# Chapter 3
# The Listening Environment Model

## 3.1 Introduction

It is well known that the acoustic environment is a determining part of the subjective quality of listening: the presence of reverberating tails has a significant influence on the vocal and/or musical message. However, the acoustic propagation in confined spaces, as seen in §1.7 is a complex phenomenon.

The scientific approach to room acoustic started in the 1890s, when Wallace C. Sabine, a physics professor at Harvard University, established the concept of *reverberation time* (RT). Sabine defined reverberation time, denoted as $T_{60}$, as the time for a decay of 60 [dB] after a stationary sound source has been stopped [1]. However, around the year 1900, for the determination of exact and repeatable measurement, there were no microphones, electronic devices and computers of today's technology.

This chapter deals with the acoustic modeling of confined listening environments. In the first part of the chapter, the natural modes of a room and its characterization, as a linear dynamic system, through the *Room Impulse Response* (RIR) are defined. Subsequently, the main measurement techniques for RIR estimating are illustrated and the main macroscopic indices for its simple characterization are defined. Finally, the main models for the *a priori* calculation of the RIR, the acoustic room correction and the acoustic quality indices of a listening environment, are briefly reported.

This modeling-approach concept is very important in modern acoustics as it lays the foundation for:

- correct measurement of the RIR. The RIR can be estimated by exciting the acoustic environment with a maximally informative signal, i.e., which has an impulsive autocorrelation function, and acquiring the system's output signal, i.e. direct signal and the one reflected from the walls with a omnidirectional measurement microphone [9]-[13].
- determination of geometric methods (such as the *image method* and the *ray-tracing method*, described in this chapter) able to evaluate *a priori* the RIR of an environment starting from its geometric description [21]-[31].

From the RIR (measured or *a priori* determined by simulation), is possible to define some significant, parameters to characterize the acoustic quality of the listening environment [32]-[38].

## 3.2 Acoustic Room Model

The acoustic environments are fully described by the wave equation (see §1.7). However, the solution of wave equations, for some boundary and initial conditions, represents a complex problem that often produces results that are not very generalized and useful for their practical use. Thus, in computational acoustics, we prefer to use simpler models and describe the acoustics of a room through a *transfer function* (TF) or through its related impulse response.

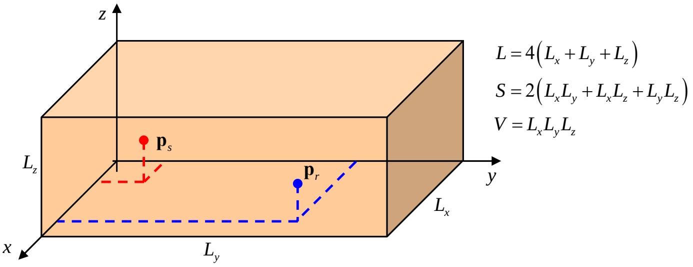


**Fig. 3.1** Rectangular room: the simplest model for real-life enclosed environments. The results obtained for perfectly rectangular room, with rigid wall, can be applied at least qualitatively to many rooms encountered in practice.

The TF represents a point-to-point causal model. A simplified models, as that illustrated in Fig. 3.1, are often used for simplified analysis in acoustics. In the model $\mathbf{p}_s = [x_s, y_s, z_s]$ represents the coordinates of the point where a source is inserted, e.g. an omnidirectional speaker (the cause); while $\mathbf{p}_r = [x_r, y_r, z_r]$ represents the coordinates of the point where the receiver is inserted, e.g. an omnidirectional microphone, (the effect).

### *3.2.1 Room Transfer Function*

An analytical solution to the wave equation, see Eqn. (1.114), is hardly available in closed form, except for simple excitation signal. The acoustic environment is considered, at least as a first approximation, a *Linear Time Invariant* (LTI) dynamic system, and it can be entirely characterized by a *Room Transfer Functions* (RTF) in terms of poles and zeros [5]-[8].

#### 3.2.1.1 Continuous-Time RTF Model

The RTFs of 3-D problems in the $s$ domain can only be obtained by eigenfunction expansion of both source and measured sound pressure, in the form of an *infinite* series of second-order rational terms, which in certain contexts are denominated as Green's functions, of the type

$$H_{RTF}(s) = \sum_{k=1}^{\infty} \frac{H_k(\mathbf{p}_s) H_k(\mathbf{p}_r)}{\left(1 + \dfrac{s}{\Omega_k Q_k} + \dfrac{s^2}{\Omega_k^2}\right)} \tag{3.1}$$

where $H_k$ is the $k$-th eigenfunction of the room, evaluated in the positions of the source and the receiver, respectively $\mathbf{p}_s$ and $\mathbf{p}_r$; $\Omega_k = 2\pi f_k$, $f_k$ being the $k$-th *eigenfrequency* of the room, and $Q_k$ the quality factor (corresponding to the damping-factor). The parameters $Q_k$ and $\Omega_k$ are independent of the source and receiver positions; their values are determined by the room size, wall reflection coefficients, and room shape. (see Eqn. (1.24)).

***Remark 3.1.*** Note that, the eigenfrequencies are sometimes referred to as *resonance frequencies* because of some sort of resonance occurring in the vicinity of those frequencies: these together implicitly identify a couple of complex conjugate poles ($p_k$, $\bar{p}_k$) in the $s$-plane, corresponding to the $k$-th room mode.

If, for some reasons, the eigenfunction expansion in (3.1) could be bound to just a finite number of terms, its expression could be re-arranged as a rational function, that can be written in terms of poles and zeros as

$$H_{RTF}(s) = \frac{B(s)}{A(s)} = \frac{\sum\limits_{k=0}^{Q} b_k s^k}{1 + \sum\limits_{k=1}^{P} a_k s^k} = G \cdot \frac{\sum\limits_{k=1}^{L_N} \left(1 + \dfrac{s}{\Omega_{zk} Q_{zk}} + \dfrac{s^2}{\Omega_{zk}^2}\right)}{\sum\limits_{k=1}^{L_D} \left(1 + \dfrac{s}{\Omega_{pk} Q_{pk}} + \dfrac{s^2}{\Omega_{pk}^2}\right)} \tag{3.2}$$

where $q_k$ are the “zeros” that model the anti-resonances and time delays in the RIR, $p_k$ are the “poles” associated with room resonances[1] and where $G$ is a scalar parameter and with a denominator depending only on the $L_D$ “room modes” (i.e. not on the specific positions of the source $\mathbf{p}_s$, nor on the receiver $\mathbf{p}_r$ which, together, define the particular considered RTF in the room).

***Remark 3.2.*** Considering the $\mathcal{L}^{-1}$-transform of Eqn. (3.2) we obtain an *ordinary differential equation* (ODE) with constant coefficients

$$\sum_{k=0}^{P} a_k \frac{d^k y(t)}{dt^k} = \sum_{k=0}^{Q} b_k \frac{d^k x(t)}{dt^k}$$

which represents a *continuos-time* (CT) linear model.

#### 3.2.1.2 Discrete-Time RTF Model

In discrete-time domain the conventional parametric models for room acoustics aims at approximating the Eqn. (3.2) as a $z$-domain rational function of the type

[1] Note that the RTF can also be factorized as: $H_{RTF} = b_n + \sum_{i=1}^{P} \sum_{k=1}^{n_i} \frac{R_{ik}}{(s+p_i)^k}$, where $n_i$ is the multiplicity of $i$-th pole, and $R_{ik}$ are the residues.

$$H_{RTF}(z) = \frac{B(z)}{A(z)} = \frac{\sum_{k=0}^{Q} b_k z^{-k}}{1+\sum_{k=1}^{P} a_k z^{-k}} = G\frac{\prod_{k=1}^{Q}(1-q_k z^{-k})}{\prod_{k=1}^{P}(1-p_k z^{-k})} \tag{3.3}$$

which transformed in discrete-time (DT) domain corresponds to the following *finite differences equation* FDE

$$y[n] = \sum_{k=0}^{Q} b_k x[n-k] + \sum_{k=1}^{P} a_k y[n-k] \tag{3.4}$$

and where $q_k$ are the zeros that model the anti-resonances and time delays in the RIR, and $p_k$ are the poles associated with room resonances.

***Remark 3.3.*** The Eqn. (3.2) is usually denoted as *Auto-Regressive and Moving-Average* (ARMA) model. The name comes from the fact that in the DT-domain the numerator from Eqn. (3.3), performs a “weighted moving average”, with the $b_i$ coeffcients of the a input samples $b_0x[n]+b_1x[n-1]$, ...; while the denominator performs an weighted average, with the $a_i$ coeffcients of past outputs samples $a_1y[n-1]+a_2y[n-2]$, ...; which is usually indicated as “auto-regressive”.

The estimation of the parameters of the Eqn. (3.3) model, and its possible factorizations and decompositions, represents a central theme of the DASP that will be extensively dealt with with different purposes and methodologies in the following of this text.

### 3.2.1.3 Common RTF’s Acoustical Poles

A pole/zero RTF as in Eqn. (3.2) or in the ARMA numerical version (3.3), represent a physical room model: poles represent resonances, and zeros represent time delays and anti resonances. Nonetheless, the overall transfer function depends on the positions of the sources and receivers here indicated as $\mathbf{p}_j = [\mathbf{p}_s \ \mathbf{p}_r]^T$, $j$=1, ..., $M$. Therefore, considering a multiple RTFs, the ARMA room model be rewritten as [6]

$$H(\mathbf{p}_j, z) = \frac{\sum_{k=0}^{Q} b_k(\mathbf{p}_j) z^{-k}}{1+\sum_{k=1}^{P} a_k(\mathbf{p}_j) z^{-k}} = \frac{Cz^{-Q_1}\prod_{k=1}^{Q_2}(1-q_k(\mathbf{p}_j)z^{-1})}{\prod_{k=1}^{P}(1-p_k(\mathbf{p}_j)z^{-1})}, \quad j=1,...,M \tag{3.5}$$

where $Q_1$ is zeros at the origin and $Q=Q_1+Q_2$ the overall order of zeros. While the zeros may depend on the acoustical source location, the poles correspond to the natural modes of a room, and they do not change even if the source and receiver positions change or people move. So in the (3.5) room modes correspond to the so called *common acoustical poles* (CAPs) of the all $M$ RTFs.

However, the usual method of modeling an RTF is usually determined by the measurement of the impulse response $h[n]$. The CAPs estimation based on impulse response is not a simple problem and different methods have been proposed to estimate the latter, but all need an order to be priorly established.

In addition, note that the mode density (i.e. the RTF eigenfrequencies density) grows with the second power of frequency, which means that the corresponding poles become closer and closer. Moreover, conventional AR-based spectral estimation methods exhibit very poor capabilities in resolving close resonances so, unless more sophisticated methods are used, high-frequency CAP are very hard to be estimated accurately. Thus, only a reduced number of low frequency CAP is targetable for consistent estimation.

For more detalis [6]-[8].

### *3.2.2 Room Modes*

At low frequencies, a listening room can be considered as a three-dimensional resonator. A sound source inside the confined environment is, in fact, subject to the reflection of the walls: it can thus generate constructive and destructive interferences giving rise to modes (stationary waves) that resonate at precise frequencies.

The listening rooms generally have rather complex geometries with walls that reflect more or less intensely the incident sound energy and the analytical study of natural ways can be done only for simple geometries. However, although a perfectly rectangular room does not exist, results obtained in this simple geometry can be applied at least qualitatively to many rooms encountered in practice.

For a rectangular and undamped room with edge lengths $L_x$, $L_y$ and $L_z$ respectively as shown in Fig. 3.1, the Helmholtz equation is

$$\nabla^2 p + k^2 p = 0, \qquad p = p(x,y,z,t) \tag{3.6}$$

where $p$ is the eigenfunction and $k$ is eigenvalue. In the case of wave $k = 2\pi f/c$ is the wavenumber and $p$ an acoustic quantity as the pressure amplitude. For perfectly parallel and rigid walls, the solutions along the three axes are independent of each other and the variables can be separated. Thus we can write the solution as

$$p(x,y,z,t) = C\cos(k_x x)\cdot\cos(k_y y)\cdot\cos(k_z z) \tag{3.7}$$

where $C$ is a constant and where the eigenvalues for the wavenumber $k$ is given by

$$k^2 = k_x^2 + k_y^2 + k_z^2 = \left(\frac{n_x\pi}{L_x}\right)^2 + \left(\frac{n_y\pi}{L_y}\right)^2 + \left(\frac{n_z\pi}{L_z}\right)^2. \tag{3.8}$$

The corresponding eigenfrequencies can be written as

$$f_{n_x,n_y,n_z} = \frac{c}{2}\sqrt{\left(\frac{n_x}{L_x}\right)^2 + \left(\frac{n_y}{L_y}\right)^2 + \left(\frac{n_z}{L_z}\right)^2} \tag{3.9}$$

where $f_{n_x,n_y,n_z}$ are the natural frequency (or normal mode), $c$ the sound speed, and $n_x$, $n_y$ and $n_z$ integers index that which can vary independently of one another. Thus, the Eqn. (3.7) represents the 3D standing wave along the three propagation axes. An example of modes (2,3,0) and (5,4,0) for a $5 \times 4 \times 3$ m room is reported in Fig. 3.2.

