

Jean-Marc Jot, Laurent Cerveau, and Olivier Warusfel
IRCAM
F-75004 Paris, France

**Presented at
the 103rd Convention
1997 September 26–29
New York**



AES

This preprint has been reproduced from the author's advance manuscript, without editing, corrections or consideration by the Review Board. The AES takes no responsibility for the contents.

Additional preprints may be obtained by sending request and remittance to the Audio Engineering Society, 60 East 42nd St., Suite 2520, New York, New York 10165-2520, USA.

All rights reserved. Reproduction of this preprint, or any portion thereof, is not permitted without direct permission from the Journal of the Audio Engineering Society.

AN AUDIO ENGINEERING SOCIETY PREPRINT

Analysis and synthesis of room reverberation based on a statistical time-frequency model

Jean-Marc Jot, Laurent Cerveau, Olivier Warusfel

IRCAM. 1 place Igor-Stravinsky. F-75004 Paris, France.

Tel: (+33) 1 44 78 48 85, Fax: (+33) 1 44 78 15 40

E-mail: [first_name].[last_name]@ircam.fr

<http://www.ircam.fr>

Abstract

This paper reviews a statistical time-frequency model of late reverberation decays, and describes an associated analysis procedure for deriving the time-frequency envelope of the reverberation from a measured impulse response, based on the notion of "Energy Decay Relief". The models and techniques discussed are applied to the representation of room acoustical quality, the characterization and equalization of transducers in diffuse fields, the restoration (denoising) of measured room impulse responses, as well as artificial reverberation and distance effects in virtual auditory displays and audio mixing systems.

0 Introduction

A stochastic model of room responses was proposed by Schroeder in the 1950s and further developed more recently by Polack [1-3]. This model is reviewed in the first section of this paper and applies to late diffuse reverberation decays, which will be the main concern of this paper. Statistical models bear a particular importance for auralization and synthetic reverberation, because they imply that it is possible to replace the later part of a room impulse response by an exponentially decaying random signal. Consequently, an artificial reverberation system can be controlled via a simple representation in terms of a *time-frequency envelope*, characterized by an initial spectrum and the reverberation decay rate vs. frequency, without necessarily relying on measured or computed impulse responses.

A real-time artificial reverberation algorithm can either be based on multiple feedback delay networks (FDNs) or on recently developed zero-delay fast convolution techniques [4]. FDNs offer the advantage of producing multiple uncorrelated output channels and arbitrarily long reverberation decays for no increase in computational cost, while providing an efficient control method for continuous adjustment of synthetic reverberation parameters. Although they do not allow exact reproduction of a particular room response, FDNs provide a valid approach to reverberator design in the framework of the above stochastic models [5-7].

To validate this model and simulate the reverberation of an existing room, it is necessary to develop an accurate analysis procedure allowing to estimate the decay characteristics from a measured impulse response, with high resolution in the frequency domain and sufficient immunity with respect to the measurement noise. The time-reversed energy integration procedure initially proposed by Schroeder [8] was extended for this purpose, leading to a time-frequency representation termed the *Energy Decay Relief* (EDR) [5, 6]. The definition and properties of the EDR are discussed in the second part of this paper, along the lines of [5]. In the third part, its relation to the time-frequency envelope is established and the development of a practical calculation procedure is outlined for accurately estimating the parameters of the reverberation decay from a measured impulse response [5, 9-11].

A measured impulse response can seldom be used directly for auralization purposes because it contains measurement noise and the influence of the directivity characteristics of the transducers used for the measurement. The above stochastic model can be used to restore the part of the later impulse response that is corrupted by measurement noise [12, 9, 10]. To correct for source and receiver directivity characteristics, the statistical time-frequency model must be extended by separating the influence of the room from the influence of the transducers. As shown in part 4, analyses of impulse responses measured in existing rooms support the following model [5]: the time frequency envelope does not depend on source and receiver locations in a given room, and the initial spectrum is proportional to the product of the diffuse-field transfer functions of the source and the receiver.

Several applications of the above models and techniques are described in part 5, including:

- the EDR as a perceptually relevant time-frequency representation of room acoustical quality,
- a new technique for estimating the diffuse-field transfer function or the directivity index of a transducer in a natural environment, rather than an anechoic or reverberant chamber (this technique also applies to the diffuse-field equalization of headphones and binaural recordings),
- restoration and denoising of a measured room impulse response for evaluation and auralization,
- faithful real-time artificial reverberation using feedback delay networks, and the control of distance effects in professional audio mixing systems and virtual auditory displays.

1 Statistical model of room responses

In a room response, the modal density D_m (average number of modes per Hz) is approximately proportional to the square of frequency f , while the echo density D_e (average number of reflections per second) is approximately proportional to the square of time t :

$$D_e(t) \approx 4\pi c^3 \frac{t^2}{V} \quad D_m(f) \approx 4\pi V \frac{f^2}{c^3}, \quad (1)$$

where c (in m/s) denotes the velocity of sound, and V (in m^3) is the volume of the room. These expressions can easily be demonstrated for rectangular rooms, and can be generalized to rooms of any geometry [13, 14, 2]. For larger times and higher frequencies, both densities become very high. This provides the foundation for statistical models of room responses, as developed by Schroeder in the frequency domain [1], and more recently in the time domain by Polack [2]. The resulting time-frequency model is valid for the later reverberation decay at frequencies above the "Schroeder frequency", and is an essential basis for artificial reverberation techniques [15, 5].

1.1 Frequency-domain statistical model

Eq. (1) implies that, at high frequencies, the normal modes of a room overlap in the frequency domain, i. e. the average separation between normal frequencies is smaller than the bandwidth Δf of a mode, which can be expressed as follows (for a -3 dB attenuation):

$$\Delta f = \delta/\pi \approx 2.2/Tr, \quad (2)$$

where δ is the damping coefficient of the mode, related to the reverberation time Tr by:

$$20 \log_{10}(e^{-\delta Tr}) = -60 \text{ (dB)} \quad \Rightarrow \quad Tr = \frac{3 \ln 10}{\delta}. \quad (3)$$

Thus, at high frequencies, any source signal will simultaneously excite several room modes. Assuming a sinewave excitation and a microphone located in the reverberant field, the signal captured by the microphone is the sum of the contributions of a large number of modes, where the phase and amplitude of each contribution varies with the recording position. Consequently, the complex frequency response can be considered as a space dependent stochastic process whose real and imaginary part are independent Gaussian processes having the same variance [1, 13, 14, 16-18]. This two-dimensional Gaussian density arises from the central limit theorem, assuming independence between modes, and implies that the amplitude frequency response follows a Rayleigh distribution. These statistical properties also apply when the complex response is considered as a frequency dependent stochastic process, for a given microphone position. They are valid irrespective of the microphone position (provided that the direct sound can be neglected compared to the reflected sound) and irrespective of the room (above a limit frequency which depends on the room).

The high modal overlap implies that the peaks in the frequency response do not correspond with the individual normal frequencies. Although the modal density increases as the square of frequency, as shown by Eq. (1), the average separation between adjacent peaks in the amplitude frequency response only depends on modal bandwidth. Accordingly, the density of peaks in the frequency domain is proportional to the reverberation time [16, 13]:

$$\text{average number of maxima per Hz: } D_f \approx \frac{\sqrt{3}}{\delta} \approx \frac{Tr}{4}. \quad (4)$$

This statistical model relies on the assumption of high modal overlap in the frequency domain, which is not verified at low frequencies. The "Schroeder frequency", above which the theory is valid, as been verified experimentally [16]:

$$f_{\text{Schroeder}} \approx 2000 \sqrt{\frac{Tr}{V}} \quad (\text{Hz}). \quad (5)$$

By combining equations (1, 2, 5), one can verify that this condition corresponds to a modal overlap $\Delta f D_m(f)$ at least equal to about 3:1 (the average spacing between normal frequencies must be less than one third of the bandwidth of a mode for the theory to be valid).

1.2 Time-domain statistical model

Moorer noted the auditory resemblance between a concert hall impulse response and a white noise multiplied by an exponentially decaying envelope, and reported that such a synthetic response can produce, by convolution with anechoic signals, a natural sounding reverberation effect [19]. To obtain a frequency-dependent reverberation time, he suggested using a filter bank and summing the subband signals after multiplying them with different exponential envelopes (Fig. 1).

Polack [2] developed a time-domain model complementing Schroeder's frequency-domain model. In this model, a room impulse response is described as one realization of a non-stationary stochastic process:

$$h(t) = b(t) e^{-\delta t} \quad \text{for } t \geq 0, \quad (6)$$

where $b(t)$ is a centered stationary Gaussian noise, and δ is related to the reverberation time Tr by (3). The random noise $b(t)$ is characterized by its power spectral density, denoted $P(f)$. According to Polack [2], the impulse response can then be observed on two different time scales:

- A small time scale corresponding to the "fast" variations of the signal $b(t)$, i. e. to the order of the millisecond (measured by the temporal spreading of the autocorrelation function of b). This corresponds, in the frequency domain, to the scale of "slow" variations of the power spectral density $P(f)$ (typically the kilohertz).
- A large time scale corresponding to the "slow" variations of the temporal envelope, i. e. to the order of the second (measured by the reverberation time), and corresponding to the scale of the "fast" variations of the frequency response of the room (measured by the bandwidth of the normal modes according to Schroeder's theory).

Since these two scales differ from several orders of magnitude, it is possible to separate the time variable t and the frequency variable f in the calculation of statistical quantities [2]. We are more particularly interested in the energy envelopes of the response in the time and frequency domains, which can be expressed as follows [2]:

$$\langle |H(f)|^2 \rangle \approx P(f) \int_0^{+\infty} e^{-2\delta t} dt = \frac{P(f)}{2\delta} = \frac{P(f) Tr}{6 \ln 10}, \quad (7)$$

$$\langle h(t)^2 \rangle \approx e^{-2\delta t} \int_{-\infty}^{+\infty} P(f) df = \sigma^2 e^{-2\delta t}, \quad (8)$$

where $H(f)$ denotes the Fourier transform of $h(t)$, σ^2 denotes the variance of b , and the brackets denote ensemble averaging.

This simple model must be extended to account for the frequency dependence of the reverberation time, by making δ a function of frequency. To generalize (6), Polack introduces the ensemble average of the Wigner-Ville distribution $W_h(t, f)$ derived from $h(t)$:

$$\langle W_h(t, f) \rangle = P(f) e^{-2\delta(f)t}. \quad (9)$$

The Wigner-Ville distribution $W_h(t, f)$ is a real-valued function defined by:

$$W_h(t, f) = \int_{-\infty}^{+\infty} h(t - \frac{\tau}{2}) h(t + \frac{\tau}{2}) e^{j2\pi f\tau} d\tau, \quad (10)$$

The properties of the Wigner-Ville distribution give it a central role in the realm of joint time-frequency distributions of non stationary signals [20, 21]. A particularly relevant property here is the preservation of the marginal distributions, i. e. the temporal energy density $h(t)^2$ (or echogram) and the spectral energy density $|H(f)|^2$ (or power frequency response):

$$h(t)^2 = \int_{-\infty}^{+\infty} W_h(t, f) df \quad |H(f)|^2 = \int_{-\infty}^{+\infty} W_h(t, f) dt. \quad (11)$$

By integration of (9) along time, (11) allows generalizing the expression (7) of the spectral energy envelope when the decay time Tr is a function of frequency. However, the expression (8) of the time-domain envelope cannot be generalized, except for $t = 0$.

