled precisely. Nevertheless, the errors are small. It should be kept in mind that the difference in energy loss at a specific angle of incidence to random-incidence may be up to 40%, until the error of the reflection level exceeds 2 dB. Only at grazing incidence, may the errors become larger. But at grazing incidence, anyway, other influences come into play (see Sect. 11.5).

If the complex wall impedance is known, the angle-dependent absorption coefficient can be calculated. This is easy for real-impedance, locally reacting walls (see Eqs. (3.3)–(3.6)). Each case of room simulation should be checked to determine if random-incidence absorption coefficients are applicable or if angle-dependent data are possibly required. The latter may apply to long or flat rooms, tunnels, corridors and coupled rooms, for example.

11.1.3 Scattering coefficients

If absorption is treated in a simplified way (assuming random incidence), this is even more justified for surface scattering. Nevertheless, angle-dependent scattering is interesting for discussing low-order reflections. The total field in front of a diffuser, depending on the signal's spectrum and coherence length, is very complicated (Sect. 3.3). Data from free-field measurements (part 2 of (ISO 17497)) or calculations can be obtained by using standardized methods, but these are valid only for harmonic signals.

Therefore, with the same argument as used above, we assume random incidence appropriate, and we accept the unavoidable uncertainties that are small in high-order reflections. But we must be aware of problems which can arise in first- or second-order reflections.

Random-incidence scattering coefficients are not available to the same extent as absorption data. However, with the measurement method recently standardized in ISO 17497, more information may be published in the future (see Annex).

11.2 Stochastic ray tracing

In this model, sound is radiated as bunches of particles – many particles (= rays). The number of particles is one of the crucial features of ray tracing. It is considered in the category of Monte Carlo methods to express the influence of probabilistic effects, such as in gambling.

Stochastic ray tracing in short

The sound source radiates an impulse. To simulate this, particles are started at the same initial time in various directions. Each ray carries certain energy, propagates at the speed of sound and once in a while hits the room boundary, for simplicity called a "wall" throughout. It is reflected from the wall, hits another wall and so on. Each wall absorption reduces the particle's energy.

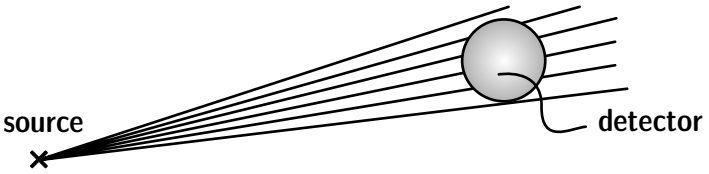

Fig. 11.6. Free field propagation and distance law "by counting"

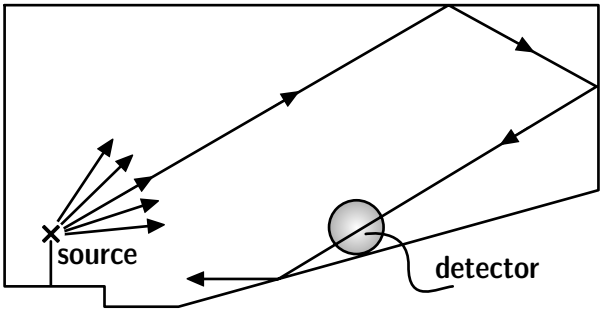

Fig. 11.7. Tracing a ray from the source to the detector

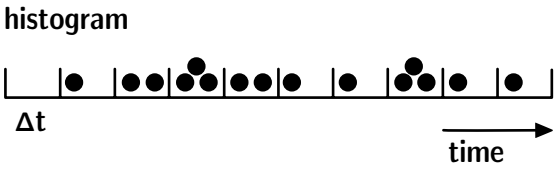

Fig. 11.8. Creating an impulse response by counting events

By means of particle detectors, the particle's energy and the time elapsed since radiation from the source are registered (see Fig. 11.8). The number of counts represents the energy detected at the receiver.

Source modelling

Sound sources are characterized by sound power and directivity (Sect. 2.4). Both depend on frequency. For ray tracing, we need just the source position and the reference (on-axis) direction. The directivity can be modelled by choosing direction-specific start energies of the particles or by variation of the density of particles, as illustrated in Fig. 11.9 in an area-accurate plot of the unit sphere.

To create spherically uniform radiation, we proceed as follows: The azimuthal angles are distributed evenly around the perimeter. The distribution of polar angles is to be chosen so that the ray density is constant. This is achieved by using a random number $z \in (0,1)$ (or by dividing the interval (0,1) into fixed steps) and taking the polar angle of radiation as

$$\theta = \begin{cases} \arccos z \\ \arccos z + \pi/2 \end{cases}, \tag{11.1}$$

the upper choice for the northern and the lower for the southern hemisphere. $\theta = 0$ at the north pole. Other weighting functions create polar

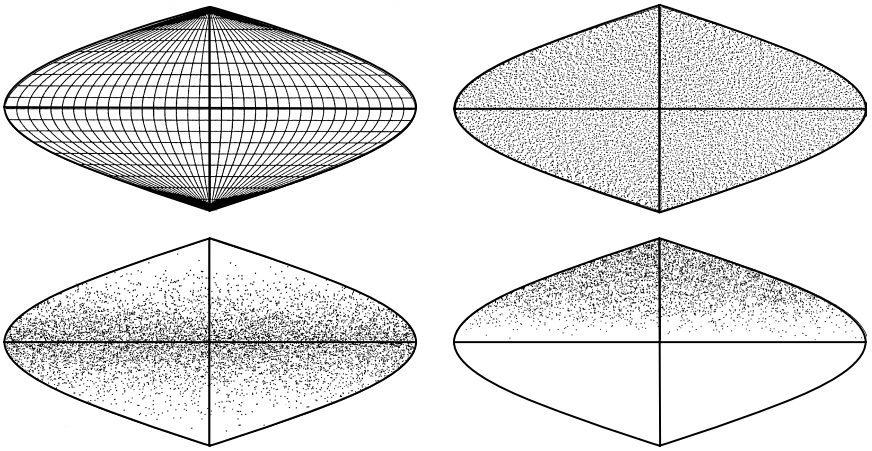

Fig. 11.9. Top left: Variants of ray sources, spherical coordinate system and regular grid. Top right: Uniform random distribut ion. Bottom left: Concentration of rays at the equator (line-array-like). Bottom right: Concentration of rays at the north pole (baffled-piston-like)

distributions with concentrations of rays at the north pole (reference axis) or at the equator.

Any measured or calculated directional pattern can also be used as a spherical weighting function of the ray energy or of the ray density.

Flow diagram

Due to the frequency dependence of absorption, ray tracing must be repeated for the frequency bands of interest, usually octave bands or one-third octave bands.[42] Air attenuation can also be modelled by reducing the ray's energy according to the flight distance and the attenuation coefficient.

We will now focus on ray tracing algorithms in more detail. There are two options for modelling absorption. It is obvious that the energy can be reduced by multiplying the incident energy by the factor $(1 - \alpha)$. The alternative is to apply stochastic annihilation of particles. Both algorithms yield identical results in the limit of large numbers, but they differ in computation time and in accuracy.

In absorption by multiplication, the particle starts with energy e_0 and is traced until a maximum travel time t_{max}, or until a minimum energy, e_{min}, is reached.

[42] Under special circumstances, parallel processing of frequency bands is possible; see below.

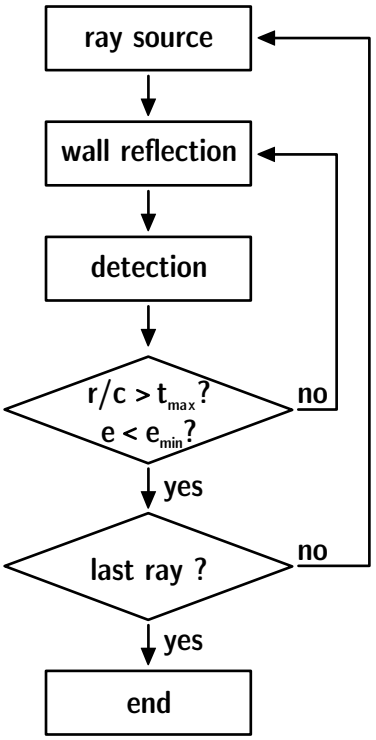

Fig. 11.10. Flow diagram of ray tracing

In the method of absorption by annihilation, a random number $z \in (0,1)$ is compared with the absorption coefficient, α. If $z < \alpha$, the particle is annihilated and the next particle is traced; see flow diagram in Fig. 11.10. As long as there is absorption somewhere in the room, we do not need to fix another truncation criterion, such as t_{max} or e_{min}.

11.2.1 Point-in-polygon test

The most inner loop of ray tracing algorithms is the test of whether a ray (particle), represented as a line segment, a) is hitting a plane and b) if the intersection point is inside or outside a wall polygon. At this step, the actual straight line and the direction of propagation, thus, a vector are known. Also known is the set of polygons.

The polygons qualified (those located in forward direction of the ray) are transformed in 2-D coordinate systems of the polygon planes and tested by calculating vector products of the vectors from the plane intersection point to vertices of the polygon. For all neighbouring vertices, the vector product

must have the same direction. Otherwise the intersection point is outside the polygon. An advantage of this algorithm is that it can be cancelled if the vector is oriented once in the opposite direction.

If the intersection point is inside the polygon, the new travel line will be input variable for the next plane hit. Before reflection takes place, wall materials affect sound by absorption, scattering and diffraction. If a particle hits a wall, it will lose energy or be annihilated,[43] and it will change its travel direction. For specular reflection, a new direction at a specular angle with reference to the normal incidence will be calculated.

When scattering occurs, which applies when a random number exceeds the scattering coefficient, the new direction is obtained from two more random numbers. Mostly, a directional distribution is assumed according to Lambert's law. On average, the polar angles of the new flight direction (with reference to the wall normal vector) must be distributed according to

$$w(\theta)\mathrm{d}\Omega = \frac{1}{\pi}\cos\theta \, \mathrm{d}\Omega \,. \tag{11.2}$$

The azimuthal angle is evenly distributed in $(0, 2\pi)$. Hence, two random numbers z_1 and z_2 in (0,1) and transformations

$$\theta = \arccos\sqrt{z_1} \tag{11.3}$$

and

$$\phi = 2\pi \, z_2 \tag{11.4}$$

yield a polar scattering distribution according to Lambert's law. Superposition of the specular direction vector and the scattering direction vector is also a possible solution to define the new direction.

11.2.2 Detectors

Surface or volume detectors are possible options. While surface detectors have an angle-dependent cross section, spherical detectors are independent of the angle of incidence. The detection cross section of spheres is

$$M_{\text{sphere}} = \pi \, r_{\mathrm{d}}^2 \,, \tag{11.5}$$

where r_{d} denotes the detector radius. Thus, the detector position (centre point) is sufficient information. Each line segment representing the ray has to be checked to determine whether it "hits" the detector. This test requires

[43] depending on the way absorption is implemented.

calculating the distance between the detector centre point and the ray vector. It is solved by finding the perpendicular line between the ray and the detector and by checking if this distance is smaller than the detector radius, r_d.

This method of ray detection represents an omnidirectional microphone. It does not represent a live audience. All physical effects of audience areas known in room acoustics are neglected. The most prominent effects are the seat-dip effect and the forward scattering at the audience heads. Including both (wave) effects in models of detectors was tried, but all attempts suffer from inconsistencies with geometrical acoustics. Anyway, audience effects can be modelled as a specific feature of audience areas, independent of detector modelling. Complex impedance and diffraction models are options for including audience effects in future simulation models, however, not in stochastic ray tracing, but in deterministic models (see below).

11.2.3 Presentation of results

Up to this point, ray tracing has been described abstractly. By discussing typical results, we can better interpret the quality, the efficiency and the benefit of ray tracing.

At each detection, the ray's energy, arrival time and direction of incidence are stored in an array called a histogram of the energy impulse response. The array must have a sampling rate that allows a sufficiently high temporal resolution, on the one hand, and sufficiently large integration intervals, on the other (see below, Sect. 11.2.5). A good compromise is a resolution of the order of magnitude of milliseconds. Psychoacoustic post-masking can be taken into account properly, but binaural attributes such as ITD cannot[44] (see Fig. 11.11). At the same time, it is clear that we cannot meet the requirement of a sampling rate adequate for audio signal processing and auralization.[45]

One could argue that auralization with stochastic ray tracing could be possible with faster computers. This, however, is not a reasonable goal since ray tracing is a rough approximation of wave propagation. Any increase of in accuracy and temporal resolution would only mimic a real gain in physical accuracy. Ray tracing still remains an approximation of wave propagation.

[44] It will be shown in Sect. 11.2.5 that the required number of rays would be unacceptably high.

[45] The temporal resolution for a sampling frequency of 44.1 kHz is 22.7 μs.