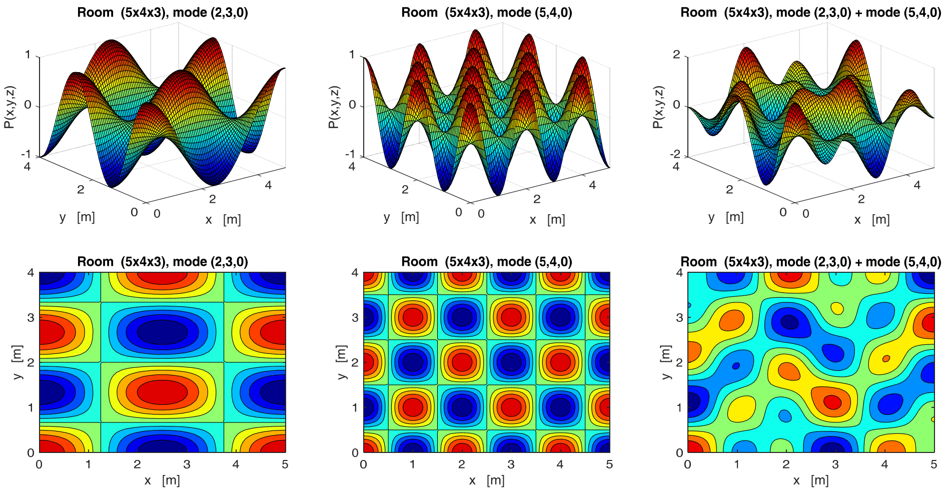


**Fig. 3.2** Example of room modes (5,4,0) and (2,3,0) in a $5 \times 4 \times 3$ m room.

For high indices values, a well know approximated formula exists, that counts the room modes up to a given frequency bound $f > f_B$, and tends to be independent of the room shape for high values of $f_B$

$$n_M(f)|_{f>f_B} \approx \frac{4\pi}{3}\left(\frac{f}{c}\right)^3 \cdot V + \frac{\pi}{4}\left(\frac{f}{c}\right)^2 \cdot S + \frac{f}{8c} \cdot L \tag{3.10}$$

where $V$ is the room volume, $S$ is the total surface area of the walls, and $L$ the sum of all edge lengths occurring in the rectangular room.

#### 3.2.2.1 Statistical Description of Room Modes

The number of natural modes grows very rapidly with the frequency, and after a certain frequency, due to the numerous constructive and destructive interferences, the field becomes diffuse: around a certain frequency, called Schroeder frequency, so a number of modes are excited and the field can be considered a diffuse field. Therefore, at sufficiently high frequencies, we may express the above mentioned variables by statistical means. The Schroeder frequency depends only on the volume and the reverberation time $T_{60}$ and can be calculated empirically with the formula

$$f_{Sh} \approx 2000\sqrt{T_{60}/V} \tag{3.11}$$

where $V$ is in $\mathrm{m}^3$. The Schroeder formula (3.11) can be understood by the fact that the resonance bandwidth is inversely proportional to the reverberation time, and the separation between the natural frequencies are inversely proportional to the volume of the room.

For example, in a room of size $V = 4 \times 6 \times 3$ m$^3$ and a reverberation time of $T_{60} = 0.5$ s the sound field can be considered diffuse for frequencies above 170 Hz.

### 3.2.3 The Reverberation

The reverberation is the phenomenon that most characterizes the acoustic behavior of confined spaces. Considering an acoustic source placed in a room, the reverberation phenomenon is due to the presence of multiple walls reflections.

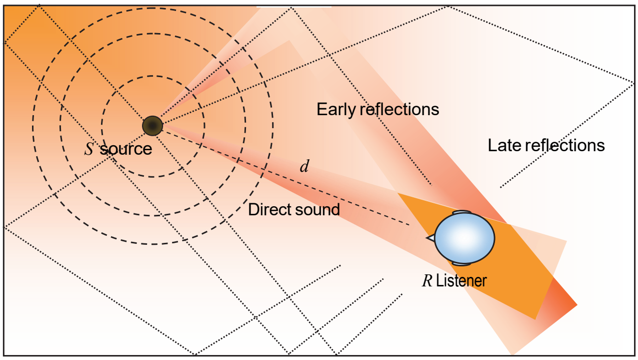


**Fig. 3.3** The sounds we perceive do not reach us as they were emitted by the source, but are always, to a greater or lesser extent, modified by the surrounding environment.

Considering the scenario described in Fig. 3.3, if the walls have a coefficient of absorption $\alpha$ (see Fig. 3.4), the resulting reflected sound is given by the reflected sound from a perfectly specular surface, multiplied by the gain coefficient $(1-\alpha)$. For example, if the ceiling is made of a very sound-absorbing material with a absorption coefficient $\alpha = 0.8$ then $(1-\alpha) = 0.2$, we have that only 20% of the sound energy incident on the ceiling is reflected and it is also for this reason that the reflected rays arrive at the receiver with a strong attenuation.

After a certain time from the emission of an acoustic signal, the repeated reflections on the walls gives rise to multiple-order rays reflections, creating a complex geometric representation of the paths of the various acoustic rays. In this case the arrival times of all reflected rays tend to become characterized by a uniform sound pressure distribution (i.e. spatial diffusion) and by a uniform incident distribution (i.e. isotropy condition).

This phenomenon gives rise to very complex vibration modes and for the study of all these effects it is preferred to use the term *reverberation*. Furthermore, the sound field consisting of the set of reflected waves, each characterized by a different time delay and attenuation level, is referred to as *reverberation field*.

When the source $S$ irradiates a message (speech, music, etc.) consisting of a succession of different acoustic sounds the receiver $R$, also, receives the residual reverberation

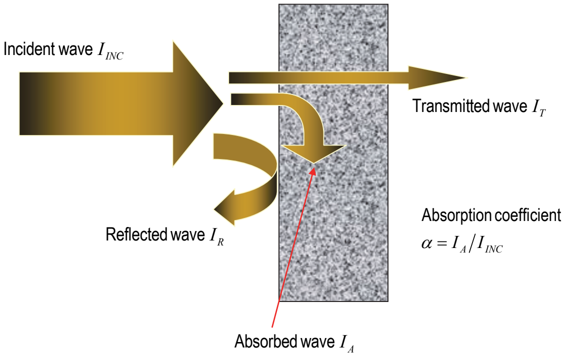


**Fig. 3.4** Schematization of a wall hit by an incident acoustic wave and representation of the acoustic quantities of interest (absorbed, reflected, transmitted wave and absorption coefficient).

of the sounds previously emitted. In a room, as illustrated in Fig. 3.5, the received sound signal can be divided into three parts: the direct sound, the first reflections, and the late reflections. Furthermore, are the first reflections that most define the perceived “timbre” of the hall.

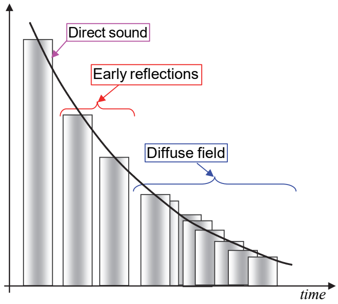


**Fig. 3.5** In the reverberation, the late reflections are no longer perceived individually, but fused together to form a “sound tail”.

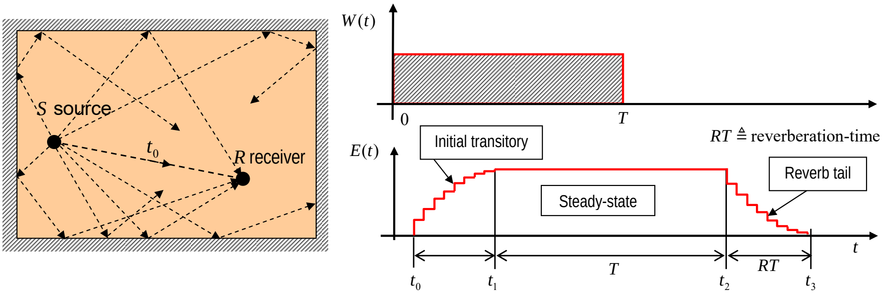


**Fig. 3.6** Power delivered by the acoustic source $W(t)$ at the point $S$, and the acoustic intensity observed $E(t)$ at the point $R$ of the reverberating room.

***Remark 3.4.*** Note that the room can be considered as a linear dynamic system. In fact, with reference to Fig. 3.6 the energy curve on the receiver $E(t)$ can be divided into the following parts:

1. $[0, t_0]$ - It is the time of the direct path (the shortest path) necessary to the acoustic wave to reach the receiver;
2. $[t_0, t_1[$ - It is the *initial transient* in which the first reflections are formed;
3. $[t_1, t_2]$ - It represents the *steady-state* in which all the modes of the room are excited and the source continues to persist. Its duration is equal to $T$. The acoustic energy in the room stabilizes at a constant level that corresponds to the balance between the power emitted by the source, and that due to wall's reflections.
4. $[t_2, t_3]$ - It represents the *final transient*. Once the acoustic source is turned off, the reflections are still present until the total energy has been exhausted. The time $[t_2, t_3]$ is referred to as reverberation time.

The perceived listening quality is greatly influenced by the reverberation. In music, for example, it is perceived as a contribution that, depending on the genre and the level, enriches the sound content. On the contrary, in theatrical representations, and for the vocal message in general, it is considered a disturbance as it can diminish the intelligibility of the spoken words. In fact, the reverberation is related to the subjective behavior of hearing and, in particular, to the fact that the ear integrates with the direct sound that part of the reverberation that reaches the listener with a delay of no more than a few tens of milliseconds (typically around at 30 ms). For greater delays, depending on the residual sound level, the indirect sound is perceived distinctly and is called echo.

## *3.2.4 Convolutional Reverberation Model*

In order to better model the reverberation, we consider a room as a dynamical system in which the acoustic source is the input or the cause, while the signal perceived by the receiver is the output, i.e. the effect. For example, in standard acoustics measurements, introduced in §1.9.2, the source is broadband loudspeaker array arranged in a dodecahedron configuration, so as to obtain an omnidirectional radiation pattern while for the receiver can be used an omnidirectional microphone [23], [24].

The acoustic source emits spherical wave fronts with equal intensity in each direction (isotropic source) and the microphone is characterized by a spherical radiation pattern. As shows in Fig. 3.3, direct and reflected acoustic ray will have different paths and times. So, from the acoustic theory it is known that the sound level is inversely proportional to the distance achieved, it is therefore evident that the sound levels are gradually lower with the arrival time increasing.

### 3.2.4.1 Room Impulse Response

Let $s(t)$ be the emitted source signal, in the hypothesis that the wall reflection is not dependent on the frequency, i.e. the reflection represents a simple attenuation, the signal on the receiver indicated as $r(t)$ can be determined by the superimposition of infinite replicas of the delayed and attenuated original signal

$$r(t) = g_1 s(t - \tau_1) + g_2 s(t - \tau_2) + \cdots$$

where $g_k$ represents the particular strength (or gain) of each reflection. So, let $\delta(t)$ be the Dirac delta function, the continuous-domain *Room Impulse Response* (RIR) can be define as

$$h(t) = \sum_k g_k \delta(t - \tau_k) \tag{3.12}$$

such that the signal at the receiver can be written as a simple convolution

$$r(t) = \int_0^\infty h(\tau)s(t - \tau)d\tau, \quad \text{for } t \geq 0.$$

In other words, considering for simplicity a discrete-time approximate model in which the time $t$, after a sampling process, is substitute with an integer $n$, the preceding expression can be expressed as a convolution sum $r[n] = \sum_{k=0}^{M-1} h[n-k]s[n]$, $n = 0, 1, \ldots,$ where as we will see below, the length $M$ is chosen according to the reverberation time; whereby the reverberation, of the acoustic path between the source and the receiver, becomes identified with the impulse response $h[k]$, $k$=0,1, ...,$M-1$; also denoted as discrete-domain RIR.