1.3 Notion of late reverberation

Equation (9) states that, in this stochastic model of reverberation decays, the *time-frequency envelope* of h (defined as the ensemble average of the time-frequency energy distribution) decays exponentially with time at each frequency, and is characterized by two functions of frequency: the reverberation time $Tr(f)$ and the *initial power spectrum* $P(f)$.

The model assumes that the time-frequency envelope varies slowly with t and f , which implies that the damping coefficient $\delta(f)$ and the spectral energy envelope $E(f)$ vary slowly with frequency, and that the reverberation time remains large, at any frequency, compared to the spreading of the initial autocorrelation function of h (inverse Fourier transform of the initial spectrum $P(f)$). Under these conditions, the "slow" variations of the amplitude frequency response are described by the spectral energy envelope, defined as the ensemble average of the power frequency response $|H(f)|^2$. According to Eq. (7), the spectral energy envelope is proportional to the product of the initial spectrum $P(f)$ and the reverberation time $Tr(f)$. The "fast" variations of the frequency response, on the other hand, can still be described by Schroeder's theory (section 1.1), although the reverberation time Tr and the damping coefficient δ are now frequency dependent. This of course still assumes frequencies higher than the Schroeder frequency (5).

In the time domain, too, there is an interval after which Polack's stochastic model becomes valid. The time-domain response can only be Gaussian if a sufficient number of reflections overlap at any time along the response. The peaks in the echogram then no longer correspond to the arrivals of individual reflections. Since the reflection density increases with time according to (1), the situation is similar to that found in the frequency domain (section 1.1), except that the "width" of a reflection in the time domain cannot be defined solely with respect to the intrinsic properties of the room (unlike the bandwidth of a mode).

The spreading of a reflection in the time domain can only be expressed with reference to the bandwidth on the source excitation (which determines the spreading of the source pulse), or to the bandwidth the receiver. If the criterion is that 10 reflections at least overlap within a characteristic time resolution of the auditory system, taken equal to 24 ms in [2], Eq. (1) leads to:

$$t_{\text{mixing}} \approx \sqrt{V} \text{ (ms).} \quad (12)$$

This value was also proposed in [22] as a reasonable approximation for the transition time between early reflections and late reverberation. Polack shows that the exponentially decaying stochastic model of section (1.2) can be established within the framework of geometrical acoustics and billiard theory [2, 3], and defines the *mixing time* as the times it takes for a set of initially adjacent sound rays to spread uniformly across the room. By that time (if the origin is taken as the time of emission of a sound pulse by the source), the reverberation process has become diffuse, i. e. the acoustical energy density and the direction of the intensity vector are uniformly distributed across the room.

The mixing character of a room depends on its geometry and the diffusing properties of the boundaries. When mixing is achieved, the echo density increases exponentially with time, rather than in t^2 [3]. Consequently, the value \sqrt{V} can be considered as an upper limit for the mixing time in typical "mixing" rooms.

Hence the conditions of validity of the stochastic time-frequency model of reverberation decays can be summarized as in Fig. 2, by limiting to frequencies higher than the Schroeder frequency, given by Eq. (5), and times later than the mixing time, for which an upper limit is given by Eq. (12). It is worth reminding that the late reverberation can be described by a stochastic model (implying no knowledge of the normal frequencies of the room) only above the Schroeder frequency, which takes a value of 100 Hz for a 400 m³ room having a reverberation time of 1 second.

Unlike typical bathrooms, concert halls are relatively absorbent with a relatively large volume, leading to Schroeder frequencies close to the lower limit of the audible range. This allows, in concert halls and similar "large rooms", for the identification of the stochastic model reviewed above with the common notion of "late reverberation".

2 Time-frequency analysis of room impulse responses

To evaluate the validity of the above stochastic model for artificial reverberation purposes, one could compare a measured room impulse response with the same response where the later part has been replaced with a synthetic random reverberation decay according to the model. This calls for an analysis method yielding accurate estimation of the reverberation time and the initial spectrum, with high resolution in the frequency domain, so that the result of the comparison will reflect only the validity of the stochastic model, and cannot be negatively affected by an inaccurate estimation of the model parameters.

A time-frequency analysis procedure was developed for this purpose in [5], based on a frequency-domain expansion of Shroeder's integrated Energy Decay Curve (EDC). The definition of the EDC and the limitations of traditional analysis methods are reviewed in the next section. The definition of the Energy Decay Relief (EDR) and its properties are discussed in the subsequent sections.

2.1 Schroeder's integrated Energy Decay Curve (EDC)

The EDC has been used traditionally to measure the reverberation time, and is defined as the RMS level recorded by a microphone after interruption of the source signal. The standard source signal is a stationary random noise emitted by a loudspeaker. The recording thus also contains a random component, which must be eliminated by averaging several recordings in order to obtain a repeatable measure, for a given position of the loudspeaker and the microphone. The reverberation time Tr is typically evaluated by doubling the time interval between -5 dB and -35 dB. The procedure is repeated with a bandpass filter inserted at the microphone output to obtain EDCs and measurements of the reverberation time in different frequency bands [8, 14].

Schroeder [8] showed that, if the excitation signal is a white random noise, the ensemble average of the level recorded at the microphone (i. e. the ideal EDC) is identical to the curve obtained from a single recording of the impulse response $h(t)$, by computing a backward integration of the echogram $h^2(t)$. The EDC is thus defined as follows:

$$EDC_h(t) = \int_t^{+\infty} h^2(\tau) d\tau, \quad (13)$$

which means that $EDC_h(t)$ equals the remaining energy in the impulse response after time t . The EDC for a given frequency band is simply obtained by filtering $h(t)$ before squaring and integration. The cumbersome previous procedure is thus replaced by a single measurement of $h(t)$ followed by post-processing with a computer, which typically involves an octave or third-octave filter bank. The accuracy of this analysis procedure is however limited by the following factors:

- The frequency resolution may be sufficient for perceptual characterization, but insufficient for analysis / synthesis. Since room responses typically exhibit reverberation times and energy levels which tend to decrease with frequency, computing EDCs in relatively wide adjacent frequency bands may result in overestimating the reverberation time vs. frequency curve (because the measure will be essentially yield the value of Tr at the lower end of each band).
- The reliability of the measurement is subject to the design of the filter bank (shape of the amplitude frequency responses of the individual filters, as well as their delay and phase properties).
- The impulse response generally contains measurement noise, which, if not attended to, distorts the EDC and biases the estimated reverberation time towards higher values.

The measurement noise can generally be considered stationary along the recorded impulse response, including when this response is obtained with current MLS measurement techniques [23]. Consequently, the effect of this noise in the integration can be largely eliminated by estimating its variance at the end and the beginning of the recorded response and modifying the summation procedure so that samples of the echogram below a given threshold (relative to the estimated noise power) are not taken into account in the integration [5, 6]. An improved method, proposed in [23], consists of subtracting the estimated noise power from all samples of the echogram before integration (assuming that the noise is uncorrelated from the actual response).

2.2 General definition of the Energy Decay Relief (EDR)

In order to formalize and generalize the above procedure for analyzing room impulse responses, we consider a general class of time-frequency representations which provide an Energy Decay Curve evaluated locally at any frequency [5, 6]. This is easily obtained, by applying Eq. (13), once one knows, at any frequency, the temporal variation of the energy in the response

A time-frequency extension of Schroeder's Energy Decay Curve can thus be defined on the basis of a time-frequency representation $\rho_h(t, f)$ obtained from $h(t)$ by means of a time-frequency distribution function ρ :

$$h(t) \xrightarrow{\rho} \rho_h(t, f).$$

Definition 1: If $\rho_h(t, f)$ is an energetic time-frequency representation of the signal $h(t)$, the associated Energy Decay Relief, denoted $EDR_h(t, f)$, is defined by:

$$EDR_h(t, f) = \int_t^{+\infty} \rho_h(\tau, f) d\tau. \quad (14)$$

Figures 4-6 show the EDR computed for a measured room response when ρ is the short-time Fourier spectrum (or spectrogram), shown on Fig. 3. The details of the computational procedure will be discussed in section 3.2. The improvement in legibility is readily apparent, and is similar to the improvement brought about by the EDC compared to the echogram represented in dB. The smoothing effect of the backward integration is further discussed in section 3.1.

The EDR can also be defined in a manner analogous to the original definition of the EDC (see section 2.1):

Definition 2: The Energy Decay Relief is defined as the ensemble average of the time-frequency representation of the reverberation decay after the interruption of the excitation signal, if this signal is a stationary white noise.

It can be shown that the two definitions above are equivalent, at least if the time-frequency distribution ρ satisfies some hardly restrictive properties [5]. Clearly, the choice of ρ influences the EDR obtained with both of these definitions, so that one should at least verify how the frequency-domain and time-domain resolutions of the time-frequency distribution ρ affect the resulting EDR. This applies of course to the traditional method of squaring the output signals of a filter bank (described in section 2.1), which can be considered as a particular choice for ρ , yielding a particular evaluation of the EDR.

The question then arises of what is the "ideal" choice of ρ for defining the EDR. Many time-frequency distributions can be considered, among which the short-time Fourier spectrum is only one particular case, which yields different results for different choices of the time-domain window used in its computation [21, 24]. In his stochastic model of reverberation decays, reviewed in section 1.2., Polack used the Wigner-Ville distribution, which satisfies the property of preservation of the marginal distributions, Eq. (11). It is known that many time-frequency distributions possess this property and that the short-time Fourier spectrum is not among them, although it deviates from it in a predictable way [21].

2.3 Archetype of the EDR

In order to select the most adequate distribution ρ , and thus the "ideal" definition of the EDR, a natural approach consists of introducing a number of desirable properties. This approach was adopted in [5] to narrow down the choice of ρ , and resulted in the definition of a unique archetype for the EDR, as outlined below.

- 1) A desirable property of the EDR is that integrating $EDR_h(t, f)$ along frequency should yield the Energy Decay Curve $EDC_h(t)$. By applying Eq. (14), it is straightforward to verify that this property is satisfied if ρ preserves the marginal distribution in the time domain, i. e. if integrating $\rho_h(t, f)$ along frequency yields $h^2(t)$:

$$\int_{-\infty}^{+\infty} \rho_h(t, f) df = h^2(t) \quad \Rightarrow \quad \int_{-\infty}^{+\infty} EDR_h(t, f) df = EDC_h(t). \quad (15)$$

- 2) The corresponding desirable property, in the frequency domain, is that $EDR_h(t, f)$ should give access to the squared Fourier transform of $h(t)$ by integration along times. This is straightforward, from Eq. (14), if ρ preserves the marginal distribution in the frequency domain:

$$\int_{t=-\infty}^{+\infty} \rho_h(t, f) dt = |H(f)|^2 \quad \Leftrightarrow \quad EDR_h(-\infty, f) = |H(f)|^2. \quad (16)$$

- 3) It is then natural to include the additional constraint that $\rho_h(t, f) = 0$ for $t < 0$ if $h(t)$ is a causal signal, i. e. that ρ should preserve the causality of signals. If ρ also verifies the above property, Eq. (16) then becomes:

$$\int_{t=0}^{+\infty} \rho_h(t, f) dt = |H(f)|^2 \quad \Leftrightarrow \quad EDR_h(0, f) = |H(f)|^2, \quad (17)$$

which means that the power frequency response is directly read on the EDR at time 0.