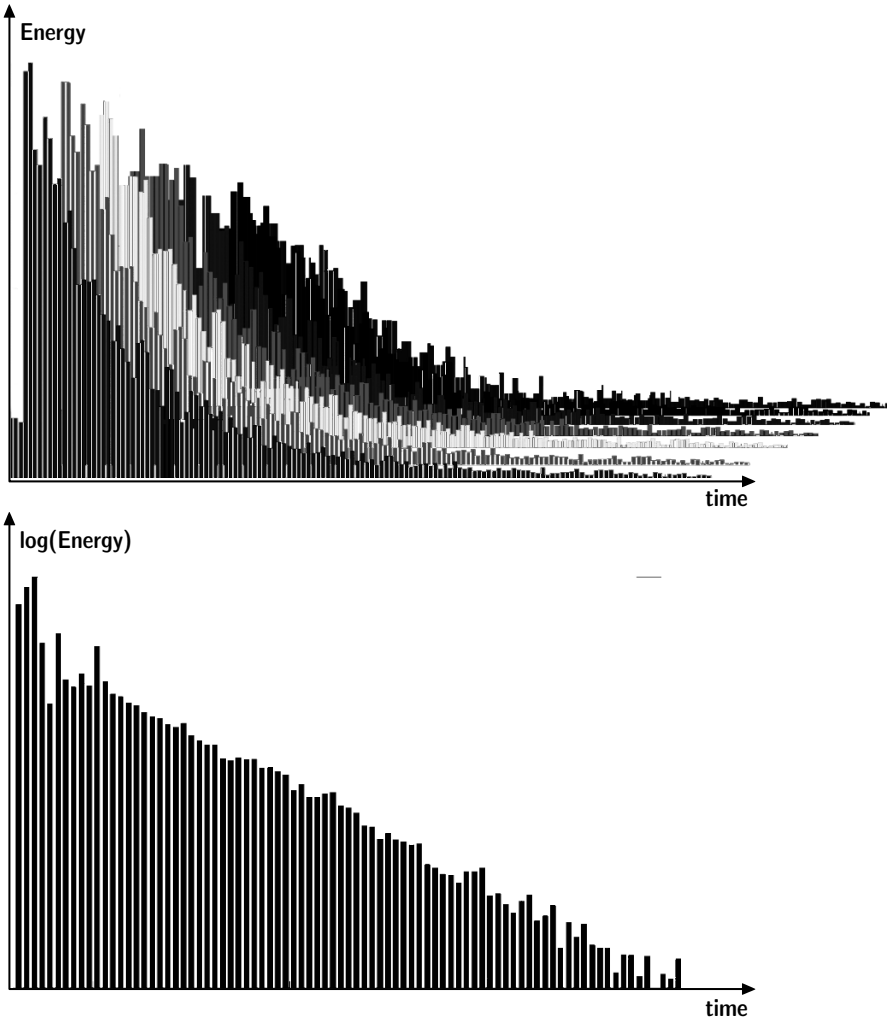

Fig. 11.11. Ray tracing energy impulse response. Top: Linear plot for several frequency bands. Bottom: Logarithmic plot for one frequency band

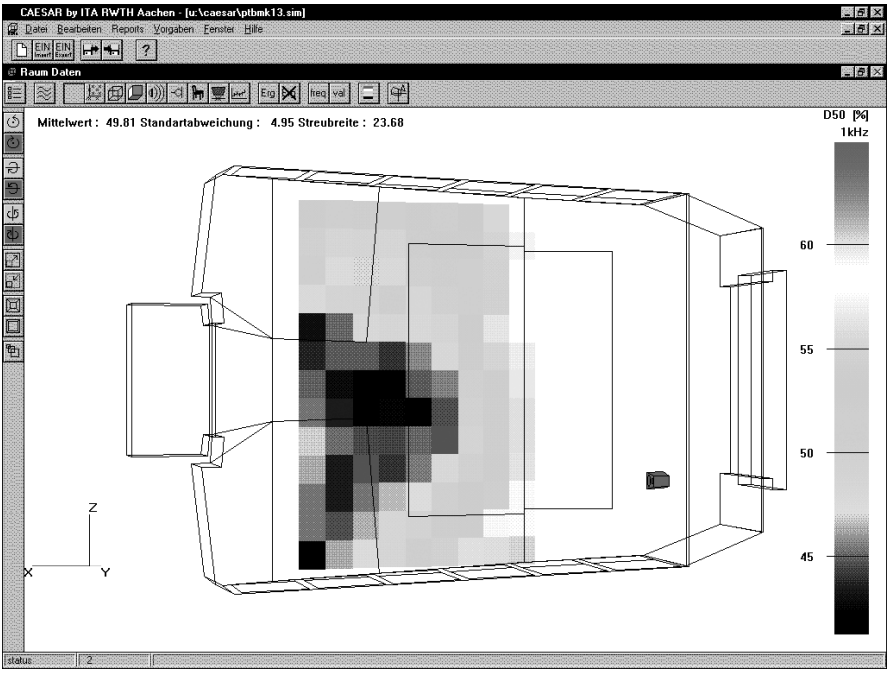

Fig. 11.12. Examples of room acoustic parameters plotted in the CAD model (audience area); see Sect. 6.4 and particularly Sect. 6.4.6

11.2.4 Curved surfaces

Inappropriate architectural room designs may lead to severe acoustic problems. Curved walls belong to the category of high risk potential. Strong spatial and temporal concentrations of sound (= "echoes") must be avoided since sound focusing is against all goals of room acoustic design to achieve the best possible uniform distribution of good acoustic quality. In rooms with cylindrical side walls or with dome-shape ceilings, we have to expect focusing. If room acoustic computer simulation shall be a tool for design, it must be ensured that occurrences of focusing and echoes are clearly visible in the results.

An exact model of the room geometry by using curved surfaces may be even faster than the approximation by planes. To obtain accurate results with ray tracing, the detectors must be small and the number of rays large. It was shown (Kuttruff 1995) that deterministic approaches with coherent image source contributions can be used. This is obvious since the sound pressure in the focal region cannot be obtained by energetic models.

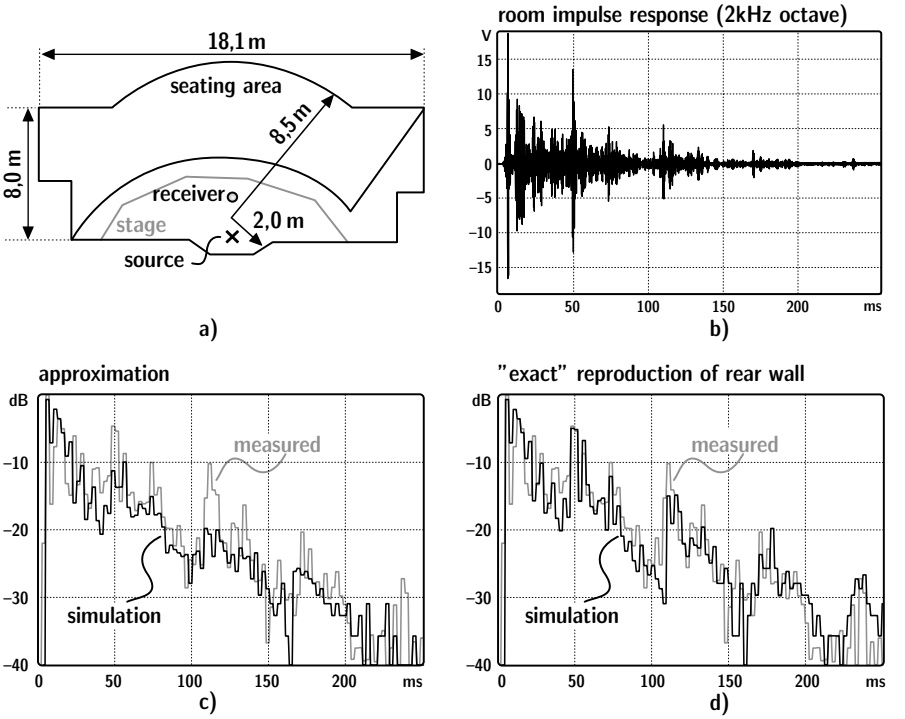

Fig. 11.13. (a) Ground plan of a small theatre (volume 900 m^3) with source and receiver positions, (b) Measured impulse response (2 kHz octave) and comparison with simulated impulse responses. (c) Approximation of five plane surfaces, (d) Mathematically exact model of the cylindrical back wall, after (Mommertz 1996)

The approximation of a cylinder with radius a by planes of width b is sufficient if

$$f\, b^2 = \frac{1}{2} c\, a \tag{11.6}$$

is fulfilled. It is interesting that this equation is frequency dependent. Thus, the condition must be checked for the highest frequency band involved, or the shape must be modelled as frequency dependent. For example, a cylindrical shape with radius 10 m must be modelled by panels 50 cm wide, if the frequency range reaches 8 kHz This means that the cylinder should be subdivided into 136 plane elements.

11.2.5 Reproducibility in stochastic ray tracing

Systematic uncertainties are present in all physical measurement and simulation methods. The approximations inherent in the input data (room geometry, absorption and scattering coefficients) and in the model itself (approximation of wave physics by a geometric model) are causes for systematic errors. These errors cannot be reduced by increased computational effort.

Independently, in Monte Carlo methods such as ray tracing, we must consider stochastic fluctuations in the results. These errors can be reduced by increasing the number of rays or by averaging repeated simulations. The number of rays launched, N, is of crucial importance because it affects both the reproducibility and the computation time t_{calc}.

The goal of this chapter is predicting the standard deviation in the simulation results in dependence on N and t_{calc}. To obtain an expectation value of the error, we remember that the energy impulse response was created by counting events of particle detection. The number of hits, k, in a detector sphere with radius r_{d}, in a time interval, Δt, is of particular interest, as well as the detection rate, r, of hits per second.

The probability of a ray hitting the detector within Δt is small but constant in time, if the number of rays in the room throughout the simulation is constant. The expectation value $<k>$, thus, is constant and follows a Poisson distribution.

$$P(k) = \frac{\langle k \rangle}{k!} e^{-\langle k \rangle} \tag{11.7}$$

The variance of the Poisson distribution, σ^2, describes the fluctuations around the expectation value. It is

$$\sigma_k^2 = \langle k \rangle = \langle r \rangle \Delta t \,. \tag{11.8}$$

The number of hits is thus related to the following expected error:

$$\sigma_k \,/ \langle k \rangle = \frac{1}{\sqrt{\langle k \rangle}} \,. \tag{11.9}$$

The expectation value of the count rate $<r>$ is calculated from the number of wall hits per second, $\bar{n}N$, the ratio of the detector surface, $4\pi r_{\mathrm{d}}^2$, and the room surface, S:

$$\langle r \rangle = \bar{n} N \frac{4\pi r_{\mathrm{d}}^2}{S} \,. \tag{11.10}$$

Herein, $\overline{n}$ is the mean reflection rate according to Eq. (4.22). The mean count rate is, thus,

$$\langle r \rangle = N \frac{\pi r_{\mathrm{d}}^2 c}{V}, \tag{11.11}$$

and the mean number of hits is

$$\langle k \rangle = N \frac{\pi r_{\mathrm{d}}^2 c \Delta t}{V}. \tag{11.12}$$

Now we calculate the expectation value of the energy, <*E*>, per time interval Δt. This can be obtained only when some features of the sound field are assumed. The diffuse sound field is chosen as a best guess for a room sound field. This approach is, by the way, exactly the same as in statistical reverberation theory (Sect. 4.4). Of course, an exponential decaying <*E*> will be found, which matches the late decay better in a real room. Thus, early reflections are not discussed here. The theory is identical to that discussed in Sect. 4.4. Therefore,

$$E(t) = e_0 N (1-\alpha)^{\overline{n}t} = e_0 N e^{-\frac{13.8t}{T}}, \tag{11.13}$$

where e_0 denotes the starting energy of rays. Depending on the absorption model used, the energy–time curve is represented by the count rate and the number of hits.

Energy multiplication

In this model, the number of rays (particles) in the room and the average hits of the detector remain constant (Eq. (11.12)). The mean energy is, thus,

$$\langle E \rangle = \langle k \rangle e_0 e^{-\frac{13.8t}{T}} \tag{11.14}$$

and

$$\frac{\sigma_E}{\langle E \rangle} = \frac{\sigma_k}{\langle k \rangle} = \sqrt{\frac{N V}{\pi r_{\mathrm{d}}^2 c \Delta t}}. \tag{11.15}$$

The relative error σ_E/<*E*> is easier to interpret when expressed in decibels. Now, we can use Eq. (11.16) which states that the time-dependent sound pressure level is

$$L_{\mathrm{ETC}} = 10 \log(E / E_0) \tag{11.16}$$

with an arbitrary reference E_0, and

$$\sigma_{\mathrm{ETC}} \approx \frac{\mathrm{d}L}{\mathrm{d}E}\sigma_E = \frac{10}{\ln 10}\frac{\sigma_E}{\langle E\rangle}. \tag{11.17}$$

Thus,

$$\sigma_{\mathrm{ETC}} = 4.34\sqrt{\frac{V}{N\pi r_{\mathrm{d}}^2 c\Delta t}}. \tag{11.18}$$

This error is constant with respect to the time variable in the impulse response. It depends on the ratio of the room volume to the number of rays, the detector size and the width of the time intervals, $V/Nr_{\mathrm{d}}^2\Delta t$.