Thus, the acoustic path between $S$ and $R$ can be modeled as a discrete-domain *Finite Impulse Response* (FIR) filter applied to the input signal with transfer function $H(e^{j\omega}) = \sum_n h[n]e^{-j\omega n}$ is the point-to-point room transfer function (RTF).

It is obvious that the impulse response is defined between two specific points. So if we want to completely characterize an acoustic environment we should consider the presence of more sources and more receivers, maybe placed on a spatial grid of appropriate dimensions.

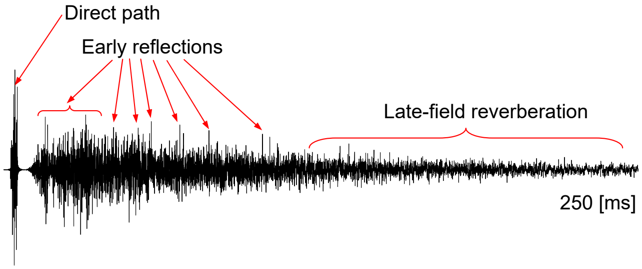


**Fig. 3.7** Room impulse response of *Teatro Alla Scala* of Milan (first 250 ms of a longer RIR), where the first reflections are quite evident.

***Remark 3.5.*** It should be noted that in real cases the propagation medium (i.e. the air) is characterized by a certain impedance, just as the reflections of the walls are also dependent on the frequencies. However, being such linear phenomena, they can be incorporated into the impulse response that contains all the information of the propagation channel. Therefore the convolution model of the reverberation is consistent even in the non-ideal but linear case. Obviously, to take into account also the complex nonlinear propagation phenomena, it is necessary to consider more sophisticated models.

For example, Fig. 3.7 shows the first 250 ms of a longer RIR of a room where the first reflections (i.e. $h(t-\tau_1)$, $h(t-\tau_2)$, ... ) are quite evident. In the first part, in fact, the impulse response has a so-called "sparse" trend, that is, the energy is concentrated only in a few peaks of the response while elsewhere it is almost null. On the contrary, in the reverberation tail, due to the presence of multiple reflections, the response is dense. In this case the acoustic field is of diffuse type (see §1.8.1).

***Remark 3.6.*** Observe that while for the first part of the reverberation it is possible to use a deterministic approach, i.e. for a simple geometry the reflections are predictable; in the reverberation tail, the presence of multiple and dense reflections tend to interfere with each other in a completely unpredictable way. In this case the characterization can be done with a statistical acoustics approach.

***Remark 3.7.*** Note that the previous development is consistent with the model in Eqn.s (3.2) and (3.3). Since the discrete-time impulse response $h[n]$ is the inverse $z$-transform of the $H_{RTF}(z)$, its measurement generally represents the starting point for the estimation of the RTF parameters in terms of poles and zeros.

### 3.2.4.2 Reflexion Diagram

The *reflextion diagram* also noted as *echogram*, that contains all significant information about the temporal structure of the sound field at a given room point, can be defined as

$$h_e(t) = 10\log_{10}|h(t)|^2$$

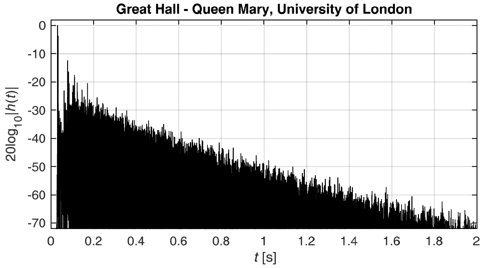


**Fig. 3.8** Example of reflexion diagram.

Note that, the echogram is sometime presented in specific frequency range, e.g. octave, or third-octave band implying that the corresponding RIR has been filtered be a specific band-pass filter.

In addition, note that from the analysis of the echogram it is possible to estimate also the possible background noise present in the room during the measurement. For example, in Fig.3.9-a) is reported the echogram in absence of background noise; while in Fig.3.9-b) is reported the echogram in the presence of a background noise at -45dB for which the decay curves tend asymptotically at this value.

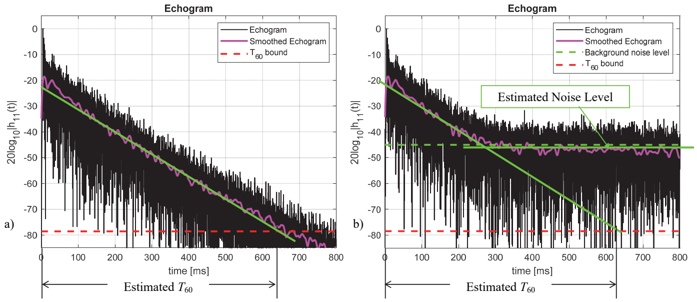


**Fig. 3.9** Example of echogram and smoothed echogram. a) Echogram measured without background noise. b) Echogram measured in presence of a noise floor of −45 dB.

## 3.3 Measurement of the Room Impulse Response

The *Room Impulse Response* (RIR) of a confined environment plays a fundamental role as it is decisive in the qualitative assessment of the acoustic environment.

The direct estimation of a RIR can be made by an impulsive excitation (generated by a loudspeaker, by an firecracker burst, a gun, etc.). However, this approach is inadvisable for several contraindications, including: 1) the difficulty of generating an infinite amplitude signal that is also exactly repeatable, for example, with a loudspeaker; 2) a high amplitude would bring the electroacoustic measurement chain (e.g. the loudspeaker) into a non-linear operating zone.

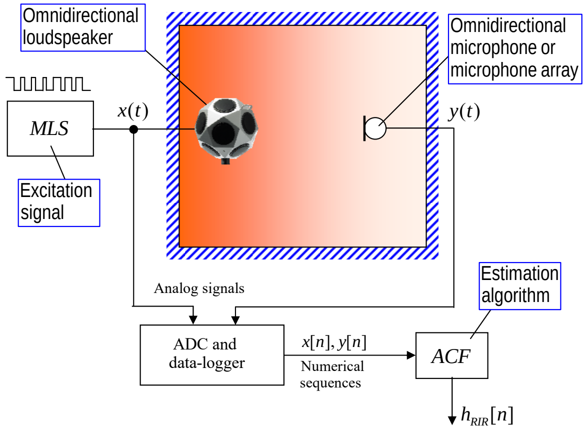


**Fig. 3.10** Typical scheme for acoustic measurement with excitation signal (e.g. a MLS sequence) irradiated by a dodecahedron loudspeaker which approximates an isotropic source, and an omnidirectional microphones.

In practice, as already seen previously for the estimate of HRTF in §2.5.3, better RIR estimates are possible by means of *correlated measures* that can derived from the Wiener-Hopf theory (see Appendix A §A.2.1). Therefore, as shown in Fig. 3.10, instead of using an impulsive signal, it is possible to estimate the impulsive response by exciting the system with an input signal $x[n]$: i) of limited-amplitude such that the system responds in a linear manner; ii) of continuative type so as to be able to obtain an adequate level above the background noise; iii) provided that it is "sufficiently informative", that is, has a sufficiently high spectrum, its autocorrelation function is very approximable with an ideal impulse (delta Dirac).

In particular, the choice of the input signal $x[n]$ has a substantial influence on the quality of the observed data. In fact, the input excitation determines the operation point and which modes and parts of the system are excited during the experiment. Two different aspects must be considered in the choice of the input signal: the first concerns its second-order statistical properties such its correlation $R_{xx}$ and its spectrum $\Phi_x^p(e^{j\omega})$; the second question concerns its waveform. Thus, you can work with different types of signals such as: sum of sinusoids, frequency modulated sinusoid with deterministic law, filtered noise, pseudo random sequences, binary signals, etc.

### 3.3.1 Excitation Signals for RIR Estimation

The most used signals that are used for RIR measurements are briefly shown below.

- The random periodic signal denoted *Pseudo Random Binary Sequence* (PRBS) also called *Maximum Length Sequence*(MLS) [9],[10], as used for the HRTF measurements, already described in §2.5.4, whose characteristics are reported in Appendix A §A.3. When choosing the MLS, bear in mind that the length of the sequence (i.e. a period) must be greater than its reverberation time. Otherwise the estimation of the autocorrelation function (ACF) would be affected by aliasing.
- The *swept sine* wave also called an FM *chirp* or *sweep* signal consists of a sinusoidal signal with a frequency that varies linearly or with exponential law between two extremes that cover the acoustic range of interest. The main advantage in the use of these signals is the possibility to characterize both the linear part and, even if partially, the nonlinear part of the [11], [12] system. In fact, since 2000, the method has been used in many applications, namely in fields of audio and acoustics, for RIR and HRTF and in many other applications in acoustics. Compared with MLS and linear sweep, the usage of exponential sine sweep, provided several advantages in term of signal-to-noise ratio and management of any nonlinear phenomena present in the measurement chain [13].
- The *multisine signal*, consists of a signal formed by a large number of sinusoids and such as to cover the spectrum of interest in the most uniform way possible. To have a minimal crest factor, the phases are determined according to an optimization criterion. Sometime, the frequencies of the sinusoids are chosen so as to avoid overlapping with the harmonics due to the non-linearity distortions that otherwise could not be detected.

The *swept sine* techniques according to many authors are the ones that best lend themselves to obtaining a high SNR. However, it should be noted that in the case of an exponential sweep, in the low frequency the signal varies more slowly than in the high frequencies. Thus the spectrum decays 10 dB per dacade. The measurement of the impulsive response must therefore be equalized [11]-[13].

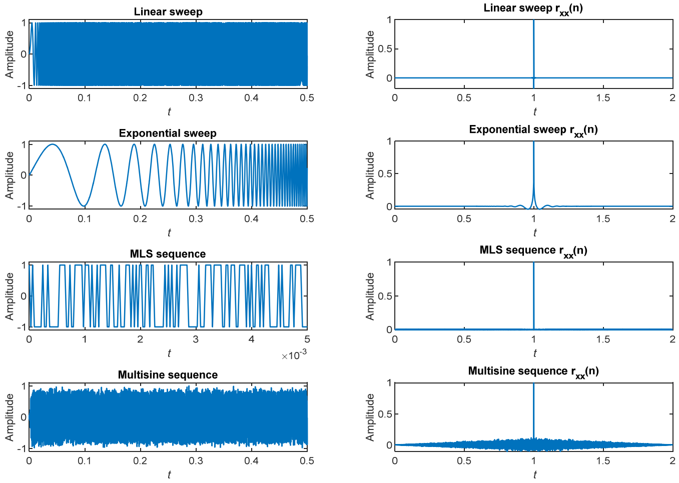


**Fig. 3.11** Typical test signals in the time domain and their autocorrelation function.

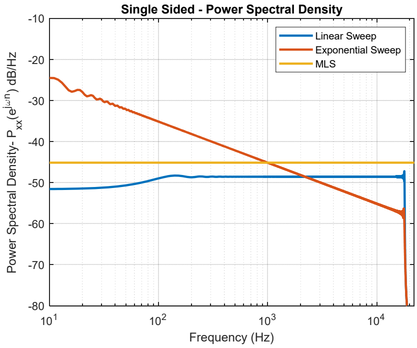


**Fig. 3.12** Power spectrum density of typical test signals.

In Fig. 3.11 some test signals and their correlation functions are reported, while in Fig. 3.12 the power spectrum density of typical test signals.

### *3.3.2 Evaluation of Spatial Information*

The external hearing organ consists of two ears (i.e. two spatially non-coincident receiver) so that the subjective quality of the perceived sound also depends on the small differences between the signals arriving at the right and left ear. Furthermore, it is known that the sensation *sound spatiality* in a listening environment can be modeled through early reflections. The criteria for subjective assessment of spatial quality are defined on the basis of correlation functions between the left and right ear signals: if the left and right signals are uncorrelated, there is a high degree of spaciousness.

For example, Ando in [14], proposes a relationship between the desired spatial response and the *inter-aural cross-correlation function* (IACF) defined as

$$\mathrm{IACF}_{t_1,t_2}(\tau) = \frac{\int_{t_1}^{t_2} h_r(\tau)\cdot h_l(t+\tau)\cdot dt}{\sqrt{\int_{t_1}^{t_2} h_r^2(t)\cdot dt \cdot \int_{t_1}^{t_2} h_l^2(t)\cdot dt}} \tag{3.13}$$

where $h_r(t)$ and $h_l(t)$ are, respectively, the impulse responses measured at the ear canal of the right and left ear. The IACF is measured with either a dummy head, or a real head with average dimensions as exemplified by dummy heads, and with small microphones placed inside the ear canals [23].