- 4) We would like to generalize the above result by finding an interpretation for the curve read on the EDR at any fixed time τ . This naturally suggests considering a truncated version of $h(t)$: $h'(t) = h(t) \mathcal{U}(t-\tau)$, where $\mathcal{U}(t)$ denotes the Heaviside step function. A simple property will then ensure that $\rho_h(t, f) = 0$ for $t < \tau$. This property is that ρ should preserve temporal shifts (i. e. shifting h causes an identical shift on ρ_h), which is classically considered for time-frequency distributions [21]. If ρ verifies this property and the two previous ones, we have:

$$EDR_h(\tau, f) = \int_{t=\tau}^{+\infty} \rho_h(t, f) dt = |H'(f)|^2 = \left| \int_{t=\tau}^{+\infty} h(t) e^{j2\pi ft} dt \right|^2.$$

which represents the spectral distribution of the energy remaining after time τ in $h(t)$.

- 5) Finally, one last constraint should be included, so that the above interpretation applies to the EDR of h , and not only of its truncated version h' : we impose that $\rho_{h'}(t, f)$ be identical to $\rho_h(t, f)$ for $t > \tau$. This is equivalent to the condition that ρ should be *anticausal*, i. e. that $\rho_h(t, f)$ only depends on values of $h(t)$ for $t > \tau$.

Thus, by introducing a limited set of constraints on the properties of the time-frequency distribution ρ , we obtain a unique and intuitive definition of the Energy Decay Relief [5]:

Definition 3: The Energy Decay Relief describes the spectral energy density remaining after any time in the signal, and coincides with the *future running spectrum*:

$$EDR_h(t, f) = \left| \int_{\tau=t}^{+\infty} h(\tau) e^{-j2\pi f\tau} d\tau \right|^2 = \int_{\tau=t}^{+\infty} P_h^+(\tau, f) d\tau . \quad (18)$$

In Eq. (18), P^+ refers to a classic time-frequency distribution, known as the Page distribution [25, 26]. More specifically, the Page distribution is defined as the derivative of the (past) running spectrum, and we are referring here to the Page distribution of the time-reversed signal, also known as the Levin distribution [26]. This distribution emerges as the only solution for ρ (except for an arbitrary additive constant) if we require that ρ should preserve the frequency-domain marginal distribution, be anticausal and preserve the causality and the shifting of signals in the time domain. It turns out that property 1, Eq. (15), is not necessary to define a unique solution for the EDR, and that this property follows from (18). However, we note that requiring that ρ should also preserve the time-domain marginal distribution forces the arbitrary constant to zero, leading to a unique solution for ρ .

3 Analyzing measured impulse responses

In the next section, we establish the relation between the EDR and the stochastic time-frequency model of reverberation decays described in part 1, and we show that the EDR is a useful tool for accurately estimating the parameters of exponential reverberation decays. We then discuss the development of a practical time-frequency analysis procedure for application to measured room impulse responses.

3.1 Relation between EDR and stochastic model of reverberation

For any stochastic signal $h(t)$ modeled as a centered random noise $b(t)$ with variance σ^2 modulated by a slowly varying envelope, it is straightforward to verify that the EDC can be indifferently computed from one realization of the signal or from its time-domain envelope:

$$h(t) = \frac{b(t)}{\sigma} \sqrt{\langle h(t)^2 \rangle} \Rightarrow EDC_h(t) \approx \frac{\langle b(t)^2 \rangle}{\sigma^2} \int_t^{+\infty} \langle h(\tau)^2 \rangle d\tau = \int_t^{+\infty} \langle h(\tau)^2 \rangle d\tau . \quad (19)$$

If, additionally, we impose an exponentially decaying envelope as in (6), then applying (8) in (19) leads to:

$$EDC_h(t) \approx \frac{\sigma^2}{2\delta} e^{-2\delta t} = \frac{\langle h(t)^2 \rangle}{2\delta} = \frac{Tr}{6 \ln 10} \langle h(t)^2 \rangle . \quad (20)$$

This equation means that the EDC is approximately equal to the temporal energy envelope of the response, multiplied by a coefficient proportional to the reverberation time. Eq. (19-20) imply that a dB plot of the EDC is an appropriate representation for verifying the exponential character of the reverberation decay in a room impulse response, and that the reverberation time is best estimated by fitting a straight line to the linear part of this plot. The backward integration leading to the EDC appears as an ideal time-domain smoothing technique to derive the envelope of the echogram, when this envelope is exponential.

In order to extend Eq. (20) to the time-frequency model, we introduce the time-frequency envelope

$$ENV_h(t, f) = \langle \rho_h(t, f) \rangle, \quad (21)$$

where, as before, the choice of the time-frequency distribution ρ is *a priori* undetermined. We then impose an exponential decay at each frequency, following Polack (9):

$$ENV_h(t, f) = P(f) e^{-2\delta(f)t}. \quad (22)$$

Eq. (21-22) imply that Eq. (20) can be extended as follows:

$$\begin{aligned} \langle EDR_h(t, f) \rangle &= \int_t^{+\infty} \langle \rho_h(\tau, f) \rangle d\tau = \int_t^{+\infty} ENV_h(\tau, f) d\tau \\ &= \frac{P(f)}{2\delta(f)} e^{-2\delta(f)t} = \frac{ENV_h(t, f)}{2\delta(f)} = \frac{Tr(f)}{6 \ln 10} ENV_h(t, f). \end{aligned} \quad (23)$$

It follows [5, 6] that the time-frequency envelope and the initial spectrum $P(f)$ can be estimated by scaling the EDR with respect to the reverberation time $Tr(f)$, which can itself be estimated by linear curve fitting at each frequency on the EDR.

Introducing the ideal expression of the EDR (18), we obtain from (23):

$$ENV_h(t, f) = \frac{6 \ln 10}{Tr(f)} \langle EDR_h(t, f) \rangle = \frac{6 \ln 10}{Tr(f)} \left\langle \left| \int_{\tau=t}^{+\infty} h(\tau) e^{-j2\pi f\tau} d\tau \right|^2 \right\rangle. \quad (24)$$

We note that it is not necessary to make explicit reference to the time-frequency distribution ρ for evaluating the time-frequency envelope. However, the underlying distribution ρ in (21) must then be the Levin distribution, introduced in section 2.3, replacing the Wigner-ville distribution used in Polack's definition of the time-frequency envelope (9).

Taking $t = 0$ in (24), we obtain that the initial spectrum $P(f)$ can be directly computed from the frequency response $H(f)$, if the decay time $Tr(f)$ is known (which requires computing the EDR):

$$P(f) = ENV_h(0, f) = \frac{6 \ln 10}{Tr(f)} \langle EDR_h(0, f) \rangle = \frac{6 \ln 10}{Tr(f)} \langle |H(f)|^2 \rangle. \quad (25)$$

Random time-domain variations are eliminated in the EDR as a result of the backward integration, as discussed above in relation to Eq. (19-20). However, we note that the EDR still contains random variations along frequency. Indeed, Schroeder's statistical model (section 1.1) applies not only to the global frequency response $H(f)$, but also to the Fourier transform of the truncated response $h(t)$ introduced in section 2.3 (property 4). Consequently, $EDR(\tau, f)$ is a stochastic signal having a Rayleigh probability density function. Eq. (24-25) express that the time-frequency envelope of the reverberation decay and its initial power spectrum are derived from the frequency-domain envelope of the EDR. Although the EDR essentially decays linearly at each frequency and directly allows accurate estimates of $Tr(f)$, evaluating $P(f)$ implies computing the EDR for several realizations of the stochastic signal h and averaging the results. However, it is also possible to exploit the hypothesis that $P(f)$ and $Tr(f)$ vary slowly with frequency (section 1.3), which also applies to the time-frequency envelope and the ensemble average of the EDR, according to Eq. (22) and (23). Consequently, the time-frequency envelope and the initial spectrum can be estimated from a single realization of $h(t)$, by averaging the EDR over frequency bands in which $P(f)$ and $Tr(f)$ are assumed to be approximately constant.

According to Schroeder [17], the autocorrelation function along the frequency axis on the power frequency response reduces to 0.1 for a frequency interval of about $7/T_r$ (in Hz). This generalizes to the EDR and gives a theoretical measure of the highest frequency resolution that can be achieved when estimating $P(f)$ from a single realization of $h(t)$.

3.2 Practical time-frequency analysis procedure

In this section, we discuss the development of an analysis procedure for computing the EDR from a measured impulse response, and deriving the time-frequency envelope characterizing the late reverberation decay, or, equivalently, the initial power spectrum $P(f)$ and the reverberation time $T_r(f)$. In practice, several reasons make it necessary to modify the procedure suggested by Eq. (24-25) in the previous section:

- The exponentially decaying Gaussian stochastic model is only valid after a certain time limit in the impulse response, which depends on the room (see section 1.3).
- Direct computation of the future running spectrum, Eq. (18), for a typical room response, involves an excessive computational effort and yields an excessive frequency resolution, requiring subsequent smoothing in the frequency domain.
- A measured impulse response is usually corrupted by measurement noise. As mentioned in section 2.1, this distorts the computed EDC and, similarly, the EDR, yielding biased estimates of the decay time and of the initial spectrum.

An approximate measure of the limit time after which the stochastic model is valid is given by Eq. (12). The limit can also be estimated practically by checking for the exponential decay in the EDR. The expression (24) of the time-frequency envelope is only valid for the later decay and, consequently, $P(f)$ cannot be computed directly by applying Eq. (25). However, the theoretical early decay of the time-frequency envelope can be recovered by linear extrapolation from the later decay. We thus define the initial spectrum $P(f)$ as the spectrum read on the extrapolated time-frequency envelope at time 0 (the time of emission of the excitation pulse) [5].

For short responses, it is appropriate to compute the EDR according to the "ideal" definition (18). In practice, this implies an FFT window whose length equals the effective length of the measured response, and which is shifted in time steps equal to the desired time-domain resolution of the EDR. If a high temporal resolution is desired, an alternative method may be more efficient, consisting of computing the DFT of $h(t)$ by a time-domain recursion initiating at the end of the response, and evaluating the squared magnitude at each step [27, 5]. These methods allow computing the EDR with arbitrary resolution along the time and frequency axes (although the accuracy of the result is subject to the presence of measurement noise).

The short-time Fourier transform (spectrogram) is a practical choice for the time-frequency distribution ρ in Eq. (14), which provides an efficient method for computing the EDR of a long impulse response. This method can be implemented so as to approximately satisfy the desirable properties given in section 2.3, the main deviation from the "ideal" EDR (18) being that one obtains a smoothed and decimated version (in both the time and frequency domains). As mentioned at the end of section 3.1, such smoothing is actually desirable in the frequency domain, since the time-frequency envelope is related to the ensemble average of the EDR, according to Eq. (24-25). The frequency-domain and time-domain resolutions of the instantaneous Fourier spectra are controlled by the shape and length of the time-domain analysis window, which should also provide substantial side-lobe rejection in order to minimize the "leakage" between the energy decay curves computed in different frequency channels.

We have used a Blackmann-Harris window of order 4 to obtain a side-lobe rejection of 80 dB or more, which proved necessary particularly for analyzing synthetic impulse responses (devoid of measurement noise) [9, 24]. With a window length of 16 ms, the effective resolution is then about 250 Hz in the frequency domain and 8 ms in the time domain.