Furthermore, on the basis of E, integral parameters can be obtained, such as clarity or strength. The strength, for instance (see Eq. (6.16)), is derived from the integral[46]

$$\begin{aligned}\langle w\rangle &= \sum_{\Delta t}\langle E\rangle \approx e_0 N\frac{\pi r_{\mathrm{d}}^2 c}{V}\int_0^\infty e^{-\frac{13.8t}{T}}\,\mathrm{d}t \\ &= \frac{4\pi r_{\mathrm{d}}^2 e_0 N}{A}\end{aligned}. \tag{11.19}$$

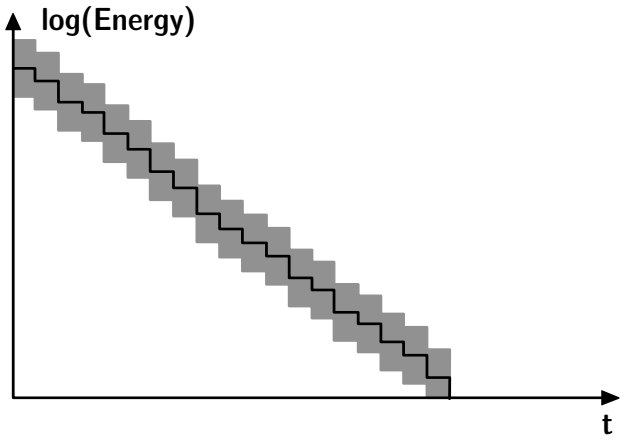

Fig. 11.14. Histogram on a logarithmic scale of stochastic ray tracing (absorption by energy multiplication), expectation values and uncertainty range according to Eq. (11.18) (example)

[46] Accordingly, the sound power of the source, P, is equivalent to $\pi c r_{\mathrm{d}}^2 e_0 N$.

The expectation value of this energy integral, <*w*>, also is affected by stochastic deviations. Due to the statistical independence of the hits in the Δt intervals, the total variance of the integral is

$$\sigma_w^2 = \sum_{\Delta t} \sigma_E^2 \ . \tag{11.20}$$

and, thus, again expressed in decibels (applying the principle of Eq. (11.17) again),

$$\sigma_L = 4.34 \sqrt{\frac{A}{8\pi N r_\mathrm{d}^2}} \ . \tag{11.21}$$

The most important fact we observe is that the level errors depend on the square root of A/N. Thus, in large and highly absorbing rooms (large A), a larger number of rays (larger sound power) is required, such as in measurements, to overcome the limit of the background noise floor. The noise, here, is represented by numerical noise given by the discrete ray formulation.

Finally, the computation time shall be estimated. To obtain the number of operations, we have a look at the flow diagram (Fig. 11.10). The inner loop is the tracing and reflection procedure (vector/plane intersection). The number of walls to be checked is n_w, the number of detectors, n_d. The computation time required for the point-in-polygon test is t_w, the time for

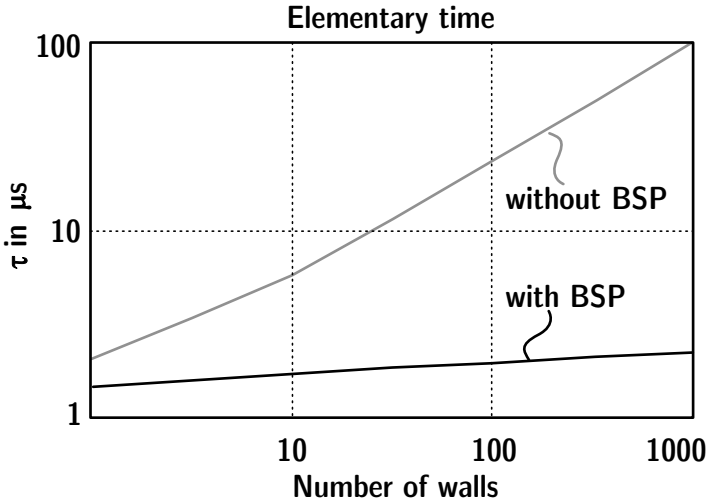

Fig. 11.15. Order of magnitude of elementary computation time per reflection as a function of the number of wall polygons, without and with binary space partitioning (example for PC with 2 GHz CPU)

checking detectors is t_d. Without spatial data structures such as the voxel technique or binary space partitioning (BSP) (see Sect. 11.3.5), the complete set of walls and detectors must be tested. The computation time for one reflection is considered elementary time, τ:

$$\tau \approx n_w t_w + n_d t_d + t_c \,, \tag{11.22}$$

where t_c denotes a constant offset for data management independent of the number of walls and detectors.

With BSP or similar spatial data structures, the elementary time reduces to

$$\tau \approx \log_2 n_w \; t_w + \log_2 n_d \; t_d + t_c \,. \tag{11.23}$$

The elementary time is to be multiplied by the events of reflections, $N\overline{n}t_{max}$:

$$t_{calc} \approx N\overline{n}t_{max}\tau \tag{11.24}$$

As a rough estimate, the following elementary computation times can be expected (see Figs. 11.15 and 11.16). These plots illustrate the dependence of the computation time (order of magnitude) on the number of walls and on the number of rays. The number of receivers is set at $n_d = 1$.

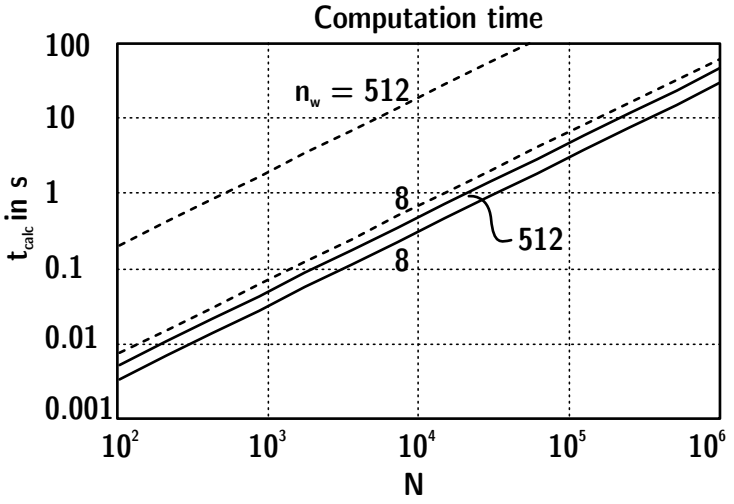

Fig. 11.16. Order of magnitude of the total computation time as a function of the number of rays, N (example for PC with 2 GHz CPU). The upper two broken lines represent classical ray tracing, the lower curves ray tracing with room subdivision strategy (example BSP). The room volume is the same in all four cases

Annihilation

When using this method, more and more rays vanish. The decision to leave the inner loop is made according to a random number. In case of absorption, the next ray is started.

The expectation value of the energy decreases with the running time in the impulse response since the number of hits decreases exponentially,

$$\langle k \rangle = N \frac{\pi r_{\mathrm{d}}^2 c \Delta t}{V} e^{-\frac{13.9t}{T}}, \tag{11.25}$$

and because of

$$\langle E \rangle = \langle k \rangle e_0, \tag{11.26}$$

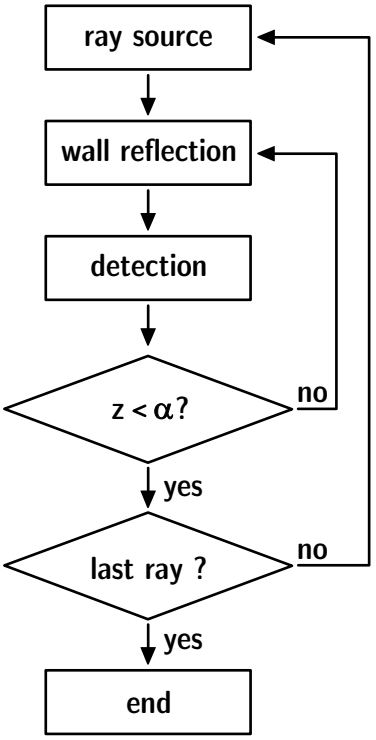

Fig. 11.17. Flow diagram of ray tracing (annihilation)

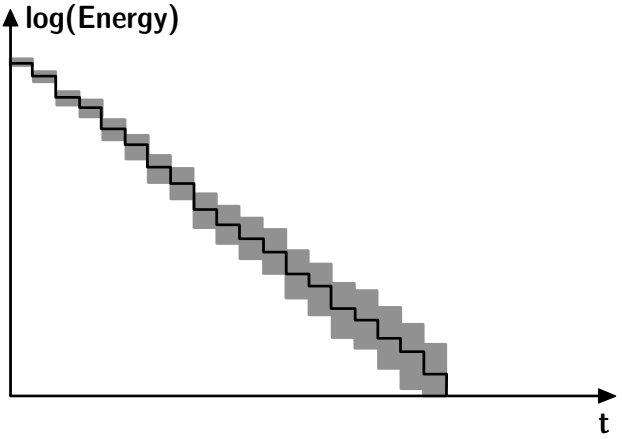

Fig. 11.18. Histogram of stochastic ray tracing (absorption by annihilation), expectation values and uncertainty range according to Eq. (11.27) (example)

there results an increasing relative variation in the energy impulse response of

$$\frac{\sigma_E}{\langle E \rangle} = \frac{\sigma_k}{\langle k \rangle} = \sqrt{\frac{V \cdot e^{\frac{13.8t}{T}}}{N\pi r_\mathrm{d}^2 c\Delta t}} = \sqrt{\frac{V}{N\pi r_\mathrm{d}^2 c\Delta t}} \cdot e^{\frac{6.9t}{T}} . \tag{11.27}$$

The increase in the uncertainty in the late reverberation tail is due to the loss of countable rays caused by absorption. The statistical error of the integral energy level (in decibels) is

$$\sigma_L = 4.34\sqrt{\frac{A}{4\pi N r_\mathrm{d}^2}} . \tag{11.28}$$

When comparing the equation with Eq. (11.21), it is observed that the difference in the integral energy error is a factor of 0.7 in advantage of the energy multiplication. This is mainly due to the fact that rays disappear and leave a smaller statistical basis.

The computation time can be estimated from the flow chart, too. But we have to determine the average number of loops for each ray. The probability that it survives $\nu - 1$ reflections and is annihilated at the νth reflection, is

$$w(\nu) = (1-\overline{\alpha})^{\nu-1}\,\overline{\alpha} , \tag{11.29}$$

where $\overline{\alpha}$ denotes the mean absorption coefficient (Eq. (4.30)). Accordingly,

$$\langle \nu \rangle = \sum_{\nu=0}^{\infty} \nu \, w(\nu) = \overline{\alpha} \sum_{\nu=0}^{\infty} (1-\overline{\alpha})^{\nu-1} = \frac{1}{\overline{\alpha}} , \tag{11.30}$$

and

$$t_{\text{calc}} = \frac{N\tau}{\overline{\alpha}} . \tag{11.31}$$

11.2.6 Computation times versus uncertainties – case studies

In the following, we consider four examples of rooms as typical cases for categories of rooms used for different purposes: A living room ($V = 100\,\text{m}^3$), a lecture room (classroom) ($V = 1{,}000\,\text{m}^3$), a concert hall ($V = 10{,}000\,\text{m}^3$), and a large church ($V = 100{,}000\,\text{m}^3$). The rooms are modelled with constant average absorption of $\overline{\alpha} = 0.15$. The increasing number of walls, n_{w}, is due to the larger complexity with increasing volume.With constant average absorption, the reverberation time also increases.

Table 11.2. Examples of room situations

	Room 1 "living"	Room 2 "lecture"	Room 3 "concert"	Room 4 "church"
Volume in m^3	100	1,000	10,000	100,000
Surface area in m^2	136	906	4,725	21,733
Mean reflection rate in s^{-1}	116	77	40	18
Reverberation time in s	0.8	1.2	2.3	5.0
Number of walls	8	20	50	100

In these categories of rooms, we compare the computation times which are required to achieve a certain quality of results (one receiver, one frequency band). At first, impulse responses are discussed. They may be intended either for visual inspection (see Fig. 11.11) of the quality of the energy impulse response or as a basis for further processing with the aim of auralization (see Sect. 11.7). The time interval in the histogram is set to 10 ms. Now the type of algorithm for absorption must be chosen. According to the increasing error illustrated in Fig. 11.18, the energy multiplication

method is clearly preferable. The truncation time is set to $t_{max} = T$. With the definition of the maximum error at each time interval of 1 dB,[47] the required number of rays, N, is defined and the computation times, t_{calc}, in the following table can be expected. The data, of course, represent only an order of magnitude, since they strongly depend on the hardware used. A standard PC with 2 GHz and 1 Gbyte RAM serves as an example. The programming language is C++.

Table 11.3. Order of magnitude of computation times in s for an expected error of 1 dB in each time interval of the energy impulse response (energy multiplication method), $t_{max} = T$.

Computation time in s (Eq. (11.24))	Room 1 "living"	Room 2 "lecture"	Room 3 "concert"	Room 4 "church"
N	200	2,000	20,000	200,000
t_{calc}	0.03	0.5	10	180
t_{calc} (BSP)	0.02	0.2	2.7	29

Next, an integral parameter such as the sound level (strength G) is the goal of the simulation.[48] Again, the limit is set to 1 dB, but in this example, it is related to the total energy in the impulse response. The time limit of ray tracing is $T/2$, which is sufficient for determining the total level (Eq. (11.39)). The number of rays is calculated according to Eq. (11.21).

Table 11.4. Order of magnitude of computation times in ms for an expected error of 1 dB in the parameter strength, G. Energy multiplication method, $t_{max} = T/2$.