The *inter-aural cross correlation coefficients*, (IACC), is defined as

$$\mathrm{IACC}_{t_1,t_2}(\tau) = \underset{-1<\tau<1\mathrm{ms}}{\arg\max}\ |\mathrm{IACF}_{t_1,t_2}(\tau)|. \tag{3.14}$$

Therefore, for the measurement of spatial information and to calculate (3.13) and (3.14), it is necessary to develop binaural measurement techniques as, for the HRTFs estimation (see §2.5.3).

The IACC measurement requires the use of the dummy-head. As alternative can be considered the *lateral fraction* $L_f$ parameter (which will be formally defined in §3.7.2.3, see Fig. 3.35), that represents the percentage of lateral energy compared to the total. Its evaluation is given by the ratio between the energy acquired by a microphone with a figure-of-eight spatial response, at 90° with respect to the source, and the energy acquired by a microphone, places in the same point (coincident microphones), with an omnidirectional spatial response.

## 3.3.3 Measurement of 3D RIR

As an alternative to classical binaural measures, methods for measuring 3D spatial impulsive response using coincident microphone arrays. The first to propose the 3D RIR registration was Gerzon [17]. The method is based on the use of a particular array referred to as Ambisonic B-format, shown in Fig. 3.13, capable of acquiring the four RIR one omnidirectional plus three directional RIRs, using figure-of-eight microphones, along the three axes ($x$, $y$, $z$) [20]. This technique, together with others, has been used by Farina and Tronchin [16], for the acoustic characterization of the most famous Theaters in the world.

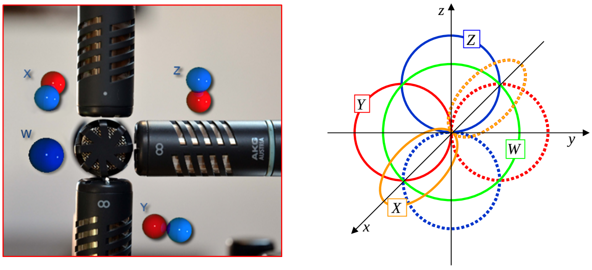


**Fig. 3.13** Ambisonic coincident microphone array consisting of four microphones a) one omnidirectional and three figure-of-eight with diagrams showing spatial response arranged along the $x$, $y$, and $z$ axes. b) Microphone polar patterns.

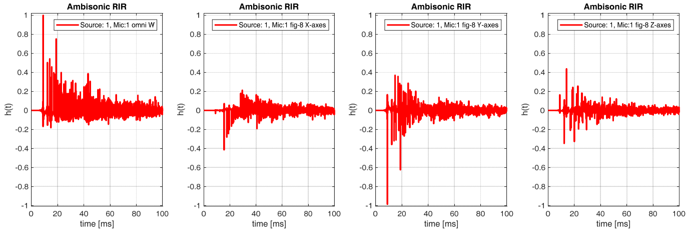


**Fig. 3.14** Example of RIRs acquired with Ambisonic microphone, first 100 ms of a longer RIR.

## 3.4 Reverb Macroscopic Indices

Some musical genres are related to the characteristics of the room acoustic as well as the string of a musical instrument to its sound box. Consider for example a chorus of voices heard in a church characterized by a very long reverberation time. The effect would certainly not be the same if the same choir were heard in a different and not reverberant environment. On the contrary, in theatrical representations, the overlapping of reverberating tails can compromise the word intelligibility.

We can say, then, that the ambient effect consists in the more or less long and gradually attenuated persistence of the sound after the sound source has ceased to act. This effect is due to the multiple reflections of the sound waves on the walls, ceiling and floor of the listening room. The duration of the reverberation time, which is one of the factors that most influence the quality of listening, depends on the shape, the volume, the various structures of the environment and the relative distance of the reflecting surfaces. Furniture, and the presence of people, affect the reverberation time, reducing its duration and intensity.

### *3.4.1 Reverberation Time* $T_{60}$

The Sabine's *Reverberation Time* (RT) definition dates back to around 1900, when there were no microphones or other electronic devices that could be used for measurements. For the RT estimate empirical methods were used, such as, the burst of a balloon, pistol shot, clapping of hands, and son on; for example, Sabine used a method with a stopwatch of four identical sets or organ pipes.

So, according to the first RT definition of Sabine, more formally the reverberation time can be defined as follows [3]-[5], [38].

**Definition 3.1.** *Reverberation time (RT)* - Let's indicate with $t = 0$ the instant in which an acoustic source, usually a stationary white Gaussian noise (WGN), is switched off, the RT is the necessary time, so that the average density of sound energy decreases by 60 dB with respect to the value reached by the acoustic source at $t = 0$ (e.g. as the $[t_2, t_3]$ interval in Fig. 3.6)). This time, indicated as $T_{60}$ parameter, allows you to evaluate how long a sound is extinguished in a closed environment. This measurement method in denoted as *interrupted noise method* [23].

Note that, the definition given by Sabine does not indicate in which exact moment the $T_{60}$ measurement should be started. In fact, the sound energy density is measured at the listening point $R$ which may be distant from the sound source point $S$. Then, considering the observation point $R$, the energy begins to decay only after a certain time necessary to the wave to cover the path that can be eavluated as $t_0 = \|S-R\|/c$.

***Remark 3.8.*** Observe that, in order to avoid too strong influence from the direct sound, no microphone position shall be too close to any source position. The minimum distance $d_{\min}$ in meters, can be calculated as

$$d_{\min} = 2\sqrt{V/cT}$$

where $V$ is the volume in m$^3$, $c$ the speed of sound in m/s and $T$ is an estimate of the expected reverberation time, in s [23].

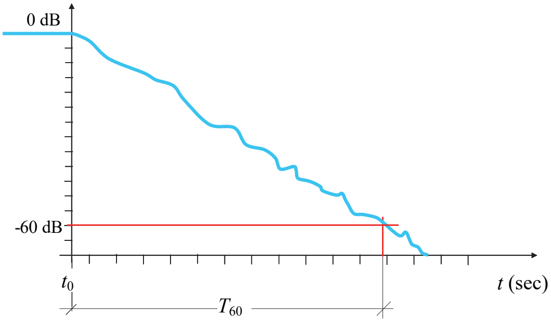


**Fig. 3.15** The energy density decay curve or *energy decay curve* of a confined listening environments. By switching off a steady-state full-bandwidth sound source, the $T_{60}$ is defined as the time required for the energy density to be attenuated by 60 dB.

In the presence of a high background noise, for which it is not possible to make a correct measurement, can be used an extrapolation of the first part of the decay curve.

In theory, if the energy density decay curve were exactly exponential, i.e. it was identical for all the natural modes, as predicted by the statistical acoustics, the curve of the level would be a straight line with a constant slope. As a result, extrapolation would not lead to significant errors. However, in many cases we have that, depending on the variation of the absorption characteristics of the walls with the frequency, to different modes correspond to different decay constants. So, sometimes, the decay curves are not straight, but presents a double slopes or not inconspicuous curves. Therefore the extrapolation is strongly influenced by the initial curve section. Thus, to overcome this problem, the decay measurement is performed between $-5$ and $-35$ dB below the steady-state noise level, and the extrapolation to $-60$ dB is performed considering these values.

The RIR contains only the information between two points and therefore, even considering the methods of integration and interpolation described in the next paragraph, does not contain all the room information. For example, if the microphone position is placed in a node of some natural mode, the computed reverberation time is approximated by default. Putting the measuring microphone close to edges would result in an exaltation of the low frequency modes. Thus, for a more consistent estimate of the decay curve, measurements can be made at several points in the room and then averaged.

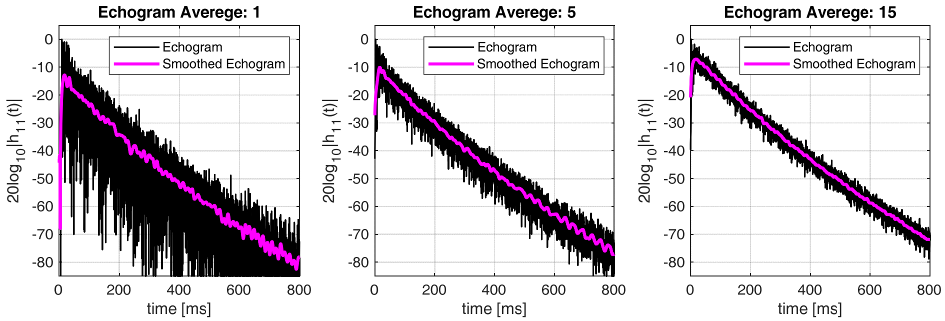


**Fig. 3.16** Averaging of energy density decay curve or *energy decay curve* acquired in, respectively 1, 5, 15, different random positions in confined listening environments.

### 3.4.2 Room Impulse Response and Reverberation Time

One of the problems in the RIR measurements is due to the possible presence of a high level of background noise. To limit its effect, supposing zero-mean noise, it is possible to repeat and average several synchronized measures. However, the drawback of this procedure is the remarkable elongation of the measurement time.

The link between the single impulse response and the reverberation time was studied by Schroeder [10] who showed how the reverberation decay law can be rebuilt through an integral of the impulse response square.

So, to minimize measurement time and compute the decay curve with on a single RIR, with the simple hypothesis that the excitation signal is a stationary WGN, Schroeder in [10] developed an alternate method, denoted *integrated impulse response* or *Schroeder "backward integration"*, that can be defined as a method of obtaining decay curves by reverse-time integration of the squared impulse responses [23].

Suppose the room is excited by zero-mean WGN with unitary variance $n(t) \in \mathcal{N}(0,1)$, let $h(t)$ be the RIR, we can define the noise curve decay as

$$\mathrm{EDC}(t) = \int_t^{\infty} h(\tau)n(t-\tau)d\tau, \qquad t \geq 0.$$

The Schroeder's method is based on the relationship between the ensemble average $\left\langle \mathrm{EDC}^2(t) \right\rangle$ of all possible decay curves. Squaring the latter expression yields a double integral, which after averaging (indicated by the operator $\langle \cdot \rangle$), can be written as

$$\left\langle \mathrm{EDC}^2(t) \right\rangle = \int_t^{\infty} h(\tau)d\tau \int_t^{\infty} h(\xi) \left\langle n(t-\tau)n(t-\xi) \right\rangle d\xi.$$

Now, as the right-hand side is the autocorrelation function of $n(t)$ that is a $\delta$ function, for the sampling property, the above double integral can be reduced to the single integral

$$\left\langle \mathrm{EDC}^2(t) \right\rangle = \int_t^{\infty} |h(\tau)|^2 d\tau.$$

In an ideal situation with no background noise the integration should start at the end of the impulse response ($t \to \infty$) and proceed in reverse mode to the beginning of the squared impulse response.

$$\left\langle \mathrm{EDC}^2(t) \right\rangle = \int_t^{\infty} |h(\tau)|^2 d\tau = \int_{\infty}^{t} |h(\tau)|^2 d(-\tau).$$

In order to minimize the influence of background noise present at the end of the RIR we can perform the backward integration from a time $t_1 > t$

$$\left\langle \mathrm{EDC}^2(t) \right\rangle = \int_{t_1}^{t} |h(\tau)|^2 d(-\tau) + C \tag{3.15}$$

where, in the case of no background noise the integration should start at the end of the impulse response: i.e. $t_1 = \infty$ and $C = 0$. On the contrary $t_1$ and $C$ are determined as a function of the background noise, depending on whether the noise level is known or not (for details see [23]).

However, if the level of the background noise is unknown, the backward integration of the squared impulse is performed as

$$\mathrm{EDC}(t) = \int_{t+T_0}^{t} h^2(\tau) d(-\tau) \tag{3.16}$$

where, as indicated in the ISO guidelines [23], the optimum value of $T_0$ is about 1/5 of the estimated reverberation time. In other words, the EDC curve is calculated as a moving average FIR filter of length $T_0$. As shown in Fig. 3.17, the EDC decays more smoothly than the RIR itself, and so it is more useful than ordinary amplitude

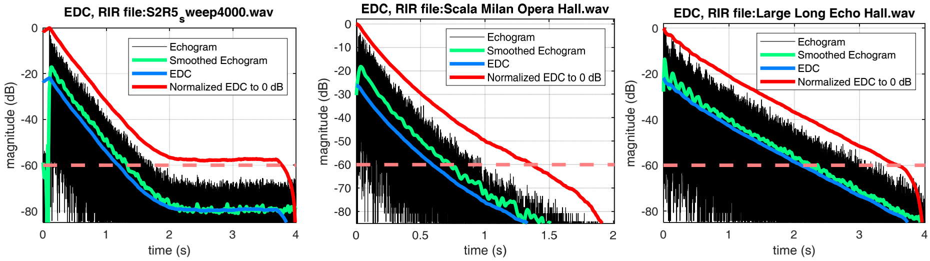


**Fig. 3.17** Example *energy decay curves* (EDC) and echograms.

envelopes for estimating $T_{60}$, that is evaluated considering the normalized EDC to 0 dB.