The short-time Fourier transform also makes it possible to compute an estimate of the EDR that is not corrupted by the measurement noise. The power of this noise can be estimated, in each frequency channel, in the end and the beginning of the measured impulse response. This estimate can then be used in the computation of the EDC for each frequency channel, as described earlier in section 2.1. Assuming a stationary measurement noise uncorrelated from the desired response, the estimated power spectral density of this noise can be subtracted from all instantaneous power spectra prior to cumulation (Fig. 5). Since the later part of the response also introduces a bias in the estimated noise power, it is possible to refine this estimate once initial evaluations of $P(f)$ and $Tr(f)$ have been derived from the computed EDR. Consequently [11], an iterative technique can be developed to compute more accurate estimates of the three parameters which describe the measured late decay: the reverberation time $Tr(f)$, the initial power spectrum $P(f)$ and the spectrum of the measurement noise $Pn(f)$.

Starting from a spectrogram computed with a N -point FFT and providing estimated power decays in $N/2$ equal-bandwidth frequency channels, the different steps of the iterative procedure yielding the EDC in each frequency channel are organized as follows:

- a) Compute initial estimates of Pn , Tr and P .
- b) Derive an estimate of the limit time $Tlim$, defined as the time where the ideal exponentially decaying power becomes lower than the noise power Pn .
- c) Compute a new estimate of Pn in each frequency channel (by averaging from $Tlim$ to the end of the response, after subtracting the extrapolated ideal decay).
- d) Compute the EDC after subtracting the new estimate of Pn from the power decay.
- e) Derive a new estimate of the reverberation time Tr and of the initial spectrum P , by a linear fit on the EDC in dB.

The steps (b)-(e) are iterated in order to satisfy a convergence criterion based on the quadratic error associated to the linear regression of the logarithmic representation of the energy decay.

Figures 3-6 illustrate this procedure for an impulse response measured in the Berlin Philharmonie hall at a 32 kHz sampling rate. Fig. 3 shows the spectrogram computed with a 512-point FFT and an temporal overlap of 75%. The corresponding time and frequency steps are respectively 4 ms and 62.5 Hz (although for clearer graphical presentation the results are shown here with a time resolution of 16 ms).

Fig. 4 shows the EDR computed without removing the measurement noise. It can be observed that the contribution of the noise is distorting the exponential decay, thus limiting the reliability of a reverberation time estimate. Fig. 5 shows the EDR obtained after subtraction of the estimated noise power in each frequency channel of the spectrogram. This figure also shows the limit curve $EDR(Tlim(f), f)$ where the time-frequency envelope meets the noise floor. The reverberation time can now be estimated with better confidence since the linear decay is observed over a larger dynamic range. Finally, Fig. 6 shows the EDR with ideal exponential extrapolation beyond the limit time $Tlim(f)$, derived from the estimated parameters $Tr(f)$ and $P(f)$.

4. Acoustical interpretation and results

The classical theory of diffuse reverberant sound fields considers a steady state excitation and assumes a constant reflected energy density across the room [14, 28]. In other words, ignoring the direct sound component, the energy density (proportional to squared pressure) fluctuates randomly with time and space, but its mean value (ensemble average) is assumed to be uniform throughout the room. According to Sabine's theory, the mean energy density can be approximated as follows (omitting the dependence on frequency, for clarity) [14]:

$$\langle E \rangle = \frac{4}{c} \frac{W}{A} = \frac{Tr}{6 \ln 10} \frac{W}{V}, \quad (26)$$

where W is the steady-state sound power supplied by the source, V is the volume of the room, and A denotes the total acoustic absorption (product of wall surface by average absorption coefficient).

As seen in section 1.3, a diffuse sound field can also be obtained with a pulse excitation in a mixing room, after a sufficient time has elapsed (the *mixing time*) [2, 3]. In the following sections, the stochastic model of diffuse room decays is studied in terms of the diffuse sound field generated across a room by an initial pulse excitation, and we derive a physical interpretation for the initial power spectrum introduced in section 1.2. We then illustrate the theory with results of EDR analyses carried out on impulse responses taken from a measurement campaign undertaken in European concert halls [29, 30]. The experimental results tend to confirm the validity of the proposed model when the situation is favorable to the establishment of a diffuse late reverberation decay. When applied to the stationary case, the model implies, in contrast to the classical theory and Eq. (26), that the reflected sound energy tends to decay exponentially with source-receiver distance, in accordance with Barron's revised theory [28].

4.1 General stochastic model of diffuse reverberation decays

Considering an impulsive acoustic excitation emitted at time $t = 0$ from any point in the room, we assume that at any time later than the room's mixing time, the acoustical energy is uniformly distributed across the room. The reverberation then dies out simultaneously for all receiver positions in the room according to the reverberation time $Tr(f)$, which is assumed to be independent of position. This implies that a unique time-frequency envelope characterizes the late reverberation decay across the room, irrespective of source and receiver positions or orientations.

Consequently, in a mixing room, not only the reverberation time $Tr(f)$ but also the initial spectrum $P(f)$ are independent of the source and receiver locations or orientations. The stochastic model of diffuse reverberation decays reviewed in section 1.2 can be considered to describe the "impulse response of the room" (considered as a time-varying acoustic pressure field), and not only the impulse response measured for a particular position of the source and the receiver.

The stochastic model of the room impulse response can be summarized as follows: the late reverberation pressure field is a position-dependent and time-dependent centered Gaussian stochastic process. It is stationary with respect to position and non-stationary with respect to time. The time dependence is completely characterized by the ensemble average of the time-frequency energy distribution observed for any position of the source and the receiver, which defines the time-frequency envelope of the reverberation of the room considered. This time-frequency envelope is fully characterized by two functions of frequency: the reverberation time $Tr(f)$ and the initial spectrum $P(f)$.

Going back to the stochastic model of an individual impulse response (section 1.2), we are now allowed to assume that different realizations of this stochastic process are obtained by varying the position of the receiver with a fixed source position or by varying the position of the source with a fixed receiver position (or of course by varying both positions). We note that the same stochastic process will be observed, irrespective of position, provided that the time origin be defined with reference to the signal emitted by the source (not as the arrival time of the direct sound at the receiver). This implies that we can assume ergodicity and evaluate the ensemble average of the EDR in Eq. (24-25) by spatial averaging.

4.2 Physical interpretation of the initial power spectrum

To derive a physical interpretation of the initial spectrum, we again consider a sound source located at any position in the room and emitting an impulsive excitation at time $t = 0$, radiating the total energy $W(f)$. If the source is a loudspeaker, we will assume that this acoustical energy impulse was produced by feeding an electrical unit pulse to the loudspeaker.

Let us first consider the ideal case where there is no energy absorption in the room, which implies an infinite reverberation time. When sufficient time has elapsed, the pressure signal at any receiving point becomes a stationary Gaussian random process, and so is the signal $h(t)$ recorded with a microphone. The signal $h(t)$ is then characterized by its power spectral density $P(f)$, as in section 1.2. $P(f)$ is related to the mean energy density $\langle \omega(f) \rangle$ via the diffuse-field sensitivity of the receiver, which we denote $R_d(f)$ and define as follows:

$$P(f) = \rho_0 c^2 R_d(f)^2 \langle \omega(f) \rangle, \quad (27)$$

where ρ_0 denotes the density of the air. Since the mean energy density $\langle \omega(f) \rangle$ is uniform across the room and the total energy in the room must be equal to the energy $W(f)$ initially supplied by the source, the mean energy observed at any point must be inversely proportional to the room volume:

$$\langle \omega(f) \rangle = \frac{W(f)}{V}. \quad (28)$$

With a finite reverberation time, we assume that the *total energy* in the room decays exponentially from $t = 0$ and is initially equal to the energy supplied by the impulsive source excitation. Note that we do not assume that the energy density $\langle \omega(f) \rangle$ at any particular receiver location decays exponentially during the early response. However, we do make this assumption for the later decay, along with the assumption that the energy density becomes uniform across the room. Consequently, we can write, irrespective of source or receiver position:

$$\langle \omega(t, f) \rangle = \frac{W(f)}{V} e^{-2\delta(f)t} \quad \text{for } t > t_{\text{mixing}}. \quad (29)$$

Finally, in order to express the source power in terms of the amplitude of the electrical input signal (which is here assumed to be a Dirac pulse), we introduce the diffuse-field transfer function of the loudspeaker, which we denote $S_d(f)$ and define as follows:

$$W(f) = \frac{S_d(f)^2}{\rho_0 c^2}. \quad (30)$$

Combining Eq. (27-30), we obtain the following expressions of the initial power spectrum and the time-frequency envelope:

$$P(f) = \frac{1}{V} S_d(f)^2 R_d(f)^2, \quad \text{ENV}(t, f) = \frac{e^{-2\delta(f)t}}{V} S_d(f)^2 R_d(f)^2. \quad (31)$$

This physical interpretation of the stochastic model of reverberation decays offers a separation of room characteristics and transducer characteristics: the temporal decay rate characterizes of the room, while the initial spectrum characterizes the source-receiver pair (irrespective of position). The initial spectrum can be factored as the product of the diffuse-field sensitivity of the receiver and the diffuse-field power spectrum of the source, as conjectured in [5, 10], with provision for a frequency-independent scaling coefficient equal to the room volume.

Note that, in Eq. (27) and (30), we have defined the functions $S_d(f)$ and $R_d(f)$ characterizing the transducers in order to relate, in both cases, a pressure signal to an electrical signal. These two functions can then be called the "diffuse-field transfer functions" of the transducers, and this choice allows exhibiting the symmetry (and reciprocity) of the contributions of the two transducers in the model, as illustrated by Eq. (31). More generally, if the source is not a loudspeaker, but any acoustical source (not necessarily having an electrical driving input), $P(f)$ can be expressed in terms of the source power by applying (28) in (27). The term $\rho_0 c^2$ is then conveniently eliminated if we define the diffuse-field sensitivity of the receiver $R_d(f)$ in terms of energy density rather than acoustic pressure (and the symmetry of the equation is also restored):

$$P(f) = \frac{\rho_0 c^2}{V} W(f) R_d(f)^2 = \frac{1}{V} W(f) R_d(f). \quad (32)$$

To conclude this section, we return to the reflected sound level and its variation with listening position. Assuming a uniform exponential decay of the acoustical energy in the room, as we do in this physical model, implies that the reflected energy tends to decay exponentially with source-receiver distance, in contrast to the classical theory and Eq. 26. This was noted by Barron [28], and follows from the fact that the reflected energy at a given location is defined as the total energy in the impulse response after the direct sound, and the arrival time of the direct sound (referenced to the time of emission of the impulse) is proportional to source-receiver distance. An expression of the reflected energy can be derived in the framework of Barron's simplified model of the impulse response [28], where reflections are assumed to follow an exponentially decaying envelope starting immediately after the direct sound: the mean reflected energy density is then evaluated by integrating (29) from $t = r/c$ to $t = +\infty$ (denoting r the source-receiver distance). Omitting the dependence on frequency as in (26), we obtain:

$$\langle E \rangle = \frac{1}{2\delta} \frac{W}{V} e^{-2\delta r/c} = \frac{Tr}{6 \ln 10} \frac{W}{V} e^{(-6 \ln 10 r)/(cTr)}. \quad (33)$$

This expression of the reflected energy coincides with the classical theory (26) when $r = 0$, and decays exponentially with distance according to the reverberation time. Eq. (33) agrees with Barron's revised theory, the validity of which was illustrated in [28] by studying the variation of a set of acoustical indexes (evaluated for midfrequencies) vs. source-receiver distance, based on extensive measurements undertaken in different British concert or conference halls. In sections 4.3 and 4.4 below, we present results intending more specifically to illustrate the validity of the stochastic model of late diffuse reverberation decays discussed in this section.