Computation time in ms (Eq. (11.24))	Room 1 "living"	Room 2 "lecture"	Room 3 "concert"	Room 4 "church"
N	16	100	500	2,500
t_{calc}	2	15	150	1,300
t_{calc} (BSP)	1	7	40	200

Now, the sound level (strength G) is the goal of the simulation again, and the limit is set to 1 dB. This time the annihilation method is used. The required number of rays is calculated according to Eq. (11.28). No time limit such as t_{max} applies since the rays are traced until their annihilation.

[47] inspired by the limit given by the jnd for a sound level of roughly 1 decibel.

[48] It is assumed that the specific impulse is not relevant in detail. Instead, the goal of the simulation is a quick estimate of the level, clarity, definition or other integral parameters.

Table 11.5. Order of magnitude of computation times in ms for an expected error of 1 dB in the parameter strength, G. Annihilation method

Computation time in ms (Eq. (11.31))	Room 1 "living"	Room 2 "lecture"	Room 3 "concert"	Room 4 "church"
N	31	200	1,000	5,000
t_{calc}	0.06	0.6	6	54
t_{calc} (BSP)	0.04	0.3	1.7	8.6

Note that computation times for detailed analysis of the impulse responses are much greater than those required for integral results. In other words, integral results can be estimated very quickly. For auralization, without reference to the specific room acoustic field, therefore, it might be of interest to estimate the room acoustic parameters quickly and to adjust artificial room impulse processors of early reflections and exponential late reverberation with reference to the integral simulation results (see Sect. 15.2.3). Note that this type of quick parameter estimation is best done by stochastic ray tracing with the annihilation method, since it offers real-time capability (time limit about 50 ms; see Chap. 15), even for the large church (Room 4).

11.3 Image source model

On the basis of the image source principle (Sect. 4.3) and extension to geometric phase superposition, the total sound pressure of direct sound and various reflections can be modelled by adding (complex) spherical wave amplitudes. However, the model is exact when $|R| = 1$. For absorbing walls, it is a good approximation, as long as the angle of sound incidence, ϑ_0, is small, far from grazing incidence (see Sect. 3.2).

11.3.1 Classical model

The reflection factor is assumed to be angle-independent, corresponding to a quasi-plane wave:[49]

$$p = \frac{j\omega\rho_0\hat{Q}e^{-jkr_0}}{4\pi r_0} + \frac{j\omega\rho_0\hat{Q}e^{-jkr_1}}{4\pi r_1}\underline{R}(\vartheta_0)\,. \tag{11.32}$$

[49] The plane-wave assumption is equivalent to the prerequisite of nongrazing sound incidence.

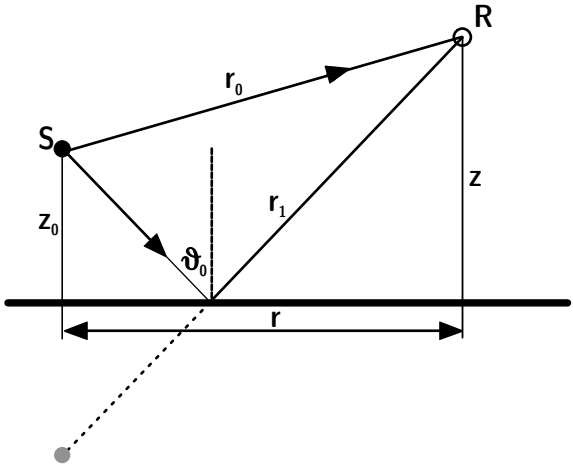

Fig. 11.19. Image source model

At the same time, it is clear that the wall must be smooth and specularly reflecting. From the definition of the single reflection, the image source model can be described as follows. For specific geometries (rectangular, triangular), it served as model for analyzing basic features of room impulse responses since the midddle of the last century (Cremer 1948). Then, the so-called "Allen-Berkley/Borish" (Allen and Berkley 1979) model was first implemented by Borish (1984) in arbitrary polyhedra and later used in numerous versions, not only in acoustics, but also in radio wave physics and in computer graphics in similar ways.

If the room reflections are purely specular, the sound paths (rays) can be backtraced from the receiver to the source. This is achieved by using virtual (image) sources. At first, they must be constructed for the room of interest. The original source is mirrored at the wall planes. Each image source is, again, mirrored at wall planes, to create image sources of higher order. All permutations of the walls must be considered, except a constellation involving the same wall subsequently. Under specific circumstances, walls can be excluded due to geometrically inconsistent ray paths (Mechel 2002).

With $\vec{S}$ denoting the source position, $\vec{S}_n$ the position of the image source, $\vec{n}$ the wall normal vector ($|\vec{n}|=1$) and $\vec{r}$ the vector between the foot point, $\vec{A}$, of the wall normal and source, $\vec{S}$, the scalar product of $\vec{n}$ and $\vec{r}$ yields the distance between the source and the wall, *d*

$$d = \vec{r} \cdot \vec{n} = |\vec{r}| \cos\alpha . \tag{11.33}$$

With this distance, we get the position of the image source,

$$\vec{S}_n = \vec{S} - 2d\vec{n} . \tag{11.34}$$

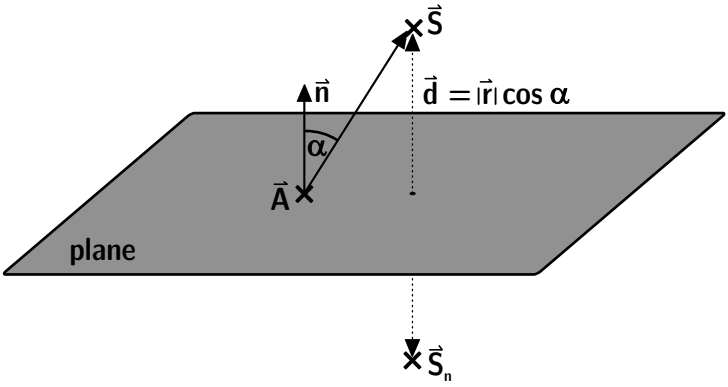

Fig. 11.20. Construction of an image source

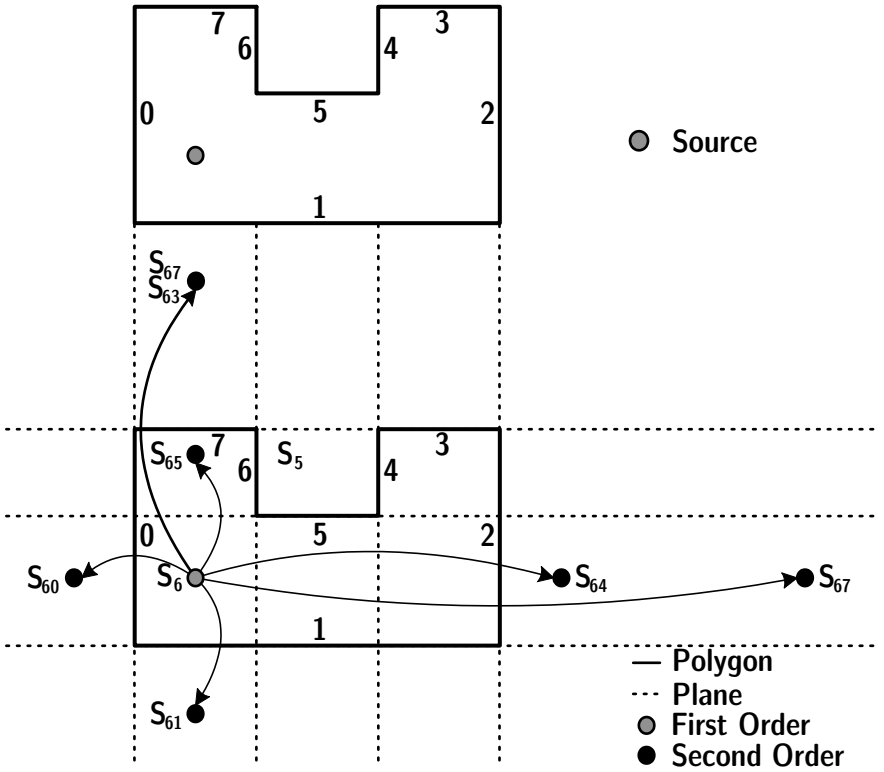

Fig. 11.21. Image sources constructed for a room (example in 2-D)

With this procedure applied to all walls we construct an image source of the first order. Image source of higher order are constructed in the same way by considering first-order image sources as “mother” sources, and so

on. The process of mirroring is continued until a certain maximum order of image sources is reached. The truncation of the process is similar to the truncation of ray tracing at a maximum time, t_{max}. Image sources of nth order correspond to rays hitting n walls.

11.3.2 Audibility test

With the set of image sources created, a so-called "audibility test" must be performed. It is necessary to check the relevance of each image source for the specific receiver position. Receivers, by the way, are points.

Each image source is interpreted as the last element of a chain of sources. The indices denote the series of walls hit on the path of the corresponding ray, while the number of indices denotes the order of the image source. A chain of ith order is

$$S \to S_{n_1} \to S_{n_1 n_2} \to ... \to S_{n_1 n_2 ... n_{i-1}} \to S_{n_1 n_2 ... n_i}\,, \tag{11.35}$$

with $n_k \neq n_{k\pm1}$ counting the walls hit, $n_k \in (1,\ n_w)$. The audibility test is started at the receiver point and the ray is backtraced from this point to the source along the chain of image sources, as shown in Fig. 11.22.

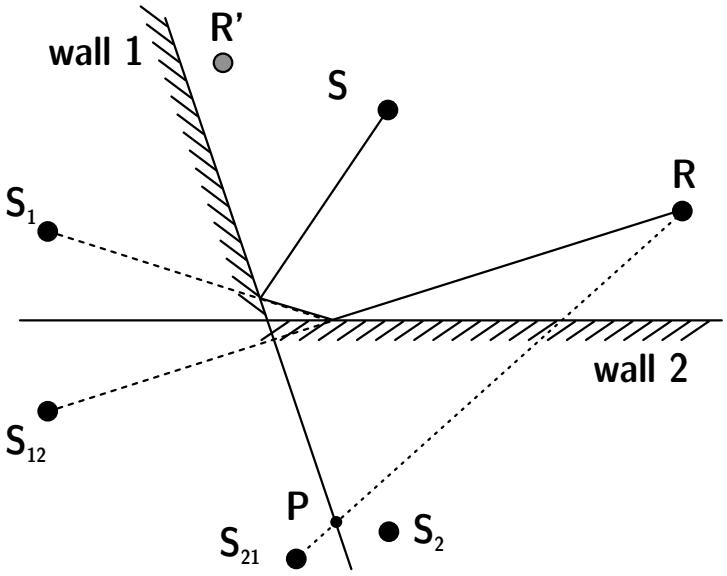

Fig. 11.22. Image source audibility test

As an example, Fig. 11.22 shows the determination of audible image sources up to the second reflection order. The current receiver point is connected with the image source under test, S_{12}. The last index indicates that wall 2 was the last hit. If the intersection of the straight line with the wall polygon 2, $\overline{RS_{12}}$, is inside the polygon, the result is preliminarily positive. We then continue by drawing a line from the intersection point to the mother source[50] of S_{12}, S_1. This procedure is repeated until the original source is reached. If all intersections hit inside the polygons, the image source (in this case S_{12}) is in fact audible from the receiver point R. Note that S_{21} is not audible from R since P is outside wall polygon 1, but it would be audible from R'. Thus each source must be tested specifically in relation to the receiver. Also, the fact that one image source is audible cannot be generalized for its predecessors. It should be noted, too, that the crucial test of this procedure is the point-in-polygon test, such as in ray tracing. All strategies, therefore, to accelerate the backtracing procedure by spatial substructures may be applied as well.

For audible image sources, the complex amplitudes are stored in the sound pressure impulse response. The delay of the impulse related to each image source is calculated from the distance between the image source and the receiver, r_{IS}.

$$p_{IS} = \frac{j\omega\rho_0 \hat{Q}\Gamma e^{-j\omega t_{IS}}}{4\pi c t_{IS}} \prod_{n=1}^{i} \underline{R}_n \tag{11.36}$$

with $t_{IS} = r_{IS}/c$ and $\underline{R}_n$ the reflection factors of the walls involved in the respective chain source. No time intervals are required at this stage of processing. The temporal resolution can be infinitely high, and the sampling rate, thus, can be freely defined.

An estimate of the reflection factor from the absorption coefficient can be done straightforwardly by using

$$|R| = \sqrt{1-\alpha} \tag{11.37}$$

and an appropriate phase. Rindel (1993) and Mommertz (1996) show that absorption coefficients in one-third octave bands serve well for reconctruction of the complex reflection factor. Mimimum or linear phases are options for the phase function, to be reconstructed from the Hilbert transformation, for instance.

[50] A mother source is the source located in the chain as a predecessor.

11.3.3 Limitations

Truncation

The image source model is a strictly deterministic method. Uncertainties are caused by prerequisites of geometrical acoustics and by necessary truncation of the image source at a certain order. Although not explicitly implemented, an average maximum time, $t_{\max}$, has to be taken into account. The maximum order is related to the computation time, which is more crucial than that in ray tracing. This is due to the dramatic increase of the number of sources to be treated with higher orders.[51] The number of audible sources, however, increases with t^3 in the impulse response (Cremer 1948).