### 3.4.3 Energetic Time-Frequency RIR Representation

For a more accurate RIR analysis, you should also take into account the frequency. Generally, low frequency natural modes have a higher decay time than higher frequency modes. In this regard, it is possible to define an extension of Schroeder's energy decay curve based on a time-frequency representation obtained from $h[n]$ using short-time Fourier transform (STFT) $H(n, e^{jw})$ defined as

$$H(n,k) = \sum_{k=0}^{K-1} w[n-k]h[k]e^{-jwk}, \qquad n = 0, 1, \ ..., \ N-1 \tag{3.17}$$

where $n$ and $k$ are the time and frequency index, respectively, and where $w[\cdot]$ is an analysis window of appropriate shape that smoothes the signal at its extremes to attenuate Gibbs' phenomenon, such as a triangular (or Bartlett) window of length 30 or 40 ms. Thus, as shown in Fig. 3.18, the STFT is equivalent to a DTFT of a portion of signal determined by a $w[n-k]$ analysis window of appropriate shape and that advances along the signal with a certain overlap (e.g. 50%) with the previous one. Then for RIR frequency analysis, is used the Energy Decay Relief (EDR) [25], defined as

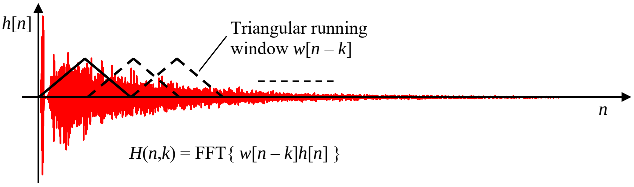

h[n]
Triangular running window w[n – k]
n
H(n,k) = FFT{ w[n – k]h[n] }

**Fig. 3.18** Representation of the *Short-Time Fourier Transform* (STFT) with a triangular “running window.

$$\text{EDR}(n,k) \triangleq \sum_{n=m}^{M} |H(n,k)|^2 \tag{3.18}$$

where $H(n,k)$ is the STFT, and $M$ denotes the total number of time frames.

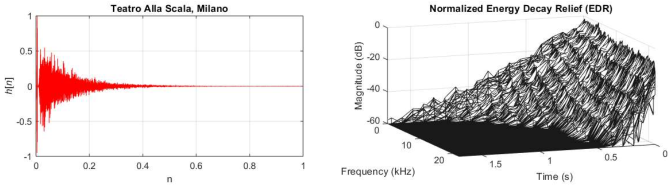


**Fig. 3.19** Energy decay relief (EDR) of Teatro alla Scala of Milan.

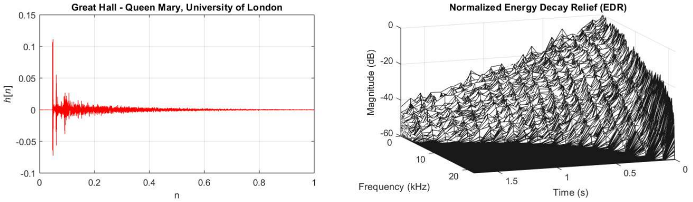


**Fig. 3.20** Energy decay relief (EDR) Great Hall - Queen Mary, University of London. Data from [26].

In Fig.s 3.19 and 3.20 are reported the EDR of the Teatro alla Scala of Milan and the Great Hall - Queen Mary, University of London, where it is possible to see how at low frequency the EDR surface decays more slowly.

### 3.4.4 Optimal Reverberation Time

In the acoustic design of a listening environment as a concert hall, a TV studio, ...; the considerations previously made must be taken into account. It is necessary, in fact, to take into account not only the architecture of the room, but its volume, the treatment of masonry structures, stable furniture, with particular regard to the choice of seats, so as to equalize the absorption, regardless of whether there is or there is no person seated.

For example, from Fig. 3.21 we can see how some environments, even with very high volumes, such as television studios and cinemas, must have a low reverberation time..

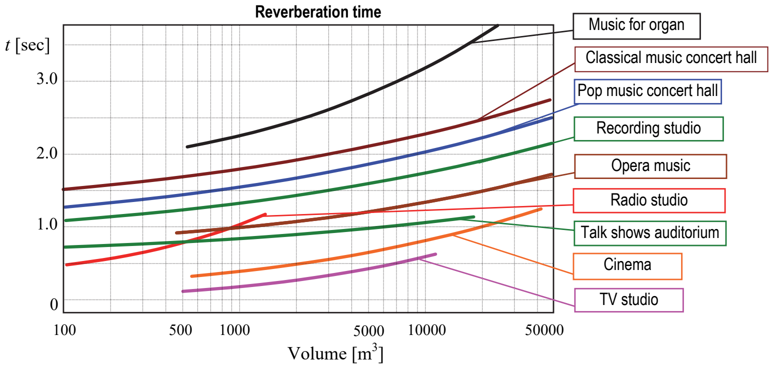


**Fig. 3.21** Optimal reverberation times, specific to some musical genres, and of characteristic acoustic environments of particular interest.

For a “good acoustic response”, the quality of reverberation is very important, that is the aptitude to act in a suitably dosed way on all the frequencies of the audible field. The “sound color” tend towards the bright if the reverberation will act mainly on the high frequencies; instead it will tend to gloomy if are exalted the low frequencies. The presence of pronounced peaks on the frequency response, especially at medium-low frequencies, will tend to perceived as an annoying coloring and "rumbling". In fact, the reverberation at low frequencies is tendentially higher, $T_{60}(bf) > T_{60}(hf)$, and for optimal listening the frequency response spectrum must be regular and without peaks.

The optimal reverberation dosage at various frequencies must be adjusted according to certain quality masks such as the one shown, for example, in Fig. 3.22.

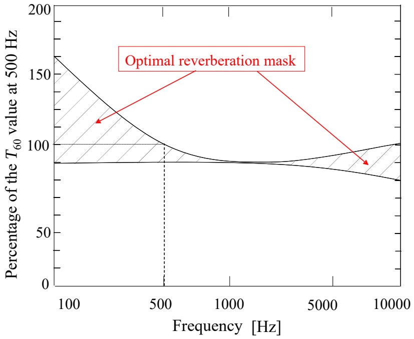


**Fig. 3.22** Reverberation time as a function of frequency. At high frequencies, walls generally have a higher absorption coefficient.

## *3.4.5 Reverberation Distance*

The reverberation distance is defined as the distance from the source for which the field can be considered diffuse.

Let's consider an reverberant ambient, for example, like the one described in Fig. 3.6 with an $S$ source and a receiver $R$. The SPL $L_t$ perceived by a listener will be given by the sum of two pressures: the direct radiation $L_d$ (sound pressure level SPL which reaches a listener directly without being reflected); the reflected radiation $L_r$ (defined as the level of sound pressure due to the only reflections). Therefore it results $L_t = L_d + L_r$.

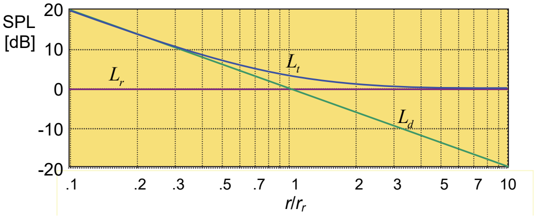


**Fig. 3.23** Reverberation distance. For the inverse-square law (§1.7.2.1), the power of direct irradiation decreases by 6 dB for each doubling of the distance (i.e. 20 dB/Dec), while the diffuse field is constant.

**Definition 3.2.** *Reverberation distance* - We define reverberation distance, the distance $r_r$ from the source so that the level of the direct field is identical to that of the reverberating field, i.e. $L_d = L_r$.

Thus, the reverberation distance can be roughly calculated from the relationship

$$r_r = 0.25\sqrt{\alpha S/\pi} \approx 0.06\sqrt{V/T_{60}}$$

## 3.5 Acoustics Room Correction

In the designing of an listening room, the main problems to be faced can be described as follows:

1. what we want to achieve;
2. what allows performers to play well;
3. what makes the sound more beautiful;
4. how we can achieve the quality desired by the performers and the public.

In principle, the design of an "acoustically interesting" environment is based on the optimal arrangement of acoustic mirrors and sound-absorbing walls.

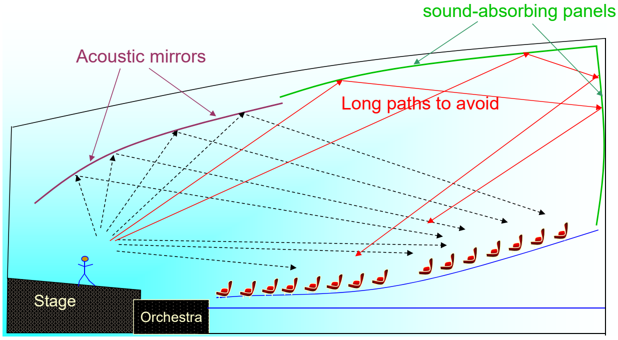


**Fig. 3.24** For a "good acoustics" it is necessary to strategically arrange acoustic mirrors and sound-absorbing panels.

If the same environment is designed to make different representations (for example cinema, music, theater), the room must be equipped with a variable acoustic response: the arrangement of the panels and mirrors is optimized according to the type of representation.

The problem is very complex and generally eliminating reverberation at low frequencies is more difficult.

The Greek theaters are still an example of how we can achieve exceptionally good results by exploiting the possibility of reinforcing the sound through appropriate reflections that are not too late compared to the direct sound.

### 3.5.1 Sound-Absorbing Materials

The possibility of adjusting the reverberation time of an environment, appropriately arranging the sound-absorbing materials, affects both the design (or the acoustic correction of a room) and the methods of noise attenuation inside a building. In general, sound-absorbing materials can be divided into three categories:

- porous materials (absorption by porosity);
- vibrating panels (absorption by membrane resonance);
- perforated-absorbent panels (absorption by cavity resonance).

When an acoustic wave affects a porous surface as a sound-absorbing panel, a good part of it penetrates into the material, the pressure variations (closely related to the variation of the speed of the sound wave), make the air molecules vibrate in the interstices. These vibrations dissipate energy in the form of friction.

As shows in Fig. 3.25-a), the absorption coefficient of sound-absorbing panel is a function of the thickness and frequency. The panel absorbs more, the sounds with wavelength proportional to its thickness. So if we want to decrease the reverberation at low frequencies, we need to use thicker panels. Moreover, as shows in Fig. 3.25-b), to increase this thickness, it is advisable to space the panel from the wall. In this case the equivalent width of the acoustic panel is given by the sum of the air gap and the effective thickness of the panel.

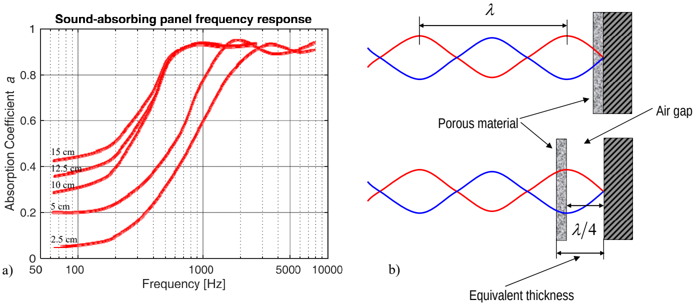


**Fig. 3.25** Sound-absorbing panel. a) Absorption coefficient a typical panel, according to the thickness and frequency. b) Equivalent thickness of a panel with a certain air gap with respect to the wall.

### 3.5.2 Selective Frequency Sound-Absorbing Panels

In addition, it is possible to selectively eliminate some frequencies using the sound-proofing anti-resonant panels that are constructed by placing a perforated panel in front of the sound-absorbing material.

Referring to Fig. 3.26-b) you can determine the the resonant frequency $f_r$ considering the surface of the panel, the holes, and the their number; while for the bandwidth of the frequencies to be deleted results: ($\Delta f \propto$ sound-absorbing material quantity).

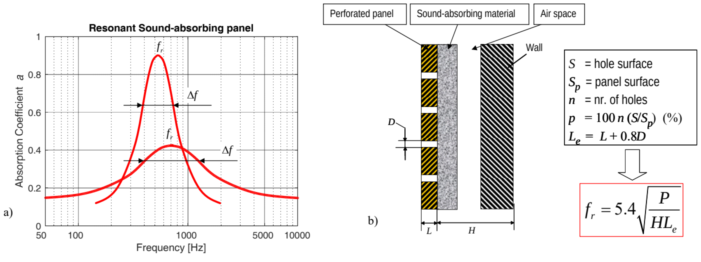


**Fig. 3.26** Resonant sound-absorbing panels. a) Absorption coefficient as a function of frequency. b) By controlling the panel geometry, it is possible to selectively eliminate the unwanted resonances of the room.