4.3 Influence of source and receiver positions in a given room

The results below are based on measurements that were undertaken in various European concert halls and opera houses. All measurements consist of impulse responses derived from MLS sequence excitation of the same source (Yamaha NS1000 loudspeaker) and recorded with the same microphone (Neuman U87) offering a choice of three spatial sensitivities: omnidirectional, cardioid and bi-directional. The results presented are curves of reverberation time vs. frequency $Tr(f)$ and initial power spectra $P(f)$, obtained by the EDR analysis procedure described in part 3 of this paper.

Fig. 7 displays the estimated reverberation time for 6 different positions of the receiver in the Berlin Philharmonie hall. Up to 10 kHz, the different estimates show little variance with respect to frequency and receiver position. Above 10 kHz, the very short reverberation time causes an increased uncertainty in the estimation of the parameters characterizing the time frequency envelope (see section 3.2). However, by averaging over the different receiver locations, an accurate estimate of the reverberation time with high frequency resolution can be obtained. Fig. 8 shows the initial spectrum $P(f)$ computed separately for different receiver locations, and an average estimate. It can be verified that $P(f)$ is independent of receiver location except at the extremes of the frequency range. At high frequencies, the variance on $P(f)$ may be explained by the variance observed for the estimated reverberation time on Fig. 7, since $P(f)$ is derived according to Eq. (25). However, at low frequencies, the reverberation time estimate appears to be more robust, which then suggests that a deviation occurs from the assumptions of the statistical reverberation model. For instance, if the source and the receiver are located in the median plane of the hall, the measurement will reflect the effects of constructive interferences between reflections, which can produce an increase in the estimated $P(f)$, particularly for longer wavelengths (i. e. low frequencies).

Another example of specific situations where a dependence of $P(f)$ on transducer location can be noted is given by Fig. 9. In this example, we compare an estimate of $P(f)$ averaged over different positions of the receiver in the main volume of the Salle Pleyel with an estimate for a receiver located under the first balcony and averaged over different source positions on the stage. For that particular receiver position, $P(f)$ drops by approximately 4 dB across the whole frequency range, as a consequence of the uneven distribution of the reverberation field in the room. Except for such particular situations, the results agree with the assumption that the late reverberation process admits the same time-frequency envelope for all positions in the room, provided that the time origin be defined as the time of emission of the sound by the source.

4.4 Influence of the transducers and of the room

As shown in section 4.2, the initial power spectrum can be interpreted as the product of the diffuse-field transfer functions (DFTFs) of the source and the receiver. This is consistent with the assumption that the late reverberation field is fed uniformly by all directions of radiation from the source and observed through the DFTF of the receiver. Eq. (31) implies that the DFTF of a loudspeaker can be estimated from an impulse response measured with a diffuse-field equalized microphone, although an absolute measurement involves the knowledge of the room volume. However, the ratio of the initial spectra associated to two different sources but the same receiver should not depend on the room, and should be equal to the ratio of the DFTFs of the two sources. This applies in a similar manner to the comparison of two receivers using the same source. As an illustration of the influence of transducer directivity, Fig. 10 shows the initial spectrum estimated by averaging over several receiver positions, for two spatial sensitivities of the microphone (omnidirectional and cardioid). Since the nominal axis response is nearly identical for the two spatial sensitivities in this case, this comparison can be interpreted as the difference in directivity index vs. frequency between these two microphone sensitivities.

According to Eq. (31), when using the same source-microphone pair to undertake measurements in different halls, the respective initial power spectra will show a frequency-independent difference corresponding to the inverse ratio of room volumes. This behavior is illustrated in Fig. 11, which shows $P(f)$ and $Tr(f)$ estimated in two different halls (Berlin Philharmonie and Amsterdam Concertgebouw). The estimated $P(f)$ curves remain essentially parallel over almost the whole frequency range, which is consistent with the assumption of a ratio equal to the inverse of the volume ratio.

5 Applications

5.1 Perceptual characterization of room acoustical quality

In various application fields involving room acoustics, it is necessary to use a visual representation allowing perceptually relevant evaluation and comparison of measured or synthetic room reverberation responses. This naturally calls for some form of time-frequency representation, since representations of the time response $h(t)$ or of the frequency response $H(f)$ give little information, for instance, on the variation of the decay time vs. frequency.

Room acoustical quality is usually represented by a set of graphs, each showing a particular acoustical index evaluated in octave or third-octave bands. Typical monaural indexes are the reverberation time, the early decay time, the clarity index, the center time, the direct to reverberant ratio, etc... (see e. g. [31]). This type of representation entails the following limitations [5]:

- It implies the selection of a limited number of perceptual attributes, and strictly specified conventions on the computation of the related indexes (such as the shape of time-domain windows, for early to late energy ratios such as the clarity index).
- The set of indexes is generally defined with consideration of the usage of the room: the most widely used acoustical indexes were designed for characterizing the perception of speech or music in relatively large enclosures.
- This method does not provide a global visual representation, but, rather, one that is divided in several separate graphs.
- As mentioned already in section 2.1, the frequency resolution is limited and the calculated values depend on conventions on the design of the filter bank.

The evaluation of a given set of acoustical indexes may yield results which differ from one measurement system to another, and is not always sufficient for distinguishing two audibly discernible acoustical qualities [32]. The lack of exhaustivity of such representations has motivated research on perceptually relevant time-frequency representations of room acoustical quality, and towards the definition of new acoustical indexes derived from such representations [33]. Griesinger has used spectrograms evaluated with a 40-ms time window and third-octave averaging, from which time-frequency plots, acoustical indexes and integrated EDCs could be derived [34].

Because the Energy Decay Curve (EDC), as defined in section 2.1 by Eq. (13), displays integrated energy values, the monaural acoustical indexes widely used in room acoustics are better represented visually by the EDC than by the echogram or any locally smoothed version of it [35]. Consequently [5, 6], a 3-D plot of the EDR (Fig. 6) offers an exhaustive, perceptually relevant representation of the (monaural) acoustical quality of a room. The variation of the classical room acoustical indexes vs. frequency (reverberation time, clarity index, etc.) can be derived from the EDR, with arbitrary resolution in the frequency domain. Furthermore, the EDR allows recovering the EDC of any filtered version of the impulse response $h(t)$. Assuming a linear-phase filter and denoting $w(t)$ and $W(f)$, respectively, its time and frequency responses, we have (ignoring the frequency-independent delay introduced by the filter) [5]:

$$h'(t) = w^*h(t) \quad \Rightarrow \quad EDR_h(t, f) \approx |W(f)|^2 EDR_h(t, f) \quad (34)$$

$$\Rightarrow \quad EDC_{h'}(t) = \int_{-\infty}^{+\infty} |W(f)|^2 EDR_h(t, f) df \quad (35)$$

Eq. (35) follows from Eq. (15), and means that the EDC in any frequency band can be computed by windowing and then integrating the EDR along frequency. Eq. (34) is only approximately exact if the EDR is defined according to (18) because the operations of filtering the signal and zeroing it up to a given time are not commutative. However, the deviation is negligible if the impulse response $w(t)$ of the filter is short compared to $h(t)$. A generalization of Eq. (34-35) exists for filters having non linear phase responses [5]. Finally, we note that Eq. (35) implies that it is licit to smooth and decimate the EDR with a frequency-domain averaging window to reduce the frequency resolution of the 3-D plot, when necessary, and that this is also a method of approximating the ensemble average of the EDR, which is meaningful for extracting the parameters of the late decay, as mentioned earlier (section 3.1).

Although the EDR only describes the monaural attributes of room acoustical quality, the approach can be extended to time-frequency representations of binaural attributes. One possibility is to replace the instantaneous spectrum $\rho(t, f)$ by the normalized instantaneous cross-spectrum of the left and right signals in a binaural impulse response, which yields a time-frequency representation of the Interaural Cross-Correlation Coefficient (IACC). A similar idea [36] is the use of time-frequency representations of the binaural sum and difference spectra.

5.2 Diffuse-field equalization of transducers

As mentioned in sections 4.2 and 4.4, the EDR analysis procedure allows estimating the diffuse-field frequency response of the transducers from impulse responses measured in a room. An additional measurement taken in free-field or with the source close to the receiver then allows deriving the directivity index. The diffuse-field transfer function (DFTF) of a loudspeaker, as defined by Eq. (30), is related to the power transmitted to the diffuse reverberant field. The directivity index is related to the ratio of direct to reverberant sound, and is therefore a key parameter in the design of sound systems [37, 38]. The DFTF also has application in recording and reproduction techniques (particularly binaural techniques): equalizing recordings with reference to the DFTF of the microphone (or the ear) offers a number of practical advantages [39-44].

The DFTF of a transducer is typically measured by one of the following two approaches, both of which involve extreme experimental conditions, usually available only in laboratory environments:

- Steady-state recording in a reverberant chamber: the DFTF is evaluated by computing the ratio of the power spectra recorded with the test transducer and a reference transducer [40].
- Impulse response measurements in an anechoic chamber: the DFTF is evaluated by power-averaging of the free-field frequency responses for a uniform distribution of directions around the transducer [42-44].

The general model of diffuse reverberation decays described in section 4.2 suggests a new technique for measuring the DFTF of a transducer in any room producing a sufficient amount of diffuse reverberation. As mentioned in section 4.4, for two impulse responses measured with different transducers in a given room, the ratio of the initial spectra is equal to the ratio of the DFTFs of the two transducers. More generally, the ratio of the two DFTFs is also obtained by dividing the two time-frequency envelopes (or the two EDRs) at any time in the late decay (referenced to the time of emission of the excitation pulse). A simple practical procedure consists in computing the spectrogram of each impulse response, estimating and subtracting the noise spectrum if necessary (as described in section 3.2), and finally summing the instantaneous power spectra to evaluate the spectrum of the reverberation after a chosen limit time (i. e. the spectrum read on the EDR for that particular time).

This approach can be considered as a variation of the conventional steady-state method (a) above. However, the new method allows in situ measurements in any typical room, because it is based on analyzing an impulse response to extract the parameters characterizing the diffuse reverberation field. If the two measurements are made at the same position, the only constraint for general validity of the measurement is that a locally diffuse field be achieved during the late reverberation decay at that position.

As mentioned above, this method can be applied to binaural techniques and headphone equalization. It is currently being used to measure the diffuse-field head-related transfer functions of human subjects, the results being compared to the traditional anechoic chamber and reverberant chamber methods [43, 44]. We note that, in a situation where a diffuse reverberant field is not realized, the technique can still be used for characterizing the coupling of a transducer to the reverberant field for that particular situation, although no general conclusions can be drawn regarding the diffuse-field frequency response of the transducer.

5.3 Restoration of measured impulse responses for auralization

When a measured impulse response is used for convolution with anechoic signals, the measurement noise at the end of the response can manifest itself as a low-level non decaying reverberation effect prolonging the natural reverberation. Furthermore, when the goal is to simulate a natural acoustic source rather than a loudspeaker playing a recorded signal, the effects of the directivity characteristics of the loudspeaker should be corrected, particularly as directivity affects the ratio of direct to reverberant sound vs. frequency [37, 38].

Eliminating the noise by windowing the measured response at the limit time where signal power becomes lower than noise power may not be adequate in all cases, since it typically has the effect of cutting off the late reverberation decay at low frequencies, while leaving a substantial portion of noise at high frequencies (where the response generally decays faster). One possibility consists of estimating the noise spectrum and using a spectral subtraction method, based on the short-time Fourier transform and overlap-add synthesis, to restore the measured response [45, 46]. This technique theoretically has the effect of windowing the response in the time-frequency domain (i. e. with a frequency-dependent time limit). However, it cannot restore the end of the decay, and may leave some audible residual artifacts instead.