At higher orders of image sources, the ratio between sources constructed and sources visible is uneconomic. With i denoting the maximum order chosen, the truncation time is

$$t_{\max} = i / \overline{n} \,, \tag{11.38}$$

and the energy missing is

$$\frac{\Delta w_{trunc}}{\langle w \rangle} = (1-\alpha)^i \,. \tag{11.39}$$

The computation time, t_{calc}, is similarly estimated from the time required for the point-in-polygon test:

$$t_{\text{calc}} = n_{\text{w}} \cdot N_{\text{IS}} \,. \tag{11.40}$$

The classical (Allen-Berkley/Borish) image source model is, thus, applicable and efficient for the following cases:

- for short impulse responses (second or third order),
- for simple geometries (small n_{w}),
- for rectangular rooms (since the audibility test can be omitted).[52]

Spherical waves at grazing incidence

Another limitation is given by the validity of Eq. (11.32). The reflection coefficient $\underline{R}(\vartheta_0) \neq 1$ in the contribution of the image source implies a constant

[51] Without strategies for excluding sources for geometric reasons, the increase is exponential.

[52] For a rectangular room, a regular lattice of image sources applies, with multiple sources (of indices in permutated order) coinciding at the lattice points. It can be shown, however, that for every receiver point, exactly one lattice source is audible.

angle of incidence and, thus, reflection of a plane wave at one reflection point. This assumption is well suited when a spherical wave[53] is considered with a large distance between the source and the wall and the receiver at the wall.

At smaller distances and corresponding grazing incidence, Eq. (11.32) carries a nondetermined uncertainty. It would then be necessary to calculate the reflected wave in the more general form.

Generally, the sound pressure at the receiver point is given by

$$p = \frac{j\omega\rho_0 Q e^{-jkr_0}}{4\pi r_0} + p_{\mathrm{IS}}, \tag{11.41}$$

whereas the standard model offers an estimate:

$$p_{\mathrm{IS,\,plane\,wave}} = \frac{j\omega\rho_0 Q e^{-jkr_1}}{4\pi r_1} R(\vartheta_0), \tag{11.42}$$

where

$$R(\vartheta_0) = \frac{\zeta\cos\vartheta_0 - 1}{\zeta\cos\vartheta_0 + 1}. \tag{11.43}$$

It includes the specific wall impedance, ζ, for locally reacting surfaces. The exact solution accounts for spherical wave propagation following Eq. (3.17).

The errors introduced by the plane-wave assumption for the impedance and the reflection factor have been investigated both by experiment and by field calculation (Suh and Nelson 1999). As a rule of thumb, we can remember that at grazing incidence ($\vartheta_0 > 60°$) and at too close distances, d, of receiver and source to the wall ($d \leq \lambda$), systematic uncertainties, Δp, of sound pressure must be considered which are clearly audible ($\Delta p_{\mathrm{IS}} > 20\%$, which corresponds to 1 dB). In the middle of a room, errors are smaller than at positions near the room boundaries. Small and medium size rooms ($50\,\mathrm{m}^3 < V < 200\,\mathrm{m}^3$) at low frequencies ($f < 200\,\mathrm{Hz}$), however, have hardly a centre area further than a wavelength from the boundaries. Here, the validity of standard (plane-wave impedance) geometrical acoustics is at its limits.

A more specific and precise uncertainty of the result in auralization cannot be predicted, since the simulation errors in the end depend not only on the room but on the type of signal. A pure tone simulation is by far more delicate than a broadband simulation over several critical bands. That walls

[53] The temporal coincidence of the wave-front phases hitting the surface is relevant here.

are not locally reacting at low frequencies and that the impedance may be distributed over the wall (by distributed screws in studs, for instance) create further complications.

11.3.4 Diffraction

Diffraction is neglected in the classical image source model (and also in ray tracing). In room acoustics, diffraction may happen for two reasons: there can be obstacles in the room space (e. g., stage reflectors), or there can be edges at surroundings of finite room boundaries. In the latter case, either the boundary forms an obstacle, such as columns or the edge of an orchestra pit, or the boundary forms the edge between different materials with different impedances (and absorption). Since diffraction is a typical wave phenomenon, it is not included in the basic simulation algorithms listed above. In the past, there were some ideas of including diffraction as a statistical feature in acoustic ray models. But the success was quite limited because the increase in calculation time is a severe problem.

In optics and radiowave physics, however, ray tracing models were generalized in the uniform geometric diffraction theory (UDT) (Kouyoumjian and Pathak 1974; Tsingos et al. 2001). Other approaches were presented by (Svensson et al. 1999), who applied the model by Biot and Tolstoy (1957) for acoustic problems. These are very powerful for determining first- or second-order diffraction. But all methods of geometric diffraction are very time-consuming for simulating a multiple-order diffraction and corresponding reverberation, and they introduce ray-splitting and an exponential factor in the multiple-scattering algorithm. Another possibility is to apply finite element or boundary element methods, of course.

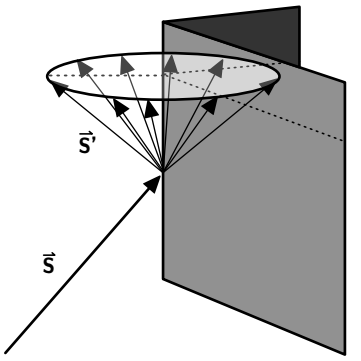

Fig. 11.23. Geometric edge diffraction using secondary sources (after (Svensson et al. 1999))

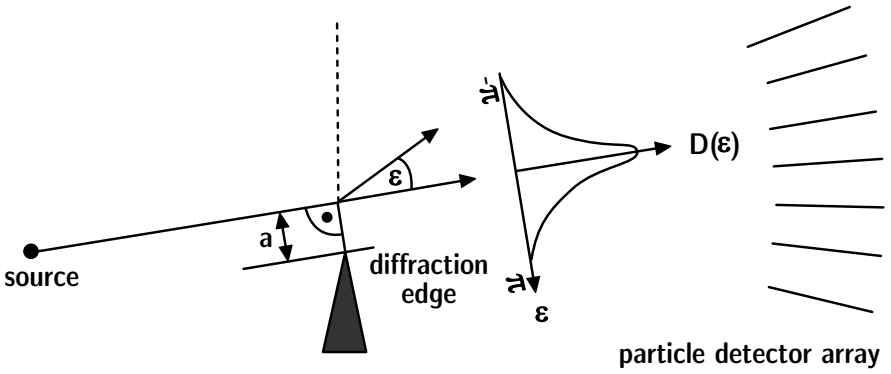

Fig. 11.24. Ray (particle) diffraction model after (Stephenson 2004b)

Stephenson (2004a, 2004b) developed two approaches for edge diffractio. One consists of the implementation of "virtual flags" mounted at the edge, which let the edges affect rays in the room interior.

The flag width is about one wavelength representing the midband frequency of the octave or one-third octave band. Rays hitting the flags can be diffracted according to statistical distribution of diffraction angles depending on the distance from the edge, the wavelength and the ray incidence angle. This distribution is derived from either slit diffraction models or Fresnel's edge diffraction models. Both methods show good results in studies, but they are, however, not perfect in the variety of all general cases in practice.

Diffraction models and their implementation in image source and ray tracing algorithms will be one of the greatest challenges in future developments of geometric room acoustic simulation methods.

11.3.5 Reduction of computational load by preprocessing

Field angle

Geometric tests help to reduce the computational load to create and check the audibility of image sources. If inconsistent sources are found, construction of daughter sources can be avoided. Mechel (2002) defines the field angle which is valid for each image source. If the receiver is not located in the field angle, the source is inaudible. Now, if a wall polygon used for creating the daughter source (see Sect. 11.3.1) is outside the field angle, the mother source cannot give birth to this daughter. The example in Fig. 11.25 illustrates that the image source S_{70} (the daughter of source S_0 and granddaughter of S) spans a field angle with the "father" wall no 7. Receivers outside the field angle do not receive sound from S_{70}. Furthermore, in this

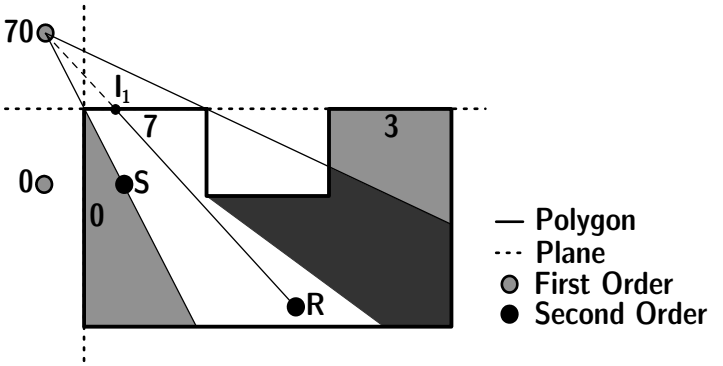

Fig. 11.25. Field angle of image souce S_{70} (reflection path with walls 0 and 7)

example, a convex corner is present and thus, the field angle is affected by a shadow region.

With these and other criteria, strategies for interrupting the creation of image sources can be reduced significantly (Mechel 2002).

Spatial data structures

Spatial data structures have been commonly used in applications of computer graphics and have already been applied in room acoustical simulation algorithms; see (Funkhouser et al. 1998; Jedrzejewski and Marasek 2004).

Different types of spatial data structures are in use, such as bounding volume hierarchies, binary space partitioning (BSP) trees and octrees. They have in common that the entire geometry is subdivided into smaller subspaces which are encoded in the data structure. In contrast to bounding volume hierarchies, which separate the scenery into single objects enclosed by bounding volumes such as spheres or boxes, BSP and octrees subdivide the scenery into subspaces without taking into account any information about single objects.

Octrees divide the scenery regularly along all three axes. Hence, every division of space results in eight new boxes. BSP trees allow a flexible partitioning of space by means of arbitrarily shaped partitioners. Both methods satisfy the same type of queries. For simple intersection tests, the BSP structure is more appropriate than octrees due to its higher flexibility.

Spatial subdivision helps to speed up required intersection tests, a fast point-in-polygon test and an efficient data structure with respect to real-time auralization to provide a faster determination of the BRIR. This goal is illustrated further by discussing the example of BSP (Shumacker et al. 1969).

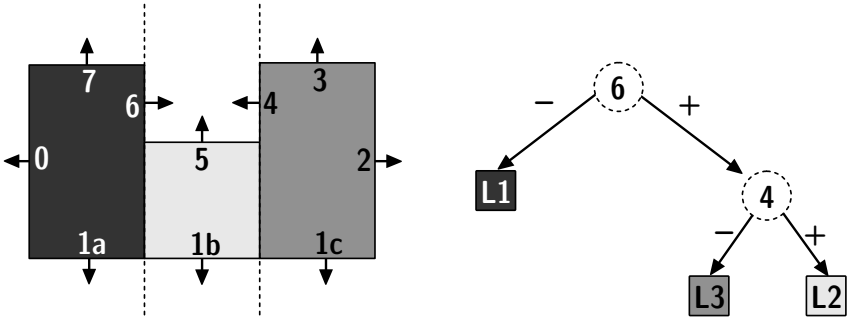

Fig. 11.26. Example of a BSP tree for a given room geometry (after (Shumacker et al. 1969; Schröder and Lentz 2006))

Binary Space Partitioning

The aim of BSP is speeding up the determination of the location of an arbitrary point relative to a geometric scene which is encoded by a so-called "BSP tree."

It is preferable to use planes spanned by polygons as partitioners (Fig. 11.26). Each tree node contains one partitioner that divides the current subspace into two smaller subspaces. In the problem of finding the next reflection point, the point in question can have three possible relative locations: 1) on, 2) in front of, and 3) behind the partitioner. In the first case, the query can be stopped. To pursue the latter two cases, the test is continued by branching to the respective son of the node, left for "behind," right for "in-front." The tree's root node therefore refers to the whole space, while leaf nodes refer to a subspace which does not contain any further scene detail. Contrary to the naive approach to test the position of the point against all planes, the tree structure allows us to determine the position by testing only a subset of planes. This subset is defined by the path in the tree. By using a balanced tree, i. e., a tree of minimum height, the number of tests can also be minimised, which drops the complexity from $O(N)$ to $O(\log_2 N)$, where N is the number of polygons.

Voxels

Like octrees, voxels are volume boxes. They are created for subdividing the actual room volume, by using cubes, for example. In one pass of preprocessing, each voxel is tested for which walls it intersects. Later, in the ray tracing process or visibility test, we look for the next intersection point with a polygon. This test can be speeded up when the complete set of polygons is not tested but only polygons in voxels along the straight line

representing the ray or image path. Stephenson (2006) derived an analytic equation for the gain in computation that results in

$$t_{\text{calc}}\big|_{\text{voxel}} \propto \sqrt{n_w} \tag{11.44}$$

which was a previously linear increase without using voxels (Eq. (11.22)).

11.4 Hybrid image source models (deterministic ray tracing)

Ray tracing and image source algorithms have opposite advantages and disadvantages. Therefore it is worthwhile to develop a combination of both to obtain

- fine temporal resolution in sampling rate quality,
- inclusion of scattering, and
- faster audibility check of image sources.