## 3.6 Geometrical Acoustic Models

Already in the early 1960s, attempts were made to simulate reverberant acoustic environments using electronic circuits. However, these were not true RIR estimates of a certain environment, but rather of simple reverberation effects [2].

Currently, for the estimation of the RIR and, consequently, the room transfer function (RTF) there are various methodologies and schools of thought [21]-[31].

The modal analysis, starting from the physical model, attempts to solve the wave equation with the boundary conditions given by the walls. However, mathematical and computational difficulties make this method practicable only for simple rooms (parallelepiped) or for limited frequency range.

Nowadays, the most used techniques are based on geometric methods, whose name derives from the fact that the sound waves are treated as rays where geometric Snell laws are valid (see §1.8.4.1).

In practice, a sound wave is considered as a light ray that is reflected on a surface (that behaves like a mirror) with the same angle of incidence. This assumption is valid if the following hypotheses are valid:

1. the dimensions of the walls are large;
2. the curvature and the irregularities of the reflecting surface are small, with respect to the wavelength of the considered sound.

The main methods developed, following the paradigm of Snell's geometric laws, are two: i) the image method; and ii) the ray tracing algorithm. In both methods all the walls are assumed flat and uneven walls are approximated with flat walls. In practice, the two methods differ only in the calculation algorithm which is based on slightly different paradigms and in boundary conditions would come to the similar result.

### *3.6.1 Coarse Evaluation of the Reverberation Time*

For some acoustic environments whose geometry and absorption coefficients of the walls are known, in the literature many different formulas are available for the calculation of the average RT corresponding to different models of approximation and related to different applications. The most used ones in practice will be mentioned below.

**Sabine equation** The Sabine equation provides more accurate results when the mean value of the absorption coefficient is $\alpha < 0.2$. Neglecting the effects of air absorption above 4 kHz, the following approximation is obtained

$$T_{60} \approx \frac{1}{6} \cdot \frac{V}{S\bar{\alpha}} = \frac{1}{6} \frac{V}{\sum\limits_n S_n \alpha_n}.$$

**Eyring equation** This type of approximation is used when the mean value of the absorption coefficients is $\alpha > 0.2$ and for frequencies below 4 kHz. The second term in the denominator is added to take into account the air absorption of high-frequency

$$T_{60} = -\frac{1}{6} \cdot \frac{V}{S \cdot \ln(1-\bar{\alpha})}.$$

***Remark 3.9.*** Observe that, in both previous equations the absorption coefficients are frequency dependent. Moreover, all the above approximations make use of the absorption coefficients of the materials. However, for the calculation of the $T_{60}$ it is necessary to pay attention to the fact that the theoretical absorption coefficients are almost always higher than the real value.

## *3.6.2 Image Method*

The image method is used in many research fields (microwaves, acoustics, etc.) and a very large literature with criticisms, improvements, specializations, etc. is available [21]. The essential idea is that every wall, or plane of reflection, behaves like a mirror (acoustic) and is replaced by an *image source* (in literature also appealed as a *phantom-*, or *virtual-source*). In the case of a rectangular room the adjacent walls are perpendicular to each other and many images coincide. Furthermore, the images in the three directions $(x, y, z)$ are independent of each other: the number of valid images and paths can therefore be easily calculated.

To understand the image method, we first see a simple case, illustrated in Fig. 3.27, with two walls not perpendicular to each other and consider the images up to the second order. In Fig. 3.27-a) with $I_2^1$ is indicated the first order image relative to the wall 2 of the acoustic source (image of order 1 of the wall 2). With $I_{1,2}^2$ we indicate the second order image generated by that of the first order and relative to wall 1: the superscript indicates the order of the image, while the subscript numbers indicate the sequence of walls that generated the image. That is, with $I_{1,2}^2$ is indicated the second-order image generated by a first reflection of wall 2 and a second reflection of wall 1.

High order reflections are calculated with a backtracking algorithm. To calculate the propagation path of the $n$ order image, it starts from the receiver (microphone) and goes back until it meets the source. The method can easily be extended to the three-dimensional case.

Again with reference to Fig. 3.27-a), starting from the receiver the ray meets the wall 1, the wall 2 and finally reaches the source. In this case, the $I_{1,2}^2$ is a *valid image*. Considering the case illustrated in Fig. 3.27-b), the ray that starts from the receiver meets the wall 1, reaches the $I_{1,2}^2$ image. Continuing with the backtracking, tracing the

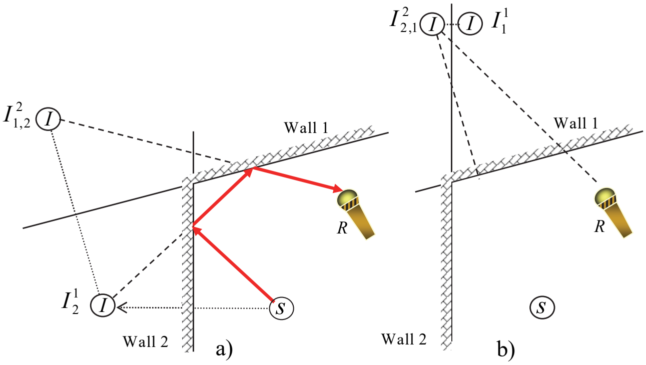


**Fig. 3.27** Image method.

perpendicular to the wall 1, the wall 2 is never encountered. It is therefore impossible to find a ray emitted by the source that passes through the wall 2, the wall 1 and arrives at the source: the image $I^2_{1,2}$ is therefore said *invalid image*. In addition, note that in the case of perpendicular walls, the images $I^2_{1,2}$ and $I^2_{2,1}$ would be overlapped.

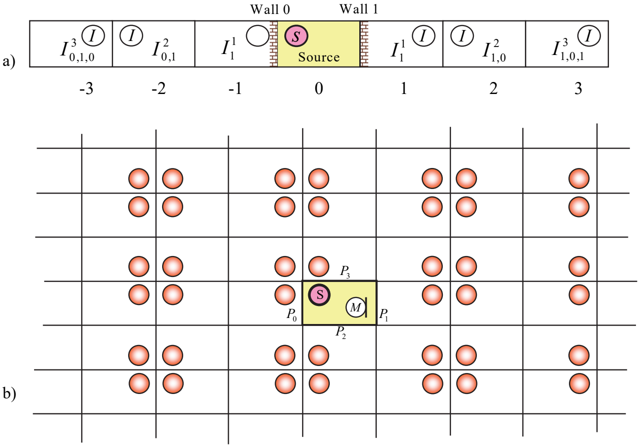


**Fig. 3.28** Image method: a) for a simple linear 1D structure; b) for a rectangular 2D structure.

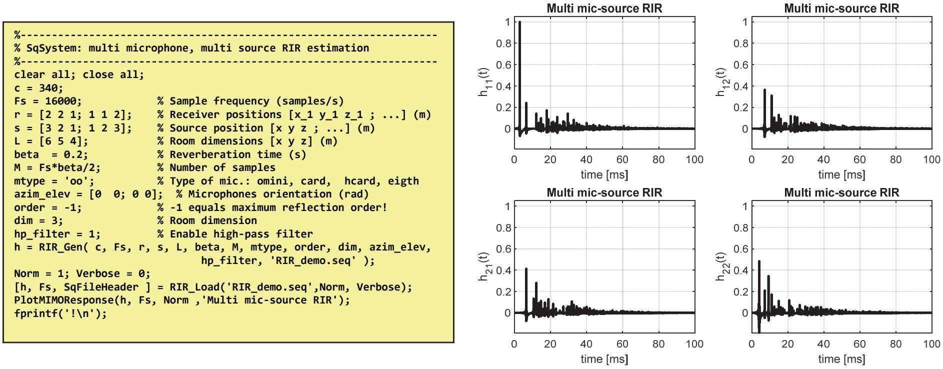
In accordance with the backtracking algorithm, and with the image validation criterion, it is possible to calculate all the acoustic paths propagated through the images of order $n$ for any closed environment. For the realization of the algorithm all images must be determined. For example, for a closed environment with 6 walls, the source generates 6 images of the first order. Each image of the first order generates, in turn, 5 images of the second order, each of which generates 5 images of the third order and so on. In total for a simulation of order $n$ the number of images to be considered and to be validated can be calculated as

$$N_{imm}^{n} = \sum_{i=1}^{n} 6 \times 5^{i-1}.$$

In the case of a rectangular structure, considering the symmetries, there is a considerable reduction in the computational cost. For example, in Fig. 3.28-a) we consider the case of a simple 1D linear structure. In the cells are shown the locations of the image sources up to the third order. Cell 0, is called “source” $S$. From the figure we can see how each cell contains only one image. For example, in cell 1 there is an image of the source $I_1^1$ relative to the wall 1. Higher orders are found in cells $\pm 3$, $\pm 4, \pm 5$, ...; in the sequence associated alternately with the wall 0 and 1. The image of order $n$ is located in the cell such that its Manhattan distance[2] from the source is equal to $n$.

In the case of a rectangular 2D structure, as Fig. 3.28-b) shows, also in this case the configuration of the sources is regular and easily calculable . For non-regular structures the image method is quite complex [27].

For example, in Fig. 3.29 is shown the trend of a typical impulse responses calculated with a simulation program that implements the image algorithm described in [21], with “Roomsim” calculation engine[3] [22].



```
%-------------------------------------------------------------------
% SqSystem: multi microphone, multi source RIR estimation
%-------------------------------------------------------------------
clear all; close all;
c = 340;
Fs = 16000;            % Sample frequency (samples/s)
r = [2 2 1; 1 1 2];    % Receiver positions [x_1 y_1 z_1 ; ...] (m)
s = [3 2 1; 1 2 3];    % Source position [x y z ; ...] (m)
L = [6 5 4];           % Room dimensions [x y z] (m)
beta  = 0.2;           % Reverberation time (s)
M = Fs*beta/2;         % Number of samples
mtype = 'oo';          % Type of mic.: omini, card,  hcard, eigth
azim_elev = [0  0; 0 0];  % Microphones orientation (rad)
order = -1;            % -1 equals maximum reflection order!
dim = 3;               % Room dimension
hp_filter = 1;         % Enable high-pass filter
h = RIR_Gen( c, Fs, r, s, L, beta, M, mtype, order, dim, azim_elev,
                              hp_filter, 'RIR_demo.seq' );
Norm = 1; Verbose = 0;
[h, Fs, SqFileHeader ] = RIR_Load('RIR_demo.seq',Norm, Verbose);
PlotMIMOResponse(h, Fs, Norm ,'Multi mic-source RIR');
fprintf('!\n');
```

**Fig. 3.29** Impulse responses for a room (6 × 5 × 4) [m] with 2 sources and 2 microphones, calculated with the *Roomsim* engine [21], [22].

## *3.6.3 Ray Tracing Method*

In the case of the *ray tracing method* (RTM) the model assumed consists in considering an acoustic source that emits a finite number of sound beams (or rays) that are modeled as “sound particles” with radial emission [28]. In their path these particles

[2] The Manhattan distance (MD) is defined as the distance between two points measured on parallel paths to the coordinated axes $x$ and $y$ making a right angle for change of direction. In a plane with the point $p_1$ in $(x_1, y_1)$ and $p_2$ in $(x_2, y_2)$, the MD is equal to $|x1 - x2| + |y1 - y2|$.

[3] To download Roomsim: https://pims.grc.nasa.gov/plots/user/acoustics/

(also called “tokens”) encounter obstacles (the walls, the floor, etc.) that when they are hit in turn emit other particles.

The main problem of this method, as already observed by Krokstad [28], consists in the fact that an infinite number of rays must be approximated with a finite number. Furthermore, the choice of such rays is critical as it is not assured that a beam emitted reaches the source.

For simple environments the number of beams to be emitted is not very high. In the case in which the structure is complicated, however, the number of beams to be emitted such as to ensure a certain probability of reception could grow uncontrollably. In this regard, various strategies have been developed for the control of the emitted rays, such as those of the methods called pyramid tracing [29] and cone tracing [30].