In the case of room impulse responses, a perceptually relevant model of the late decay can be synthesized from estimates of the initial spectrum $P(f)$ and the reverberation time $Tr(f)$ (preferably with a high frequency resolution, which is possible with the EDR analysis procedure described in part 3). The resulting time-frequency envelope defines a time-dependent filter which transforms a centered and normalized stationary Gaussian white noise into a synthetic late reverberation decay which can be substituted to the part of the impulse response that is corrupted by measurement noise. As shown on Fig. 12, the substitution is carried out in the time-frequency domain and the restored time response is obtained by inverse Fourier transform and overlap-add synthesis [9, 10]. The time-domain windows for Fourier analysis and synthesis must be chosen in order to ensure exact reconstruction of the non-corrupted part of the time response [47]. We have used a rectangular window for Fourier analysis and a raised cosine (Hann) window for overlap-add synthesis, with 50% temporal overlap [9]. The cross-fading effect of the Hann window eliminates signal discontinuities at window edges caused by time aliasing (which results from windowing the noise spectra in the frequency domain). Time aliasing can be otherwise ignored here since the phase information contained in the original random noise signal is not important. Fig. 13 shows an example result of the restoration procedure. A similar technique was developed by Polack & al. [12] to restore impulse responses measured in small-scale models.

Simulating a change of source and/or receiver directivity in a measured impulse response can be achieved by windowing and spectral correction of the direct sound and the later reverberation, separately. The spectral correction on the direct sound can be derived from free-field measurements, while the correction on the reverberation is derived from the ratio of diffuse-field transfer functions (see sections 4.4 and 5.2). When the diffuse-field transfer function of the source cannot be defined (such as with a natural acoustic source or instrument), its power $W(f)$ can be evaluated from a steady-state recording in the reverberant field, according to Eq. (33). Theoretically, early reflections should be corrected individually (which would imply identifying the associated acoustical paths and directions of emission and incidence, as well as the corresponding free-field transfer functions of the sources and the receivers). However, this may be unnecessary in many situations, where one can assume that early reflections are not perceived individually and span several directions of emission and incidence [42].

These spectral corrections can be implemented in the frequency domain, in the left branch of the diagram shown on Fig. 12, with proper zero-padding and choice of time-domain windows [47] since useful phase information is present especially in the beginning of the reverberation decay. This approach can be extended to implement a time-varying filter simulating a modification of the reverberation time (this technique was used in [12] to simulate a modification of air absorption characteristics, thus allowing auralization based on impulse responses measured in scale models).

5.4 Natural-sounding artificial reverberation using feedback delay networks

Convolution by long impulse responses can be implemented in real time with no input-output delay and affordable computational resources [4, 48]. Room impulse responses can be measured in a real room, or derived from a computer model of a virtual room [49], or synthesized according to the stochastic model described in part 1 (as shown on Fig. 1 or Fig. 12). Convolution techniques are particularly well suited for comparative or predictive auralization of concert halls, auditoria or sound systems, for evaluation purposes. This application typically involves headphone reproduction or a controlled listening environment, which allows taking full advantage of the accuracy of the convolution technique.

Applications of digital reverberation techniques also include professional audio production, sound systems, computer music, multimedia or virtual auditory displays. In these contexts, the essential design criteria, above the ability to accurately reproduce an existing room effect, are the naturalness of the reverberation effect, the efficiency of the control interface for dynamic control of the synthetic reverberation effect, and the possibility to adapt the signal processing system to different reproduction formats and listening conditions. Convolution techniques rely, by definition, on a cumbersome low-level representation (the impulse response), which implies heavy storage requirements and/or complex updating schemes for manipulating the synthetic reverberation. A more efficient parametrization is offered by artificial reverberation algorithms based on digital delay networks, combining feedforward paths to render early reflections and feedback paths to synthesize the later reverberation [15, 19, 50]. Multichannel feedback delay networks (FDNs) provide a general approach for implementing natural sounding, computationally efficient and scalable reverberation algorithms [51, 5, 7].

Physically exhaustive reproduction of a measured or computed impulse response is not a necessary condition for ensuring the naturalness of an artificial reverberation effect. Indeed, the use of a feedback delay network for rendering the late reverberation decay is justified fundamentally by the stochastic model reviewed in part 1 of this paper: a FDN reverberator can be designed so that its late impulse response is physically undistinguishable from an exponentially decaying Gaussian random process, and that both the decay time $Tr(f)$ and the initial spectrum $P(f)$ can be tuned with arbitrary accuracy [5-7].

This excludes features deviating from the stochastic late reverberation model, such as non-exponential decays or flutter echoes. However, we note that this limitation is common in current auralization software based on physical computer models. Generally speaking, in any auralization system where the late reverberation decay is artificially synthesized on the basis of computed reverberation times and reverberation levels, feedback delay networks can offer a more efficient and flexible approach than convolution techniques.

A general reverberator design method is described in [51, 5-7], based on energy-preserving multiple feedback delay networks. Such networks can produce several uncorrelated non-decaying reverberation responses in which the reflection density increases, converging towards a Gaussian amplitude distribution. The time-frequency envelope is controlled by associating a spectral corrector in cascade with the network and an absorptive filter with each delay unit in the feedback loop. The average number of modes per Hz is equal to the total duration of the delay units in the feedback loop, which should therefore be at least equal to the reverberation time in order to ensure a modal overlap of 3:1, and thus satisfy the assumptions of Schroeder's frequency-domain model (see section 1.1). In practice, following Schroeder's suggestion [15], it is generally sufficient to match the average number of maxima per Hz, which only implies a total delay length of $Tr/4$ according to Eq. (4). The computational cost is essentially proportional to the number N of delay units, which is eventually determined by the requirement of ensuring both a sufficient total delay length and a fast increase of the echo density in the beginning of the impulse response.

5.5 Simulating real rooms with an efficient tunable artificial reverberator

As illustrated by the structure of the *Room* module shown on Fig. 14, the basic FDN (denoted *reverb*) can be extended to allow separate control of early reflections and later reverberation [5, 7]. On this example, the *early* module is a delay line producing several delayed copies of the mono input signal, which are used both to render the early reflections R_1 and feed the subsequent stages of the reverberator. Optionnally [7, 42], a second stage (*cluster*), comprising an energy preserving mixing matrix and a bank of delay lines, can be included to provide control over an intermediate and denser group of reflections R_2 .

These elementary reverberation modules (*early*, *cluster*, *reverb*) have been implemented in the IRCAM *Spat* software, and can be interconnected in different ways to build scalable reverberation algorithms or mixing architectures configurable for various contexts of application, ranging from professional audio and sound systems to multimedia and virtual reality [5, 7, 52, 53]. As discussed in [42], the *Room* module can be directly used as a stereo or binaural reverberator providing individual adjustment of each early reflection (in time, amplitude and direction of incidence). Alternately [7], as shown on Fig. 14, it can produce a 3/2-stereo or 3/4-stereo compatible output, which can then feed a separate module (*Pan*) for directional distribution of the direct sound and the early reflections, and diffuse rendering of the late reverberation decay (this *Pan* module is configurable to allow rendering in various formats and setups, using binaural, Ambisonic or pairwise intensity panning techniques [7, 52, 53]).

The total signal processing cost of the *Room* module of Fig. 14 is equal to $2N(\log_2 N + 3) + 48$ numerical operations (multiply-adds) per sample period, where N is the number of delay channels across the three stages of the reverberator [7]. In high fidelity applications, a suitable value for N is 8, i. e. 144 operations per sample (i. e. 7 MIPS at a sample rate of 48 kHz). This includes the calculation of 15 second-order (biquadratic) parametric shelving filters allowing spectral correction of the reverberation time and the energy of the reflections R_1 , R_2 , R_3 , independently (in three frequency bands, with adjustable crossover frequencies). To allow continuous adjustment of these parameters, the additional expense is 5 operations per filter, which implies a maximum theoretical cost of about 210 numerical operations per sample (10 MIPS at 48 kHz).

The *Room* algorithm also allows adjustment of the individual times of the early reflections and of the overlap between the sections R_1 , R_2 and R_3 in the synthetic reverberation (with the constraint that section R_n must start later than section R_{n-1} and be at least as long).

Based on the EDR analysis procedure described in part 3, the parameters of the reverberation algorithm can be optimized from measured or computed impulse responses [6, 9, 54]. One approach, described in [6], consists of directly optimizing the coefficients of the elementary absorptive filters and spectral correctors of the FDN reverberator, to derive an accurate and computationally efficient IIR model of the late reverberation decay of an existing room. A second approach, adopted in [54] and [9], consists of deriving optimal settings of all control interface parameters, i. e., in the case of the *Room* module described above, the reverberation time and the levels of the different temporal sections of the generic reverberation model, at low, medium and high frequencies. For a given reverberator implementation, this second approach is less accurate than directly optimizing filter coefficients, but provides an initial setting (or "preset") of interface parameters allowing subsequent continuous modifications by the user. In [54], formal perceptual tests were carried out over headphones, in which convolution with binaural impulse responses was compared with a binaural implementation of the *Room* module [42], using this analysis / synthesis technique. This study, in which the simulated room was a small teleconference studio, showed that the reverberator was able to reproduce the real reverberation very convincingly.

5.6 Physical room models and distance rendering in virtual auditory displays

Spatial sound processors for computer music or virtual reality require dynamic control of panning effects and artificial reverberation parameters. The control interface should provide a higher-level description of the sound scene, in terms of the positions of the sources and the listener and the acoustical characteristics of the room.

Chowning [55] introduced the principle of a perceptually-based control interface for independent control of the angular localization and the perceived distance of the source. Angular panning was applied to the direct sound and a fraction of the reverberation, while the distance cue involved simultaneous attenuation of the power of the direct sound by a factor $1/r^2$ and of the reflected sound by a factor $1/r$. Moore [56] proposed a physically exhaustive approach using more sophisticated reverberation algorithms, providing individual control of each early reflection in time, intensity and direction, according to the geometry and physical characteristics of room boundaries, the position and directivity pattern of the source, and the listening setup.

More recently [52, 53], a perceptually exhaustive approach was developed, in which the artificial reverberation effect is exclusively controlled via a set of mutually independent perceptual attributes, derived from earlier research carried in the IRCAM Room Acoustics group [57, 58]. This perceptual approach is more efficient than an exhaustive physical approach, from a computational point of view, and has also proven to be effective in the context of musical production or composition [53]. Experience has also confirmed that it remains useful, as proposed by Chowning [55], to provide a geometrical representation of the sound scene, typically as a 2-D graphical interface simultaneously allowing to manipulate the angular panning and the distance of each sound source relative to the listener. Generally speaking, in this context, the difficulty is to maintain a realistic reverberation effect and a distinct localization of the source for large distances (particularly with 3-D sound techniques allowing a surrounding reproduction of the reverberation, over headphones or loudspeakers). The general model of reverberation decays described in section 4.2 can be regarded as the most complete physical model that can be implemented without introducing an exhaustive description of room geometry and wall materials. It also suggests, as shown below, a physically-based extension of Chowning's approach.