This combination is called a "hybrid method." The term "ray tracing" also in use, however, is somewhat confused with the stochastic ray tracing described above. The key to combining the methods is the audibility test of image sources in the forward direction.

> If we run a specular ray tracing process and find a receiver hit by a ray, the corresponding image source must be audible.

Variants of this approach are often summarized as "ray tracing." Specific algorithms are also known as "cone tracing," "beam tracing," "pyramid tracing," etc. All have in common the idea of a forward audibility test of specular reflections, represented physically by image sources. The specific physical reason is the fact that the energy contributions of the reflections are calculated by using Eq. (11.36). The main difference between this type of ray tracing and the previous stochastic ray tracing method is the way of detecting sound energy.

Of course, in a running ray tracing process, rays hitting the receiver can have the same history of wall reflections. In this case, we observe energetic overlap between rays and it must be ensured that each image-source reflection is counted just once, for instance, by checking the indices in the wall list. Exactly this point is the reason for the differences in the variants listed above. In the various concepts, rays are extended toward geometrically diverging objects such as cones or pyramids.

This way it is also ensured that image source contributions are counted only once. While cone or beam tracing still creates slight problems with overlap (Vian and van Maercke 1986), pyramid tracing using the wall

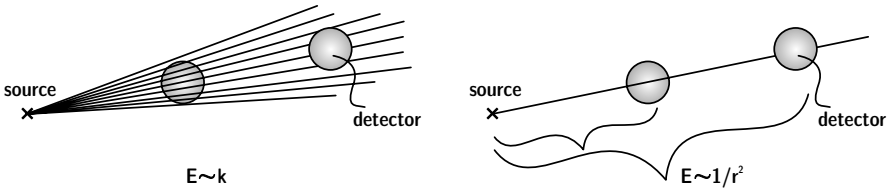

Fig. 11.27. Energy detection in stochastic (left) and deterministic (right) ray tracing

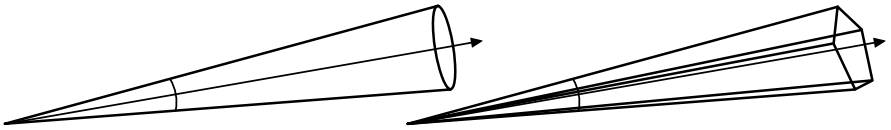

Fig. 11.28. Cone (left) and pyramid (right) tracing

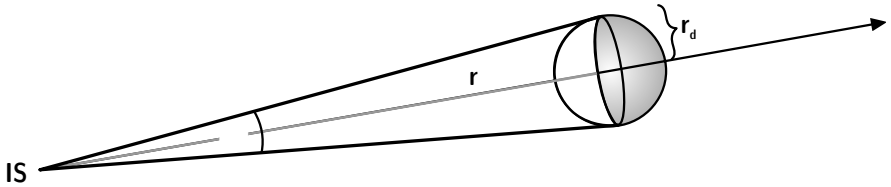

Fig. 11.29. Geometry of source, cone and receiver

polygons as pyramid base avoids double-counts (Stephenson 2004b). At the same time, with extended and diverging ray geometries, a point-like receiver can be used.

The required number of rays or cones depends on the maximum distance between image source and receiver and, thus, on the order of the image source and the corresponding average delay in the impulse response, as illustrated with the example of a cone.

The figure shows the path between image source and receiver in an expanded version.[54] N cones are started into the full space of 4π. From these,

$$k = N\,\Omega/4\pi \tag{11.45}$$

will hit the receiver, with Ω denoting the spatial angle of the cone:

$$\Omega r^2 = \pi r_d^2 \tag{11.46}$$

[54] The real path is folded from reflection to reflection. The total path length is identical in the expanded version.

for $r \gg r_d$. Hence, the probability of cones hitting the receiver is

$$k = \frac{N r_d^2}{4r^2} = \frac{N r_d^2}{4(ct)^2}. \tag{11.47}$$

To ensure that at least one ($k = 1$) cone hits the receiver after time t_{max},

$$N_{min} = \frac{4(ct_{max})^2}{r_d^2}. \tag{11.48}$$

The computation time of this kind of ray tracing (see also Eq. (11.24)),

$$t_{calc} = \frac{4c^2 \bar{n} t_{max}^3}{r_d^2} \tau, \tag{11.49}$$

increases with the third power of the impulse response length.

In the last few decades, several authors have presented strategies to offset the computational load of the image source method (Mechel 2002; Funkhouser et al. 1999; Stephenson 2004a). Other authors have presented interesting approaches for generalizing the image source principle with regard to diffraction and scattering (Martin and Guignard 2006). It should be stressed again that the standard image source method is based on non-absorptive walls or, at least, on absorption in the plane-wave approximation. Approaches or extensions which do not consider wave theoretical extensions still suffer from limits set by the theory of specular reflections.

Before we focus on strategies for speeding up the simulation algorithms for auralization, the physical limits of room acoustics simulation by ray

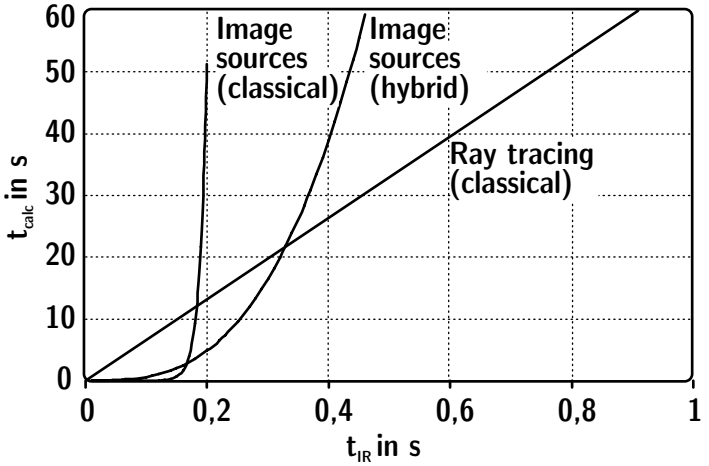

Fig. 11.30. Computation times of room acoustic simulation algorithms

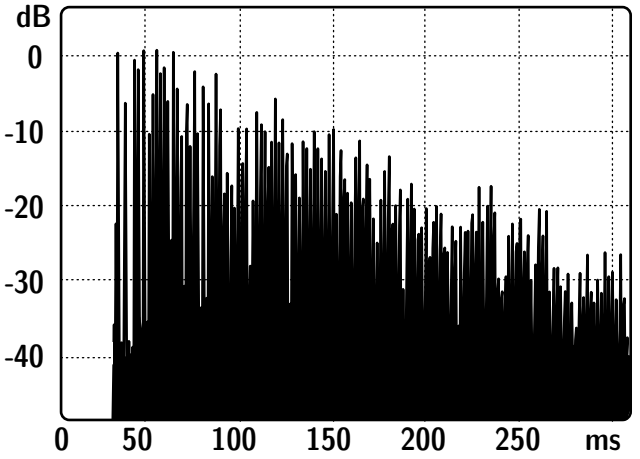

Fig. 11.31. Results with a hybrid image source algorithm

tracing and image sources should be discussed. Uncertainties of simulation methods applied in practice were checked and documented. The basic material data available and the necessary psychoacoustical rating of the results must also be considered.

11.5 Systematic uncertainties of geometrical acoustics

Geometrical acoustics (ray tracing, image sources) allows simulating sound fields in rooms and outdoors. These models yield correct results under specific circumstances (mostly academic cases) and fulfil the wave equation. But absolute physical correctness is not always a reasonable goal. In most cases, geometrical acoustics offers approximate solutions, solutions which are sufficient in practice. But it should be kept in mind that the limits are

– large rooms, nongrazing incidence,[55]
– low absorption coefficients, and
– broadband signals.

Room acoustical results such as reverberation time, clarity, strength, etc., are usually expressed in relation to frequency bands. The reason for using frequency bands is that sound signals related to room acoustics are of broadband character, such as speech and music. Also from the physiological point of view, one-third octave bands are a good compromise since they are an approximation for critical bands (Sect. 6.2.1). These arguments

[55] or tangential modes

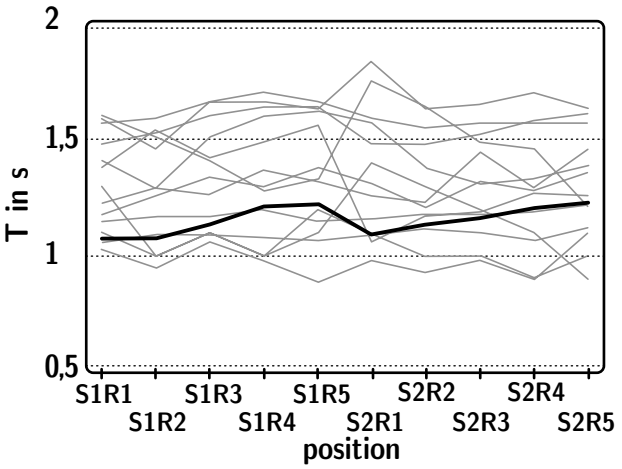

Fig. 11.32. Round-robin I, PTB auditorium (Vorländer 1995). Reverberation times calculated and measured (thick line) for source/receiver combinations

are the justification for using energetic methods (neglecting phases). At low frequencies, however, the situation is different. Transmission of pure tones, possibly the fundamental frequency of musical instruments, is surely of interest in optimizing the placement of recording arrangements in studios or to find optimum loudspeaker positions. Simulation and auralization of harmonic signals in such cases may largely differ from the sound pressure in the real case.

In 1995, room simulation software solutions were tested in an international round-robin project. Eight parameters defined in ISO 3382 (T, EDT, D, C, T_S, G, LF and LFC) were calculated at two source positions and five receiver positions in a test room.[56] At first, the results were compared with measurement results in only one frequency band.

The simulated reverberation times of the first phase of the project were generally too large, thus indicating that the users underestimated absorption. Then, after performing measurements, the absorption coefficients were published and the software was run again with harmonized input data. The agreement was much better, of course, but now the efficiency of the software could be identified. The overall accuracy of the computer simulations can be estimated by a single number rating. This quantity is defined by using the absolute difference between the result from the simulation and the measurement on each of the 10 source and receiver combinations. The average of these location-dependent differences is then related to the jnd

[56] Auditorium of the Physikalisch-Technische Bundesanstalt, Braunschweig, Germany.

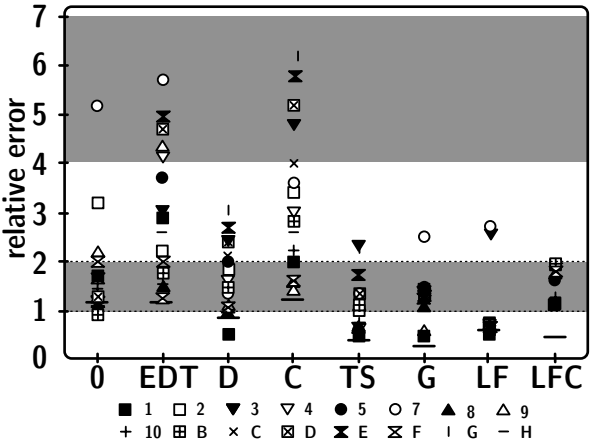

Fig. 11.33. Overall results of the "Round-Robin I" (Vorländer 1995). The ordinate scale refers to a normalized deviation from the jnd[57]

for the respective acoustic criterion. The resulting relative error for each acoustical parameter is shown in Fig. 11.33. Note that a value of 1 indicates the order of magnitude of the jnd. Only 3 out of 17 programmes could deliver results within similar accuracy as given by the standard deviation of the measurements and by the just noticeable differences (Table 6.1).

In the second intercomparison project, Bork (2000) investigated the results of simulations in a larger room and for more frequency bands. The example room was the ELMIA hall in Jönköping, Sweden. The intervals of the jnd related to the specific psychoacoustic dimensions were refined. The results generally supported the conclusion from 3 years before, that simulation algorithms require a module for treatment of scattering. Also, in the second project, one CAD room model was provided to run software with identical input data.

Finally the third round-robin was focused on a smaller room with expected modal effects. The results again support the fact that a) a good scattering model is essential and b) the overall prognosis accuracy, when results are compared with experimental results, can be in the range of 1 to 2 jnd. Programmes are available for achieving this accuracy, so that a certain quality of he results can be guaranteed. The operator of the software, however, still has crucial influence on the choice of absorption and scattering coefficients.

[57] At that time, the jnd for clarity was assumed to be around 0.5 dB. Later, it was shown that 1 dB is a better approximation; see also Table 6.1.

After all, it has been proven that room acoustical computer simulation yields reliable results. Scattering was identified as an important factor. With careful data choice of absorption and scattering coefficients, it is possible to obtain good prediction results which deviate from experimental results by the same order of magnitude as the uncertainty of the experiment or the jnd.

11.6 Hybrid models in room acoustics

Deterministic models such as the image source model suffer from inherent systematic errors and from limitations in software implementation. Stochastic models suffer from poor temporal resolution, but they can handle scattering. Impulse responses from image-like models consist of Dirac pulses arranged according to their delay and amplitude and sampled with a certain temporal resolution. However, in the intercomparions of simulation programs (see Sect. 11.5), it soon became clear that pure image source modelling would create too rough an approximation of physical sound fields in rooms, since a very important aspect of room acoustics – surface and obstacle scattering – is neglected. This fact is further supported by the observation that scattered energy dominates the reverberation process after a few reflections (order of 3 or 4), even in rooms with rather smooth surfaces (see Fig. 11.34 and (Kuttruff 1995)). Also the audibility of the characteristics of diffuse reflections was proven by (Torres et al. 2002).