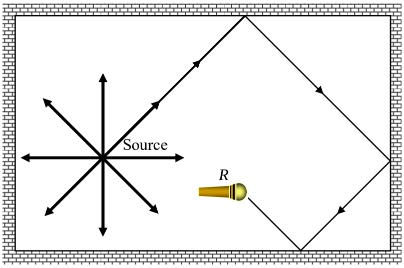


**Fig. 3.30** Calculation of the impulse response of a room with the ray tracing method. Every emitted beam by the source is reflected, attenuated and refracted by every obstacle it encounters in its path.

Another not negligible problem of RTM-based methodologies concerns the size of the receiver. A receiver of infinitesimal dimensions, in fact, would not intercept any ray. Among the various proven forms it seems that the spherical one gives the best acoustic results. Lehnert *et al.* [31], for example through a simple reasoning (shown in Fig. 3.31), proposed a receiver size with variable size and proportional to the length traveled by the beams.

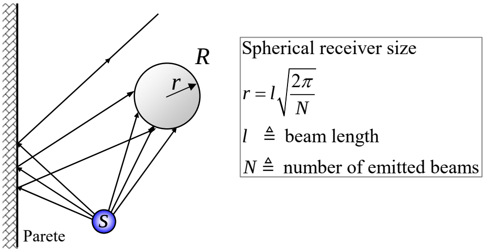

R
r
S
Parete
Spherical receiver size
$r = l\sqrt{\frac{2\pi}{N}}$
$l \triangleq$ beam length
$N \triangleq$ number of emitted beams

**Fig. 3.31** The three direct beams emitted by the source intercept the receiver (spherical). Considering the three reflected rays only two intercept the receiver.

One of the advantages of the RTM is that it is able to easily insert the modeling of the phenomena associated with the propagation as diffraction effects, diffusion phenomena that can also be frequency dependent.

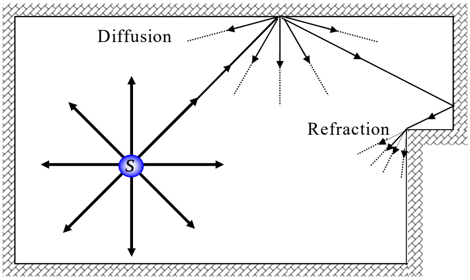


**Fig. 3.32** The RTM allows to model the phenomena of refraction and diffusion.

In order to obtain an even more realistic simulation, it is possible to integrate the RT method with a CAD software, where both the listening environment and the furniture are designed, in which all the characteristics of the materials are stored in a database.

#### 3.6.3.1 Moving Sound Source Simulation

The simulation of a moving source (or a moving receiver) consists in determining the impulse response variable over time according to the position of the source. Considering the coordinates of points $\mathbf{p} = [x\,y\,z]^T$, the impulse response can be expressed as $h(\mathbf{p},t)$. The estimate of the $h(\mathbf{p},t)$ can be done by sampling the trajectory and applying one of the previously described techniques for each point.

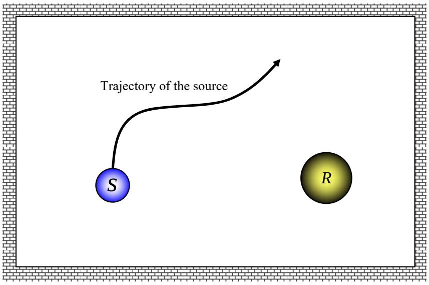


**Fig. 3.33** Model of a moving sound source used to compute the time-variant impulse response.

The time-variant impulse response is approximated as a sequence of time-invariant impulsive responses $h_{\mathbf{p}_i}(t)$ indicated as

$$h_{\mathbf{p}_i}(t) \approx [h_{\mathbf{p}_1}(t),\; h_{\mathbf{p}_2}(t), \ldots].$$

Thus, the computational cost increases linearly with respect to the number of points considered.

***Remark 3.10.*** With the RIR so approximated it is impossible to obtain physical phenomena such as the Doppler effect. In this case the speed of the source rather than its position is decisive. As we will see later (Chapters 7 and 9) this can be done by considering acoustic propagation models using digital transmission lines of varying time length.

## 3.7 Quality Indexes of a Listening Room

The characterization of a listening environment or an auditorium in terms of quality is a rather complex task as it depends on the formal definition of the end user, i.e. the human listener, which by its nature has subjective characteristics. Furthermore, for the overall quality, it is also necessary to take into account the perceived quality of the performer, i.e. the musicians on stage, also strongly influenced by the presence of reverberation and other acoustic parameters.

However, over the years, we have tried to establish objective criteria based on a number of measurable parameters, on average linked to subjective quality perception. The formal definition and the exact procedures for measuring these parameters are defined by ISO (International Organization for Standardization) in the standards [23], [24] and in other Standardization organization as the IEC (International Electrotechnical Commission) [36]. In the past, the evaluation of the quality index of a room was based only in the reverberation time. For example, some parameters referred to as a "good balance" between direct-field energy and reverberation, are defined as the early-reflection/diffused-sound ratio.

However, reverberation by its global nature, proved insufficient for a complete listening room characterization and, although, almost all the parameters are derived from the RIR or echogram structure, all parameters that can be measured, are frequency-dependent.

A distinction must be made between the evaluation of the acoustic quality of rooms intended for listening to music and rooms for listening to speech. Subsequently, several variants were proposed, almost always defined considering the RIR as a starting point, some of which are shown below

### 3.7.1 Speech Quality Index

In the case of speech signal, the fundamental requirement is correct comprehension of the transmitted message, that is, its intelligibility understood as an understanding of the transmitted message, that is its intelligibility, understood as a percentage of words or phrases correctly understood by a listener compared to the totality of the sentences pronounced by a speaker.

#### 3.7.1.1 Definition Index $D_{50}$

The *definition index* $D_{50}$, introduced by Meyer and Thiele [32], is defined as the ratio between the useful and the most disturbing sound. Based on subjective evidence, assuming that the useful sound is the one that arrives in the first 50 ms, the $D_{50}$ is defined as

$$D_{50} = \frac{\int_0^{50ms} h^2(t)dt}{\int_0^{\infty} h^2(t)dt} \tag{3.19}$$

where $h(t)$ indicates the RIR response (i.e. the acoustic pressure), measured at the listening point.

The intelligibility of the syllables in the spoken word is greater the higher the value of $D_{50}$. The optimal values of the definition index are higher than 50% for the music; while for speech, less than 50% for music.

### 3.7.1.2 Clarity Index $C_{50}$

Proposed by Reichard *et. al.* [32] the *clarity index* $C_{50}$ is the sound-of-interest/reverberation that can be defined as

$$C_{50} = 10\log_{10}\frac{\int_0^{50ms} p^2(t)dt}{\int_{50ms}^{\infty} p^2(t)dt}. \tag{3.20}$$

This is exactly related to $D_{50}$ by the following relationship

$$C_{50} = 10\log_{10}\left(\frac{D_{50}}{1-D_{50}}\right) \text{ dB}$$

optimal values are: $C_{50} > 3$ dB.

### 3.7.1.3 Speech Transmission Index

Among the numerous parameters regarding intelligibility, the most important is the *Speech Transmission Index* (STI). The STI is a physical-perceptive metric based on the estimation of some physical measurable parameters of a generic transmission channel (a room, electro-acoustic equipment, telephone line, etc.), the metric is well correlated with the intelligibility of speech degraded by additive noise and reverberation.

The method, developed since 1971 by Houtgast *et. al.*, [34], [35], is based on a amplitude modulation model of speech production

In accordance with the IEC-60268-16 [36], the STI method is based on the determination of the modulation index, here denoted *modulation transfer funcion* (MTF) and indicated as $m(F)$, for a 98 points, obtained for 14 modulation frequencies over the range $F \in (0.63, 12.5]$ Hz spaced as a one-third octave interval, for seven frequency bands defined by an octave filter banks from 125 Hz up to 8 kHz (included).

***Remark 3.11.*** Note that, these low modulation frequencies correspond, roughly, to the envelope of the average vocal signal. In addition, the used octave frequency bands are related to the typical frequency range of a human voice. The female voice spectrum model does not include the 125 Hz octave band.

For the effects of the reverberation, as illustrated in Fig. 3.34, at the listening point the received signal has a modulation index less deep, in fact the “valley” are partially filled by the background noise in the room and by the reverberation tail of the previous “mountain”.

For each of the bands considered and for each value of the modulation, from the envelope of the received signal it is possible to estimate the signal-to-noise ratio (S/N) as

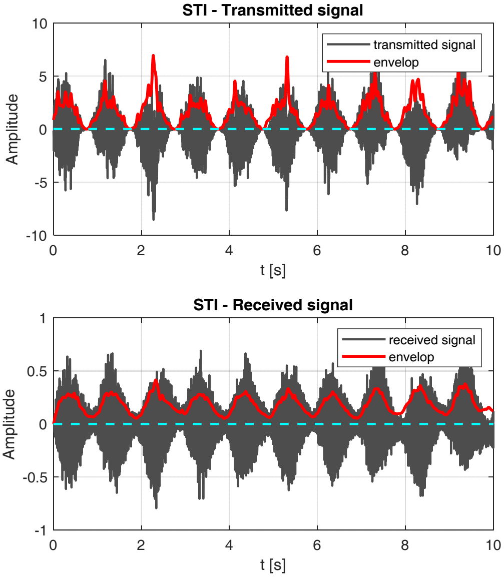


**Fig. 3.34** Transmitted source and received signal at a distance of 4 m, for a room $(5 \times 7 \times 3.1)$ with $T_{60} \approx 1.5$, with a modulated signal of 1 Hz. For the effects of reverberation at the listening point, the received signal has a less profound modulation index. Note that in the example, the envelope is calculated as smooth $\left(\sqrt{(r[n])^2}\right)$, where with ”smooth” we mean a fifth-order Butterworth low-pass filter with $f_t = 16$ Hz, implemented in zero-phase.

$$\left(\frac{S}{N}\right)_{f_o,F} = 10\log_{10}\left(\frac{m(F)}{1-m(F)}\right)$$

where with $m_F$ is indicated the frequency modulation index at frequency $F$ .

For each band of the filter bank the he overall S/N, is calculated, as average (with some precautions to eliminate any calculation errors), for all 14 frequencies considered. Then these values are averaged in order to obtain only one value, the STI.

If the calculations are limited to the 500 and 2k Hz bands only, the resulting average is called *rapid-STI* (RASTI).

A more immediate way to estimate the STI and RASTI defined by Schroeder in [37], indicated as *Schroeder indirect method*, it is based on the direct estimation of the modulation transfer function, defined by the ratio between the component at frequencies $F$ of the Fourier transform of the echogram and its integral, for which

$$m_k(F) = \frac{\int_0^\infty h_k^2(\tau)e^{-j2\pi F\tau}d\tau}{\int_0^\infty h_k^2(\tau)d\tau}, \quad k \in [1,7], \quad f_m \in [1,14]$$

where $h_k(t)$ is the response to the filtered impulse for the $k$-th noise carrier band, and $f_m$ is the $m$-th modulation frequency.

However, this value of $m(F)$, does not take into account the effect of background noise, but only the effect of the reverberating tail of the previous “modulation valley”. Therefore, if we want to derive the value of MTF from impulse response measurements, the value of the $m(F)$ above must be corrected (reduced) to take into account the signal-to-noise ratio.

$$m_k^*(F) = m_k(F) \cdot \frac{1}{1+10^{\left(\frac{-\mathrm{SNR}_k}{10}\right)}}$$

where $\text{SNR}_k$ is the signal-to-noise ratio in dB of the $k$-th analysis band.

The advantage of the indirect method is that with only three measurements, against the 98 measurements necessary for the direct method, it is possible to estimate the STI index.

#### 3.7.1.4 STI for Public Address Systems

There are situations where the indirect method is not applicable, and as mentioned the STI direct method is too complex to use. Therefore, simplifications of the direct STI method have been developed which use only a subset of 98 carrier/modulating combinations for the calculation.

Among these, the *Speech Transmission Index for Public Address* (STIPA) systems, which is defined as a subset of the STI, is sensitive to the distortions typical of environments and/or electroacoustic reinforcement systems.

The particularity of the STIPA is that the choice of modulation frequencies and the modulation index for each carrier, means that it is possible to superimpose all 7 carriers modulated in a single signal with the possibility to filter the output signal to the system and demodulate each band in parallel. This makes it possible to estimate the STIPA index through the direct method with only one measurement lasting about 20 seconds against the 98 measurements required by the complete STI.

Finally, it should be noted that the intelligibility measures are of great importance for classrooms, auditoriums and theaters. In fact, it is known that a high intelligibility entails a high signal-to-noise ratio and a reduced reverberation time.

For more details, refer to the IEC-60268-16 [36].