In the following, t will denote absolute time in the impulse response (relative to the time of emission of the excitation pulse), as before, while τ will denote the arrival time of a reflection relative to the arrival of the direct sound at the listener:

$$\tau = t - \frac{r}{c}. \quad (36)$$

The level of the direct sound, expressed in terms of squared pressure, can be written as follows:

$$E_{dir}(f) = S_\phi(f)^2 \frac{(\mu(f))^r}{4\pi c r^2}, \quad (37)$$

where $\mu(f)$ denotes the intensity absorption coefficient vs. frequency for 1 meter propagation in the air, and $S_\phi(f)$ denotes the free-field transfer function of the source in the direction ϕ of the receiver. Similarly, according to Eq. (31) or (33), the level of the reverberation after time τ in the impulse response can be expressed as follows:

$$E_{rev}(\tau, f) = EDR(t, f) = \frac{Tr(f)}{13.81} \frac{S_d(f)^2}{V} \exp(-13.81 (\tau+r/c)/Tr(f)). \quad (38)$$

where we note, for clarity, $\exp(x) = e^x$. Eq. (38) is equivalent to Eq. (33) and describes the exponential decay of reflected sound level vs. distance mentioned before in section 4.2, which differs from the $1/r$ law used by Chowning in [55]. It should be noted, particularly, that, since the reverberation time is usually shorter at high frequencies (as a result of the combined effects of air absorption and wall absorption), Eq. (38) implies that the reverberation should be increasingly low-pass filtered with distance, compared to the direct sound.

In this model, the complete description of a sound scene is given by the following control interface parameters:

- the position of each source relative to the listener; its distance r and orientation ϕ (as well as its angular localization, which controls the directional panning of the direct sound);
- the radiation characteristics of each sound source, represented by its directivity $S_\phi(f)$ and its diffuse-field transfer function $S_d(f)$,
- the room, represented by its volume V and its reverberation time $Tr(f)$ (which can be expressed in terms of the absorption coefficient of the air and walls and the surface of the walls, according to classical formulas of acoustics such as the Sabine or Eyring formulas [13, 14]).

This general model can be refined by providing separate control over a number of individual early reflections, as in the *Room* module described in section 5.5 and Fig. 14. The level of an isolated reflection, still expressed in terms of squared pressure, can be derived from the expression (31) of the time-frequency envelope of the later reverberation decay and the expression (1) of the echo density:

$$E_{refl}(\tau, f) = \frac{ENV(t, f)}{D_e(t)} = S_d(f)^2 \frac{\exp(-13.81 (\tau+r/c)/Tr(f))}{4\pi c^3 (\tau+r/c)^2}. \quad (39)$$

Eq. (39) is similar to Eq. (37), except that r/c is replaced by $\tau+r/c$ and that the air absorption is replaced by the combined absorption of the air and walls. This corresponds, as suggested earlier by Gerzon [59], to the mean attenuation vs. distance undergone by a sound wave emanating from an image source, in the framework of the geometric theory of room acoustics [13, 14]. Accordingly, the free-field transfer function $S_\phi(f)$ is also replaced by the diffuse-field transfer function $S_d(f)$. We note that, while each individual reflection undergoes the geometric decay in $1/r^2$, the echo density increases in t^2 to compensate for this effect, so that the envelope decay of the reverberation is exponential, with a rate only determined by physical absorption phenomena.

These equations (37-39) can be used to control source receiver-distance with the *Room* algorithm described in section 5.4 and Fig. 14, and, more generally, in any application involving an artificial reverberator to create and control distance effects, including the Chowning and Moore models [55, 56]. In this statistical approach, the room is fully parametrized by its volume V and its reverberation time $Tr(f)$. The only parameters which remain undetermined relate to the fine temporal structure of the impulse response, i. e. the arrival times of the early reflections and the later reverberation relative to the direct sound (denoted τ in Eq. (38-39) and τ_i in Fig. 14). Depending on the application, these temporal parameters can be fixed according to the size of the virtual room or variable with the source-receiver distance r (for instance via a simplified image source model).

Conclusion

We have reviewed a statistical time-frequency model of reverberation decays, derived from earlier work by Schroeder and Polack, in which the later part of a room impulse response is modelled as an exponentially decaying Gaussian random noise. We have shown that this model can be considered to describe a diffuse reverberation decay, over the whole room, by two functions of frequency: the reverberation time $Tr(f)$ and an initial power spectrum, denoted $P(f)$. The physical interpretation of this time-frequency model has been studied theoretically and experimentally, confirming that the initial spectrum is proportional to the product of the diffuse-field transfer functions of the source and the receiver, and inversely proportional to the room volume.

A time-frequency representation particularly relevant to room responses and the characterization of room acoustical quality, the Energy Decay Relief (EDR), has been introduced. The development of a practical computational procedure for EDR analysis of measured impulse responses has been discussed. The relation between the EDR and the stochastic reverberation model has been outlined, showing that the EDR allows accurate estimates of $Tr(f)$ and $P(f)$, with high frequency resolution. This results in a new method for *in situ* measurement of the diffuse-field frequency response of a transducer, based on impulse response measurements in typical rooms (and therefore not requiring reverberant or anechoic chambers).

The relevance of the stochastic time-frequency model to auralization, artificial reverberation and virtual acoustic displays has been outlined, leading in particular to design criteria for natural sounding artificial reverberation algorithms and efficient physically-based room models and distance effects. An analysis/synthesis approach to artificial reverberation has been described, based on the EDR analysis procedure and the stochastic time-frequency reverberation model. Future extensions of this work include objective and perceptual tests aiming at validating the following applications: the simulation of concert hall reverberation using feedback delay networks and the proposed analysis/synthesis procedure, the measurement of diffuse-field head-related transfer functions for application to the equalization of headphones and binaural recordings, or the *in-situ* measurement of the directivity index of a transducer.

Acknowledgments

The authors wish to thank Olivier Cappé for extensive proofreading of an early version of section 2.3, as well as Patrick Flandrin and Yves Grenier for introduction to time-frequency distributions and bibliographical references. The support of Studer Digitec during the early stages of this work at Telecom Paris, and, since then, of Espaces Nouveaux and France Telecom (CNET), is gratefully acknowledged. Applications described in sections 5.5 and 5.6 are covered by issued and pending patents assigned to France Telecom.

References

- [1] Schroeder, M. R. (1954). "Eigenfrequenzstatistik und anregungsstatistik in Räumen", *Acustica* 4: 456-468. English translation in "Statistical parameters of the frequency response curves of large rooms". *J. Audio Eng. Soc.* 35: 299-306 (1987).
- [2] Polack, J.-D. (1988). "La transmission de l'énergie sonore dans les salles". Thèse de Doctorat d'Etat, Université du Maine, Le Mans.
- [3] Polack, J.-D. (1993). "Playing billiards in the concert hall: the mathematical foundations of geometrical room acoustics". *Applied Acoustics* (special issue on Computer Modelling and Auralisation of Sound Fields in Rooms) 38: 235-244.
- [4] Gardner, W. G. (1995). "Efficient convolution without input-output delay". *J. Audio Eng. Soc.* 43(3).
- [5] Jot, J.-M. (1992). "Etude et réalisation d'un spatialisateur de sons par modèles physiques et perceptifs". Doctoral dissertation, Télécom Paris.
- [6] Jot, J.-M. (1992). "An analysis/synthesis approach to real-time artificial reverberation". *Proc. 1992 IEEE Int. Conf. Acou. Speech and Signal Proc.* II: 221-224.
- [7] Jot, J.-M. (1997). "Efficient models for reverberation and distance rendering in computer music and virtual audio reality". *Proc. 1997 International Computer Music Conference*.
- [8] Schroeder, M.R. (1965). "New method for measuring reverberation time", *J. Acoust. Soc. Am.* 37: 409-412.
- [9] Vandernoot, G. (1995). "Analyse temps-fréquence de réponses impulsionales de salles. Application à la simulation sonore sur écouteurs". IRCAM tech. report. Master's thesis, Télécom Paris.
- [10] Jot, J.-M., Vandernoot, G. and Warusfel, O. (1996). "Analysis and synthesis of room reverberation in the time and frequency domains - application to the restoration of room impulse responses corrupted by measurement noise". Presented at the 131st Meeting of the Acoustical Society of America (paper 4aAA5). *J. Acou. Soc. Am. (Abstracts)* 99(4).
- [11] Cerveau, L., Jot, J.-M. and Warusfel, O. (1996). "Auralization of room acoustic quality". Presented at the 132nd Meeting of the Acoustical Society of America (paper 2aAA3).
- [12] Polack, J.-D. and Grillon, V. (1993). "Auralization in scale models: processing of impulse response". *J. Audio Eng. Soc.* 41(11).
- [13] Kuttruff, H. (1979). *Room Acoustics*, 2nd edition. Applied Science Publishers Ltd, London, 1979.
- [14] Cremer, L. and Müller, H. (1982). *Principles and Applications of Room Acoustics*, Applied Science Publishers Ltd.
- [15] Schroeder, M. R. (1962). "Natural-sounding artificial reverberation". *J. Audio Eng. Soc.* 10(3).
- [16] Schroeder, M. R., and Kuttruff, H. (1962). "On frequency response curves in rooms. Comparison of experimental, theoretical and Monte-Carlo results for the average frequency spacing between maxima", *J. Acoust. Soc. Am.* 34: 76.
- [17] Schroeder, M. R. (1962). "frequency-correlation functions of frequency responses in rooms", *J. Acoust. Soc. Am.* 34(12): 1819-1823.
- [18] Ebeling, K. J., Freudenstein, K., and Alrutz, H. (1982). "Experimental investigations of statistical properties of diffuse sound fields in reverberation rooms". *Acustica* 51(3): 145-153.

- [19] Moorer, J. A. (1979). "About this reverberation business". *Computer Music Journal* 3(2): 13-18.
- [20] Flandrin, P. (1985). "Analyse temps-fréquence par transformation de Wigner-Ville". *Traitement du Signal* 2(2): 144-150.
- [21] Cohen, L. (1989). "Time-frequency distributions - a review". *Proceedings of the IEEE*, vol. 77(7): 941-981.
- [22] Reichardt, W., and Lehmann, U. (1978). "Raumeindruck als oberbegriff von raumlichkeit und halligkeit, erlauterungen des raumeindrucksmaßes". *Acustica* 40: 174-183.
- [23] Vanderkooy, J. (1994). "Aspects of MLS measuring systems". *J. Audio Eng. Soc.* 42(4).
- [24] Harris, F. J. (1978). "On the use of windows for harmonic analysis with the discrete Fourier transform". *Proceedings of the IEEE* 66(1): 51-83.
- [25] Page, C. H. (1952). "Instantaneous power spectra". *J. Appl. Phys.* 23(1): 103-106.
- [26] Grace, O. D. (1981). "Instantaneous power spectra". *J. Acoust. Soc. Am.* 69: 191-198.
- [27] Flandrin, P., Escudie, B., and Martin, W. (1985). "Représentation temps-fréquence et causalité". *Proc. 10ème Colloque Traitement du Signal et Applications (GRETSI)*: 65-70.
- [28] Barron, M. and Lee, L. J. (1988). "Energy relations in concert auditoriums". *J. Acoust. Soc. Am.* 84: 618-628.
- [29] Warusfel, O. and Jullien, J.-P. (1992). "Une campagne de mesures objectives et perpectives en acoustique des salles". *Proc. 2nd French Congress on Acoustics* 1: 151-154.
- [30] Kahle, E. (1995). "Validation d'un modèle objectif de la perception de la qualité acoustique des salles dans un ensemble de salles de concert et d'opéras". Doctoral dissertation, Univ. du Maine, Le Mans.
- [31] Jordan, V. L. (1981). "A group of objective acoustical criteria for concert halls". *Applied Acoustics* 14: 253-266.
- [32] Pelorson, X., Vian, J.-P., and Polack, J.-D. (1991). "Stabilité et variabilité des critères objectifs utilisés en acoustique des salles", *Journal d'Acoustique* 4: 599-619.
- [33] Martin, N., Mars, J., Pierron, J.-C., and Martin, J. (1991). "Représentation temps-fréquence et acoustique des salles", *Proc. 13ème Colloque Traitement du Signal et Applications (GRETSI)*:
- [34] Griesinger, D. (1995). "How loud is my reverberation?". Proc. 98th Conv. Audio Eng. Soc., preprint 3943.
- [35] Jullien, J.-P. (1986). "Correlations among objective criteria of room acoustic quality". *Proc. 1986 International Congress on Acoustics* (Toronto) 2: E4-E9.
- [36] Griesinger, D. (1996). "Spatiousness and envelopment in musical acoustics". Proc. 101st Conv. Audio Eng. Soc., preprint 4401.
- [37] Warusfel, O. (1990). "Etude des paramètres liés à la prise de son pour les applications d'acoustique virtuelle". *Proc. 1rst French Congress on Acoustics* 2: 877-880.
- [38] Warusfel, O., Derogis, P., and Caussé, R. (1997). "Radiation synthesis with digitally controlled loudspeakers". Proc. 103rd Conv. Audio Eng. Soc..
- [39] Blauert, J. (1983). *Spatial Hearing. The Psychophysics of Human Sound Localization*. MIT Press.
- [40] Thiele, G. (1986). "On the standardization of the frequency response of high-quality studio headphones". *J. Audio Eng. Soc.* 34(12).