Fortunately, the particular directional distribution of scattered sound is not relevant in the first place and can well be assumed to be Lambert scattering. Solutions to the problem of surface scattering are given by either

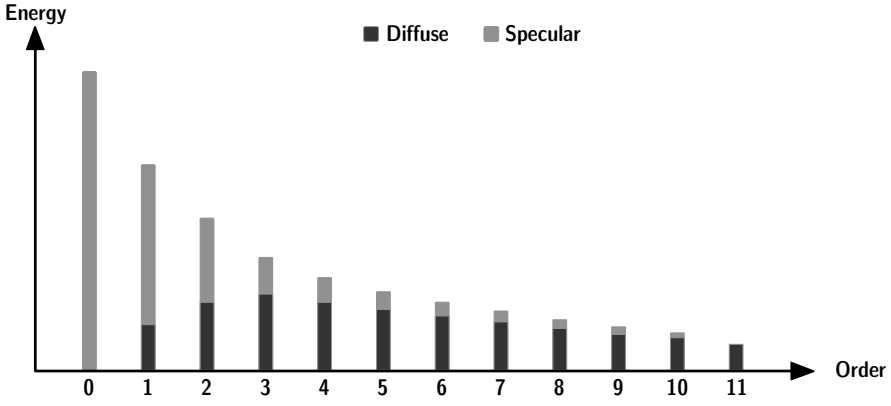

Fig. 11.34. Conversion of specularly into diffusely reflected sound energy, illustrated by an example (after (Kuttruff 1995))

stochastic ray tracing or radiosity. Another effect of wave physics – diffraction – can be introduced into geometrical acoustics in principle. Big problems, however, arise from extending diffraction models to higher orders (Svensson et al. 1999). In applications other than outdoor noise propagation,[58] diffraction models have not yet been implemented. New algorithmic schemes, such as presented by Stephenson (2004b), have not yet been implemented, too. It should be mentioned here that numerous algorithmic details have been published in the field of wave field simulation, until today. Sound transmission and diffraction, too, must be implemented in cases of coupled rooms or in room-to-room transmission; see Chap. 12.

11.6.1 Hybrid deterministic-stochastic models

The general principles of room acoustic modelling were discussed in the previous sections. They yield results of high temporal resolution (on scales of μs), typically represented by reflections of directly used or implicitly involved image sources. Due to numerical constraints and also because it is unnecessary to create such a high resolution for the late reverberation (> 100 ms), stochastic methods such as ray tracing and radiosity move into the focus of our interest. It is obvious to combine the tedious but exact deterministic methods with a quick estimate of late reverberation.

In the following, some well-known software solutions are briefly discussed. This overview is not intended to cover all kinds of software used today. The examples are particularly listed since they represent the historic process of software development for room acoustic computer simulation using hybrid models in the 1990s.

CATT-acoustic

In his thesis, Dalenbäck (1995) studied aspects of an extended radiosity approach merged with specular reflection techniques (see also (Dalenbäck 1996)). Implementation of the software already began in the late 1980s and has been continuously improved and extended. The main model consists of predicting high-order reflections using randomized tail-corrected cone-tracing and direct sound, first-order specular and diffuse and second-order specular-to-specular reflections using deterministic methods and image source sources and are thus created with full temporal resolution and bandwidth. Higher order reflections are the result of independent ray/cone-tracing for each octave-band taking into account the frequency dependence of diffuse reflections via randomly selecting specular or diffuse via the

[58] almost free-field propagation with at most one diffraction edge.

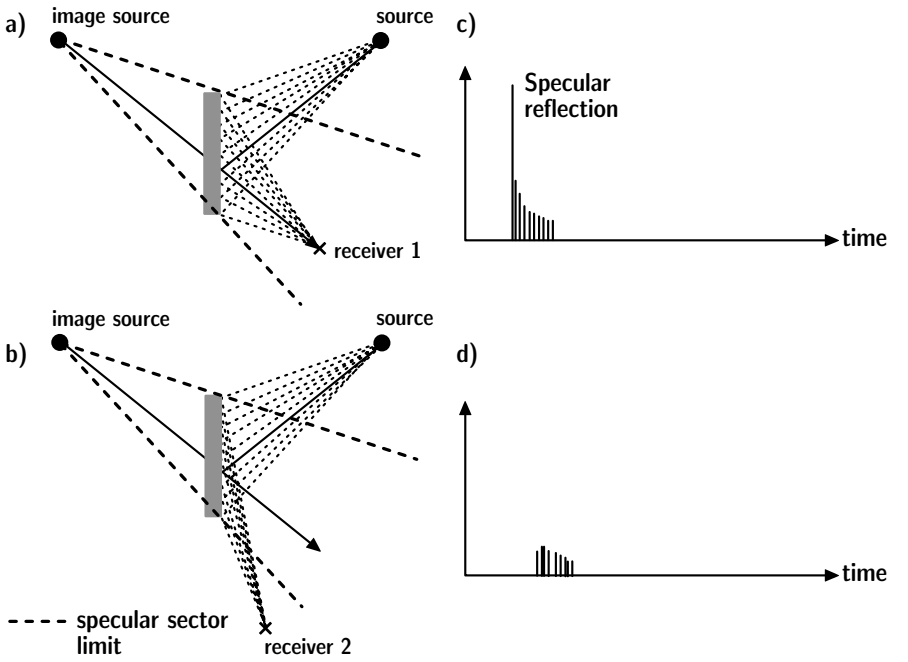

Fig. 11.35. Scattered energy radiated from surface patches in CATT. a) First-order diffuse and specular reflections. b) First-order diffuse reflections. c) Schematic echogram at receiver 1 (direct sound omitted). d) Schematic echogram at receiver 2 (direct sound omitted)

scattering coefficient magnitude where a diffuse ray is reflected using the Lambert distribution. Dalenbäck and McGrath 1995) also presented tools for dynamic auralization using headphones with head-tracking in 1995.[59]

ODEON

The first version of ODEON was published in Naylor (1993). It is a hybrid image source model with a stochastic scattering process using secondary sources. The secondary sources are assigned a frequency-dependent directionality, the so-called reflection-based scattering coefficient; this implies that the direction of the ray that created the secondary source is taken into account. The polar scattering patterns are created from a vector-based process which adds Snell's (specular) and Lambert's (diffuse) into one final scattering direction for each ray (Rindel and Christensen 2003).[60] The process of creation of secondary sources is shown in Fig. 11.36 (top) for

[59] http://www.catt.se
[60] http://www.odeon.dk

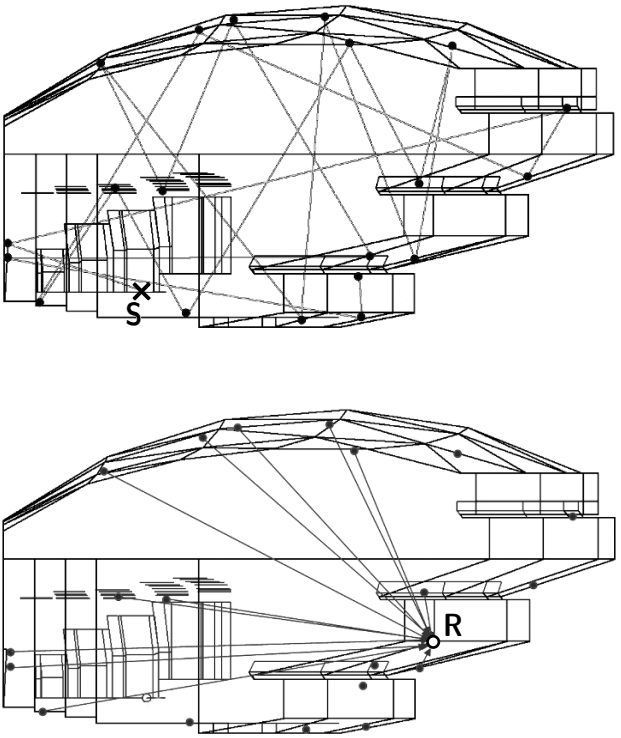

Fig. 11.36. Secondary source model in ODEON (after (Rindel 1993)). S is the source, R the receiver point. Top: Creation of secondary sources. Bottom: Only the secondary sources with free line of sight contribute to the sound pressure at the receiver on the balcony

the example of one single ray. The irradiation of sound from the secondary sources to a receiver is shown in Fig. 11.36 (bottom) for the example of a receiver on the first balcony. Some of the secondary sources cannot contribute to this receiver due to the audibility check.

QPBT

QPBT is the abbreviation for quantized pyramidal beam tracing (Stephenson 1996, 2004). The geometric principle of the method is pyramidal beam tracing based on the wall polygons that form the pyramidal base area. Successive reflections are created by extending and splitting the pyramids toward the other wall polygons. The problem of exponential increase of pyramids is inherent in the geometric imaging method. In QPBT, this is solved by unifying spatially closely located pyramids. Furthermore, by

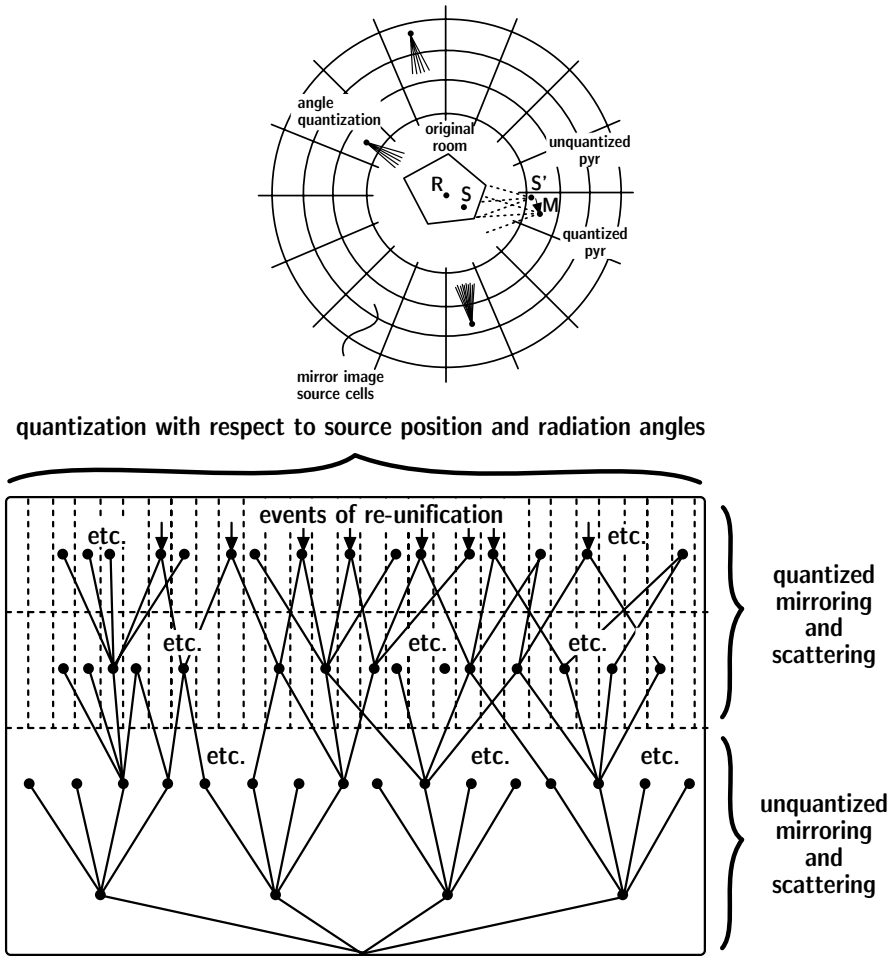

Fig. 11.37. Quantized pyramidal beam tracing and corresponding pyramid tree (after (Stephenson 1996))

also discretizing the time frame, it is possible to recombine sound energy into finite elements of time and spatial angle. The particular way of energy travelling via a pyramid to the discrete element is irrelevant. The model is thus open for inclusion of scattering and edge diffraction. In which way edge diffraction is implemented best in the pyramidal beam was actually tested in Stephenson and Svensson (2007).