### *3.7.2 Quality Indices for Music*

The evaluation of the perceived sound quality depends on the performed music genre (see Fig. 3.21) for which the measured parameters must be contextualized according to the type of representation considered. In general, for the listener in the audience, we have five sound parameters denoted as :

1. Subjective level of sound (neither too high nor too low).
2. Perceived reverberation (neither too dry nor too reverberant).
3. Perceived clarity (the preferred value varies from high for speech to low for choir and organ music).
4. Apparent source width (sound reflections from the side walls contribute to an audio perception that the sound comes from a wide source, not from a point).
5. Listener envelopment (the feeling of being embedded in sound).

For the musicians on stage, two other parameters are :

1. Ensemble conditions (how well the musicians can hear each other).
2. Perceived reverberation amount (how well the musician can hear the room's response to his/her instrument).

#### 3.7.2.1 Clarity Index $C_{80}$

Proposto da Reichard *et. al.* [32] the *clarity index* $C_{80}$ is an extension of the clarity index $C_{50}$ applied to music signals, introduced to evaluate “musical transparency” (clear perception of musical notes played in rapid succession) and harmonic transparency (ability to distinguish notes of one or more instruments played concurrently). The clarity index is defined by the following ratio

$$C_{80} = 10\log_{10}\frac{\int_0^{80ms} p^2(t)dt}{\int_{80ms}^{\infty} p^2(t)dt}. \tag{3.21}$$

the increase of the integration time of the sound useful to 80 ms, derives from the considerations: 1) the integration time of the ear is longer for the music than for the speech; 2) the transients, for most of the instruments, lasts less of 100 ms.

Like many of the acoustic qualities discussed, $C_{80}$ is dependent upon frequency. For example the parameter indicate as $C_{80}(3)$, has been defined as the average of $C_{80}$ values at 3 frequency octave bands centered at 500 Hz, 1000 Hz, and 2000 Hz.

In general, acceptable values per una corretta trasparenza è per alcuni autori $C_{80} \pm 1.6$ dB mentre per altri $-4 \leq C_{80} \leq 2$ dB.

#### 3.7.2.2 Sound Strength $G$

The robustness index $G$, analyzes the sound intensity that the listener perceives in a point of the room, comparing it with the intensity response that would give the same omnidirectional source in the free space. It is defined as the ratio between the RIR in the observation point at an impulse emitted by an omnidirectional source on the stage and the response to the same impulse in a fixed point of the room at $s$ distance from the source (located at distance of 5 m or 10 m); $\Delta t$ represents the duration of the direct impulse.

$$G = 10\log_{10}\frac{\int_0^{\infty} p^2(t)dt}{\int_0^{t+\Delta t} p^2(s,t)dt}. \tag{3.22}$$

It is therefore the parameter that represents the amplification effect of a room. Of course it is a parameter that varies according to the frequency is then measured for six octave bands. It is considered particularly important the average value of $G$, which is indicate as $G_m$, measured with the octave band filters for frequencies ranging from 500Hz to 1kHz.

It is also considered extremely important the $G_{low}$, or its average value, linked to the low frequencies ranging from 125 to 250 Hz, as this index is related to the perceived intensity of the bass and certain aspects of the spatiality of the sound.

Optimal $G$ values suggested in literature are the following:

- great symphony orchestra, very trained singers $G_m \geq -4$ dB
- little orchestra, singers $G_m \geq 1$ dB
- speakers, trained actors $G_m \geq 6$ dB
- weak instruments, weak speakers $G_m \geq 11$ dB.

### 3.7.2.3 Lateral Fraction $L_f$

The lateral fraction $L_f$ represents the percentage of lateral energy compared to the total. As shown in Fig. 3.35, it is measured considering the relationship between the energy acquired by a figure-eight microphone, places at 90° with respect to the source, and the energy energy acquired by an omnidirectional microphone. In accordance with [23] the lateral fraction defined as

$$L_f = \frac{\int_{0.05ms}^{80ms} p_\infty^2(t)dt}{\int_0^{80ms} p^2(t)dt} \tag{3.23}$$

where $p_\infty^2(t)$ is the instantaneous sound pressure in the auditorium impulse response measured with a figure-of-eight pattern microphone.

Optimal values, considering the average between 125Hz and 1kHz, should be $0.1 < L_f < 0.35$.

Observe that the lateral fraction is related to the IACC as

$$L_f \approx 1 - IACC(\tau).$$

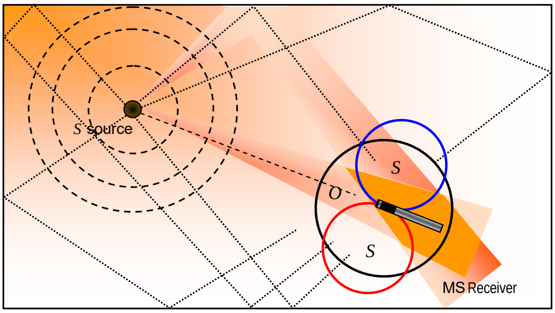


**Fig. 3.35** Measure of lateral fraction with two coincident microphones omnidirectional and figure-of-eight. The evaluation is given by the energy ratio between the O and S microphones.

### 3.7.2.4 Support Index $ST1$

For the podium or stage area, support $ST1$ corresponds to the subjective feeling how a musician perceives his/her instrument with respect to other instruments

$$ST1 = \frac{\int_{20ms}^{100ms} p_\infty^2(t)dt}{\int_{0ms}^{10ms} p^2(t)dt}. \tag{3.24}$$

It is an index that tries to quantify the right balance between the sound emitted directly by the musician and the one that the orchestra chamber and the room give back to him.

# References

1. W. C. Sabine . “Reverberation”, in Acoustics: Historical and Philosophical Development, ed. by R. D. Lindsay, Dowden, Hutchinson, and Ross, Stroudsburg, PA (1972), 1900.
2. A.D. Pierce, “Acoustics”, American Institute of Physics, for the Acoustical Society of America, 1989.
3. G. Mondaca Lo Giudice, S. Santoboni, “Acustica”, Masson 1995.
4. W.W. Seto, “Acoustic”, McGrawHill, Inc. New York, 1971.
5. H. Kuttruff, “Room Acoustics”, Fourth edition, Spon Press, ISBN 0-419-24580-4, 2000.
6. Y. Haneda, S. Makino, and Y. Kaneda, “Common acoustical pole and zero modeling of room transfer functions, “ IEEE Transactions on Speech and Audio Processing, vol. 2, no. 2, pp. 320 - 328, April 1994.
7. Y. Haneda, Y. Kaneda, and N Kitawaki “Common-Acoustical-Pole and Residue Model and Its Application to Spatial Interpolation and Extrapolation of a Room Transfer Function”, IEEE Transactions on Speech and Audio Processing, vol. 7, no. 6, pp. 709 - 717, Nov. 1999.
8. G. Bunkheila , R. Parisi and A. Uncini, “Model order selection for estimation of Common Acoustical Poles”, IEEE International Symposium on Circuits and Systems, May 2008
9. D.D. Rife and J. Vanderkooy, “Transfer Function Measurement with Maximum-Length- Sequence”, J. Audio Eng. Soc. Vol. 37, June 1989.
10. M.R. Schroeder, “New Method of Measuring Reverberation Time,” J. Acuost. Soc. Am., Vol. 37, pp. 409-412, 1965.
11. A. Farina, “Simultaneous measurement of impulse response and distortion with a swept-sine technique”, in AES 108th convention, (Paris), Feb. 2000.
12. A. Farina, “Advancements in impulse response measurements by sine sweeps”, in Audio Engineering Society Convention 122, Audio Engineering Society, May 2007.
13. A. Farina, “Impulse Response Measurements,” 23rd Nordic Sound Symposium, Bolkesjø (Norway), September 2007
14. Y. Ando, “Subjective Preference in Relation to Objective Parameters of Music Sound Fields with a Single Echo”, J. of Acoust. Soc. Am., Vol. 62, pp.1436-1441, Dec. 1977.
15. Y. Ando, “Concert Hall Acoustics”, Springer-Verlag, 1985.
16. A. Farina, “Acoustic quality of theatres: Correlations between experimental measures and subjective evaluations”, Applied Acoustics 62(8):889-916, 2001.
17. M. Gerzon, “Recording Concert Hall Acoustics for Posterity,” JAES Vol. 23, Number 7 p. 569, 1975.
18. A. Farina, R. Ayalon, “Recording Concert Hall Acoustics for Posterity,” 24th AES Conference on Multichannel Audio, Banff, Canada, 26-28 June 2003
19. A. Farina, L. Tronchin, “Measurement and numerical simulation of Binaural and B-format impulse responses in concert halls,” Proc. of 1st Inter. Symposium on Temporal Design, Kobe, 2003.
20. A. Avni, B. Rafaely, “Interaural Cross Correlation and Spatial Correlation in a Sound Field Representation by Spherical Harmonics”, Ambisonic Symposium, Graz, June 25-27, 2009
21. J.B. Allen, D.A. Berkley, “Image method for efficiently simulating small-room acoustics”, J. Acoustic Soc. Am., Vol. 64(4), pp. 943-950, 1979.
22. D.R. Campbell, K.J. Palomäki, G.J. Brown, “Roomsim, a MATLAB Simulation of “Shoebox” Room Acoustics for use in Teaching and Research”, The 2004 European Signal Processing Conference (EUSIPCO-2004), 2004.
23. ISO 3382-1997, “Acoustics - Measurement of the reverberation time of rooms with reference to other acoustical parameters”, 2nd ed, International Organization for Standardization, Genéve, 1997.
24. ISO 3382-1:2009, “Acoustics – Measurement of room acoustic parameters – Part 1: Performance spaces”, International Organization for Standardization, Genéve, 2009.
25. J.-M. Jot, “An analysis/synthesis approach to real-time artificial reverberation,” in Proceedings of the International Conference on Acoustics, Speech, and Signal Processing, San Francisco, (New York), pp. II.221-II.224, IEEE Press, 1992.

26. R. Stewart, M. Sandler, Mark, "Database of Omnidirectional and B-Format Impulse Responses," in Proc. of IEEE Int. Conf. on Acoustics, Speech, and Signal Processing (ICASSP 2010), Dallas, Texas, March 2010.
27. J. Borish, "Extention of the image model to arbitrary polyhedra",J. Acoustic Soc. Am., Vol. 75, pp. 943-950, Aprile, 1984.
28. A. Krokstad, S. Strom, and S. Sorsdal, "Calculating the Acoustical Room Response by Use of a Ray Tracing Technique", Journal of Sound and Vibration, Vol. 8, No. 1, pp. 118-125, 1968.
29. T. Lewers, "A computer model of auditoria acoustics using a combination of ray-tracing, images, and radiosity methods", Int.Symp. On Computer Modelling and Prediction of Objective and Subjective Properties of Sound Fields in Rooms, Copenhagen, Denmark, & Gothenburg, Sweden, 1991.
30. J.P. Vian, and D. Van Maerke, "Calculation of the Room Impluse Response Using a Ray-Tracing Method", Proceedings of the Vancouver Symposium on Acoustics and Theatre Planning, Vancouver, August 1986.
31. H. Lehnert, J. Blauert, "Principles of binaural room simulation", Applied Acoustics, Vol. 36, pp. 259-291, 1992.
32. E. Meyer and R. Thiele. "Raumakustische untersuchungen in zahlreichen konzertsälen und rundfunkstudios unter anwendung neuerer messverfahren", Acustica, 6, 425-444 (1956).
33. W. Reichardt, O.A. Alim, and W. Schmidt, "Definition und Messgrundlage eines objektiven Masses zur Ermittlung der Grenze zwischen brauchbarer und unbrauchbarer Durchsichtigkeit beim Musikdarbietung," Acustica, 32, 126-137, 1975.
34. T. Houtgast, and H.J.M Steeneken, "Evaluation of speech transmission channels by using artificial signals", Acustica 25, 355-367, 1971.
35. T. Houtgast, and H.J.M Steeneken "A review of the MTF concept in room acoustics and its use for estimating speech intelligibility in auditoria", J. Acoust. Soc. Am. 77, 1069-1077 1985.
36. IEC 60268-16:2011 "Sound system equipment – Part 16: Objective rating of speech intelligibility by speech transmission index," International Electrotechnical Commission, Geneva, Switzerland 2009.
37. M. R. Schroeder, "Modulation transfer function: definition and measurement," Acustica 49, 179-182, 1981.
38. J. H. Rindel,"Improving acoustics from the concert hall to the office", 'ISO Focus+, Magazine of the International Organization for Standardization (ISO), ISSN 2226-1095, Vol. 3, No. 10, Nov.Dec, 2012.