EASE

The software EASE stems from the application of simulating installations of professional audio and sound reinforcement systems (Ahnert and Feistel

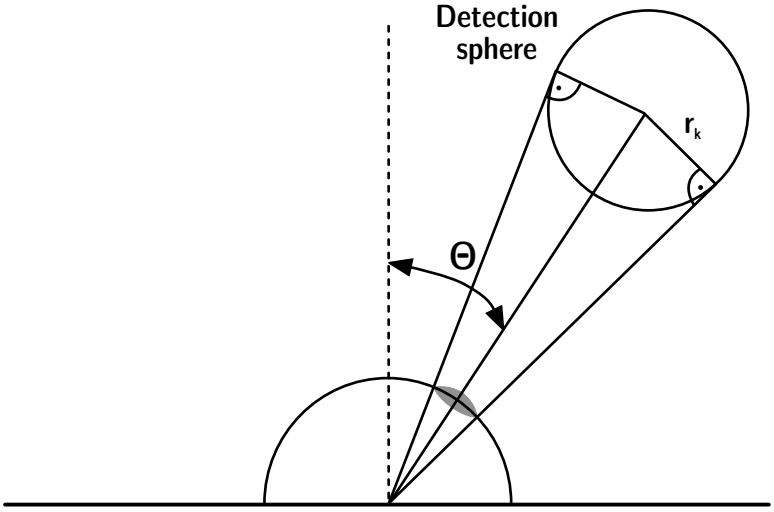

Fig. 11.38. Diffuse rain implemented for calculating the portion of energy scattered to a detector sphere (see (Heinz 1993) or the recently extended version in (Schröder et al. 2007); see also Fig. 15.5)

1991).[61] The database of loudspeaker frequency responses and directivities is hence very large, and it includes innovative approaches for detailed and accurate loudspeaker representation. It was implemented at first in an image source model capable of simulating particularly the direct sound and the early response.

In a later version of EASE, the room acoustic simulation was extended by integrating elements from a hybrid image source/ray tracing model called "CAESAR" (Vorländer 1989; Schmitz et al. 2001). The key to the latter method is parallel operation of image source and ray tracing algorithms. This delivers the specular reflections precisely and the diffuse reflections via the stochastic approach ("diffuse rain" after (Heinz 1993)) with lower spectral and temporal resolution in the reverberation tail. The main feature of this method is a closed solution of an inherent transition from early specular to late specular/diffuse reflections. It is not required to define transition orders or secondary sources.

State of the art

Still today, the programmes listed above are subject to permanent development of the components of impulse response computation and other elements such as source and material databases. Most of them also offer

[61] http://www.ada-acousticdesign.de/set_en/setsoft.html

signal processing tools, visualization tools and, auralization tools in various reproduction formats.

More recently, algorithms of geometric room acoustics modelling were also published by (Farina 1995; Alarcão and Bento Coelho 2003; Lokki 2002; Camilo et al. 2002), besides others. Many of these programmes were inspired by (Kulowski 1985; Stephenson 1985; Vian and van Maercke 1986; Vorländer 1989; Lewers 1993). In some methods. building room impulse responses from reflection statistics was tried (Vorländer 2000; Bento Coelho et al. 2001). Due to the rapid development in acoustic room simulation also, this list cannot be complete.

Now, the basics of hybrid room acoustic modelling techniques are known, so that we can discuss the next step, creation of impulse responses for auralization. This is reasonable, including binaural technology.

11.7 Construction of binaural room impulse responses

Impulse responses suitable for signal processing and particularly for auralization must have a sampling rate appropriate for the audio frequency range, typically about 40 kHz. Thus, image source algorithms or hybrid models are applicable. The impulse responses are fed into FIR filters for convolution with dry source signals.

Room auralization must be based on binaural hearing. Otherwise, the spatial information would be lost. By using the HRTF (see Sect. 6.3.1), the directional information of a human listener is taken into account.

During the historical development of auralization of rooms, the first studies were based on purely specular reflections. After the principles were proposed by Schroeder et al. (1962) and Schroeder (1973), one of the first who created a signal processing concept for room acoustics auralization was Pösselt (Pösselt et al. 1986), who used image source models for rectangular rooms.[62] The model was further extended by (Lehnert and Blauert

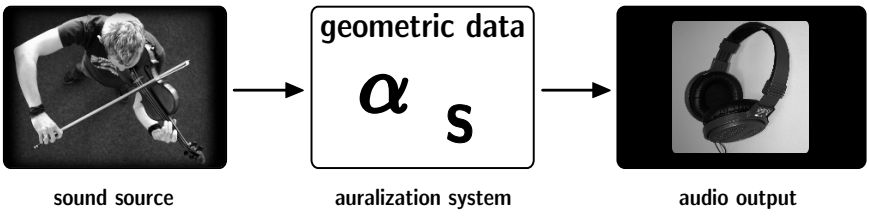

Fig. 11.39. Auralization of sound in rooms

[62] The reason for this was clearly the limitation by processing time in the late 1980s.

1989, 1992; Lehnert 1992; Vian and Martin 1992). The purely specular reflection model, however, was not sufficient due to lack of scattering, as we know today. But his important and novel contribution was the rigorous description of the binaural room impulse response based on spatial direct sound and reflections. With the deeper understanding of the details of room impulse responses and their relevance for auralization available, extended model approaches, including hybrid models and scattering models in particular, have been developed.

Finally, a binaural impulse response is created from direct sound, early reflections and scattered components by using the concept of binaural synthesis (Sect. 9.3). All components are added. When formulating in the frequency domain,[63] this process is described by multiplication of the transfer and filter functions representing sound travelling from the source to the receiver:

$$\underline{H}_j\Big|_{\text{left,right}} = \frac{\mathrm{e}^{-j\omega t_j}}{ct_j} \cdot \underline{H}_{\text{source}}(\theta,\phi) \cdot \underline{H}_{\text{air}} \cdot HRTF(\vartheta,\varphi)\Big|_{\text{left,right}} \cdot \prod_{i=1}^{n_j} \underline{R}_i \quad (11.50)$$

where H_j denotes the spectrum of the jth reflection, t_j its delay, $j\omega t_j$ the phase, $1/(ct_j)$ the distance law of spherical waves, $\underline{H}_{\text{source}}$ the source directivity in source coordinates, H_{air} the low pass of air attenuation, $\underline{R}_i$ the reflection factors of the walls involved, and HRTF the head-related transfer function of the sound incidence in listener coordinates at a specified orientation.[64]

Since data for air attenuation and particularly absorption coefficients of walls are available in frequency bands, the spectral representation required for Eq. (11.50) must be created by interpolation.

Interpolation

Several interpolation algorithms are possible candidates. The cubic spline interpolation is one example of a well-qualified method to create interpolated data. For interpolation of reflection factors from octave band data (see also Sect. 11.1.2), the effects of interpolation in the frequency and time domains must be discussed. Too rough interpolation may cause audible ringing. Onset and decay of the interpolation filter creates a kind of tonal reverberation.

---

[63] In the time domain, the same can be formulated, of course, by convolution, but the spectral product terms are easier to interpret.

[64] Usually the listeners are oriented toward the original source or a similar reference point.

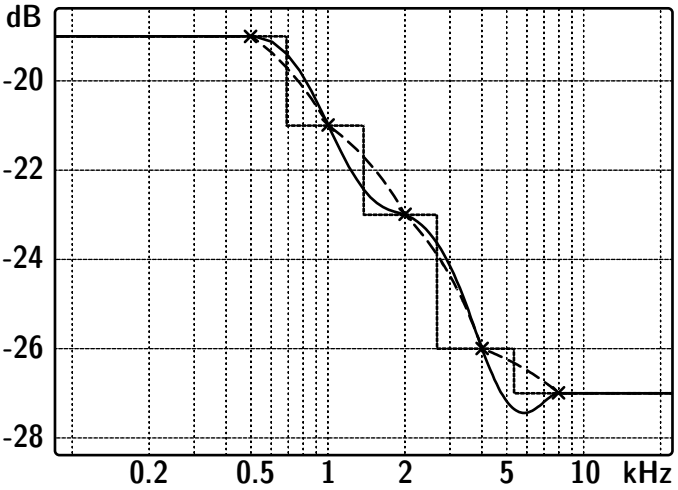

Fig. 11.40. Step (dotted line), linear (broken line) and spline interpolation of spectra based on octave band data (example)

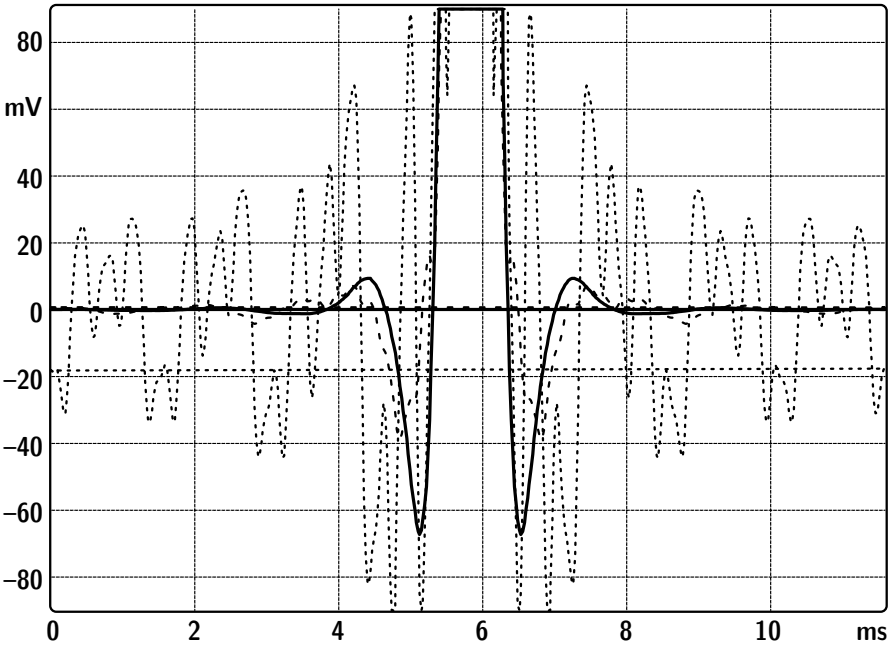

Fig. 11.41. Result of an interpolation of reflection factors created from one-third-octave band absorption coefficients. Plot in time domain (maxima clipped). Spline interpolation between one-third octave bands corresponds to the continuous line. Too rough interpolation (step function) corresponds to the dotted line (ringing effects)

Mommertz (1996) showed that absorption coefficients in one-third octave bands serve well for reconstructing the complex reflection factor. While the modulus of the reflection factor can be obtained directly, a plausible phase must be added. Minimum or linear phases are options. In only very few cases such as for focuses, the specific choice will be important.

After inverse Fourier transformation, a spectrum consisting of M complex coefficients corresponds to a signal of M samples. Convolution of signal and transfer function of M lines each in the frequency domain creates a result of M lines, too. In contrast, in the time domain convolution, the result contains $2M$ samples.

Late reverberation

As explained above, scattering plays an important role particularly in the late response. Apparently, Eq. (11.50) 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 $\underline{H}_j$, whose arrangement 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, $M^2(\Delta t, f)$, a function of frequency and time[65]. By adding an adequate fine structure which represents the actual reflections statistics, the binaural impulse responses can be created (Heinz 1993).

The parameters of the fine structure are the average amplitudes, $\underline{H}_j$, and the density, n_j, of the reflections in the time interval Δt. The set of reflections yield the correct modulation if

$$M^2(\Delta t) = \left|\underline{H}_j\right|^2 \cdot n_j \tag{11.51}$$

for all frequency bands.

The choice of the specific algorithm for estimating the late response, the bandwidth, the size of the time intervals and spatial resolution of the HRTF are subjects of research and development in many places. It is clear that fine details of the late response can usually not be perceived. If, however, the room creates an echo in the late response and this was not detected and modelled in the simulation, the auralization will miss an important feature of the room. Hence there exists no simple rule on how the parameters of the late response and the transition time between the early and the late response must be chosen.

[65] similar to a modulation function.

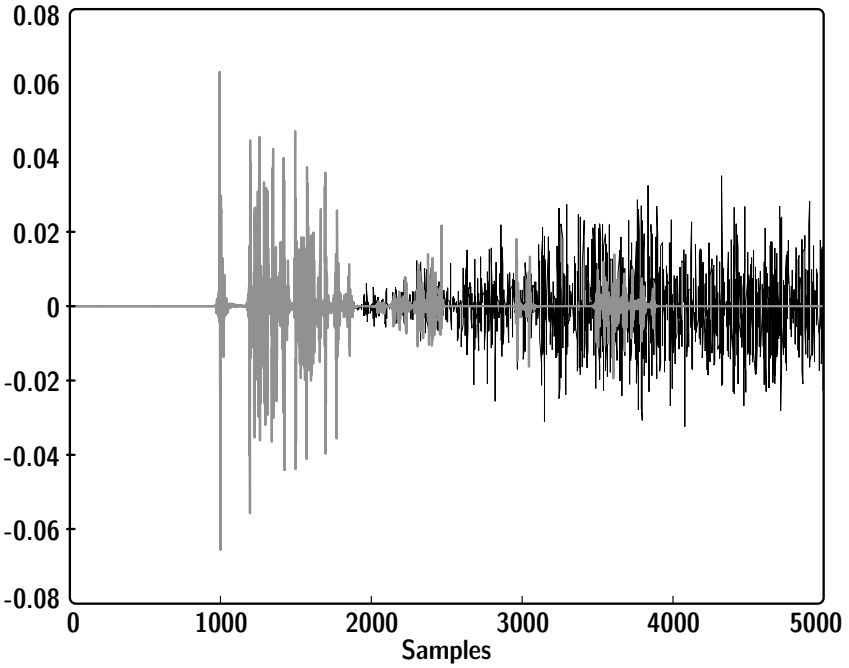

Fig. 11.42. One channel of a binaural room impulse response created from specular (grey) and diffuse (black) reflections (example after (Schröder et al. 2007))

Many more details of signal processing are found in the specific programmes listed in Sect. 11.6.1 above. For further information on these details, refer to the Internet links listed in Sect. 11.6.1.