- [41] Möller H., Jensen, C. B., Hammershoi, D., and Sorensen, M. F. (1995). "Design criteria for headphones". *J. Audio Eng. Soc.* 43(4).
- [42] Jot, J.-M., Larcher, V., and Warusfel, O. (1995). "Digital signal processing issues in the context of binaural and transaural stereophony". Proc. 98th Conv. Audio Eng. Soc., preprint 3980.
- [43] Larcher, V. (1997). "Correction individuelle pour la reproduction binaurale d'enregistrements effectués dans l'habitacle d'un véhicule automobile". IRCAM tech. report.
- [44] Larcher, V. and Jot, J.-M. (1997). "Equalization methods in binaural technology". In preparation.
- [45] Boll, S. F. (1979). "Suppression of acoustic noise in speech using spectral subtraction", *IEEE Trans. Acoust., Speech and Sig. Proc.* 27(2): 113-120.
- [46] Cappé, O. and Laroche, J. (1995). "Evaluation of short-time spectral attenuation techniques for the restoration of musical recordings". *IEEE trans. Speech and Audio Processing* 3(1): 84-93.
- [47] Crochiere, R. E., and Rabiner, L. R. (1983). *Multirate Digital Signal Processing*. Prentice Hall.
- [48] Reilly, A., and McGrath, D. (1995). "Convolution processing for realistic reverberation". Proc. 98th Conv. Audio Eng. Soc., preprint 3977.
- [49] Kleiner M., Dalenbäck, B.-I., Svensson, P. (1993). "Auralization - an overview". *J. Audio Eng. Soc.* 41(11): 861-875.
- [50] Griesinger, D. (1989). "Practical processors and programs for digital reverberation". *Proc. 7th Audio Eng. Soc. International Conference*.
- [51] Jot, J.-M. and Chaigne, A. (1991). "Digital delay networks for designing artificial reverberators". Proc. 90th Conv. Audio Eng. Soc., preprint 3030.
- [52] Jot, J.-M. and Warusfel, O. (1995). "A real-time spatial sound processor for music and virtual reality applications". *Proc 1995 International Computer Music Conference*.
- [53] Jot, J.-M. (1997). "Real-time spatial processing of sounds for music, multimedia and human-computer interfaces". Submitted to *ACM Multimedia Systems J.* (special issue on Audio and Multimedia).
- [54] Marin, M. (1996). "Etude de la localisation en restitution du son pour la téléconférence de haute qualité". Doctoral dissertation, Univ. du Maine, Le Mans.
- [55] Chowning, J. (1971). "The simulation of moving sound sources". *J. Audio Eng. Soc.* 19(1).
- [56] Moore, F. R. (1983). "A general model for spatial processing of sounds". *Computer Music Journal* 7(6).
- [57] Jullien, J.-P., Kahle, E., Winsberg, S., and Warusfel, O. (1992). "Some results on the objective and perceptual characterization of room acoustical quality in both laboratory and real environments". *Proc. Institute of Acoustics XIV*(2).
- [58] Jullien, J.-P. (1995). "Structured model for the representation and the control of room acoustical quality". *Proc. 15th International Conf. on Acoustics*.
- [59] Gerzon, M. A. (1992). "The design of distance panpots". Proc. 92nd Conv. Audio Eng. Soc., preprint 3308.

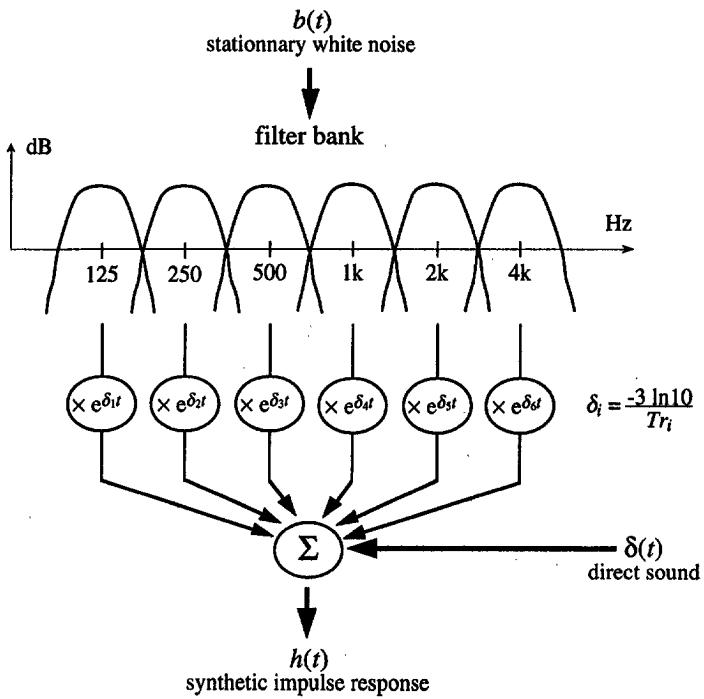


Figure 1: Method suggested by Moorer for synthesizing a room impulse reponse with a frequency-dependent reverberation time.

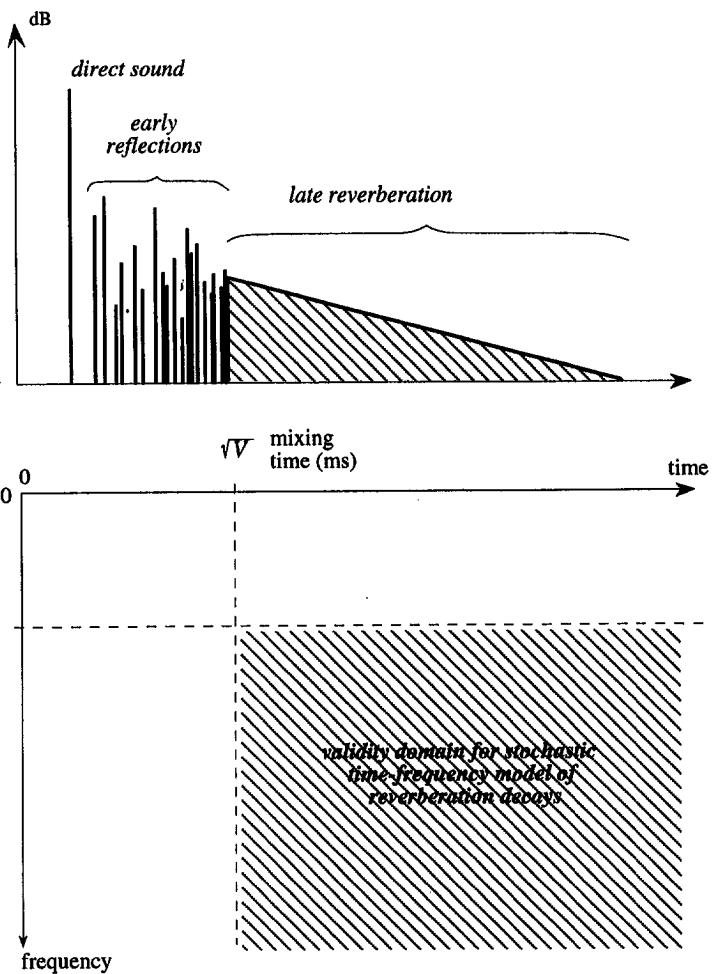


Figure 2: Validity domain of the stochastic model in the time-frequency plane (hatched).

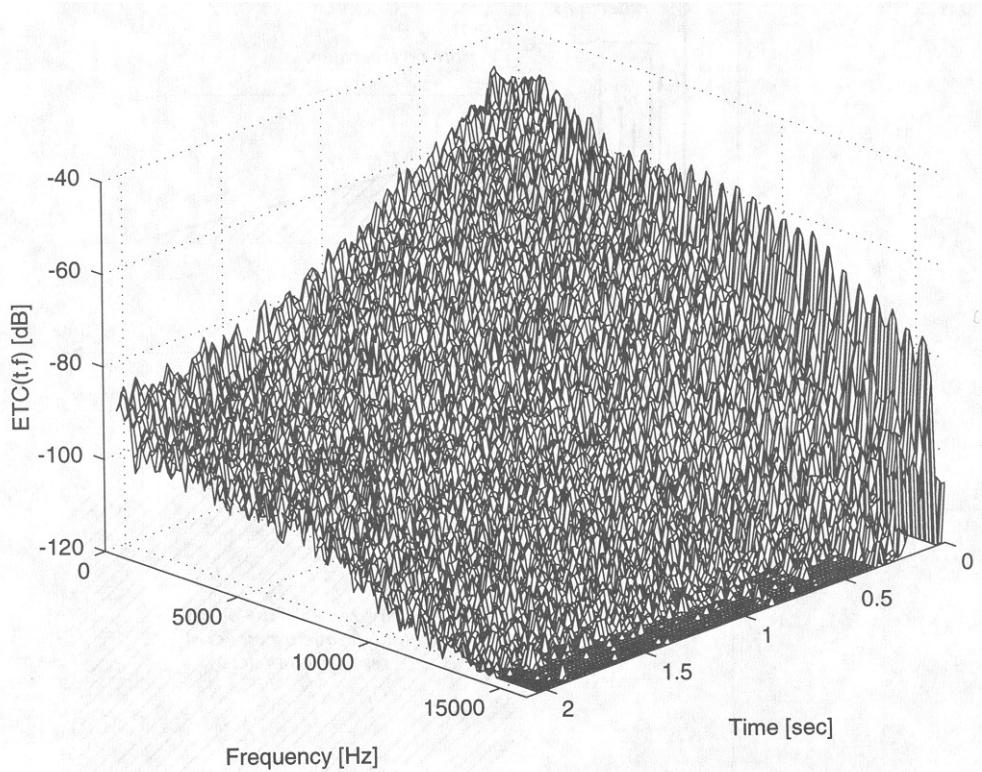


Figure 3: Short-time Fourier spectrum (spectrogram) computed from impulse response measured in Berlin Philharmonie hall at a sample rate of 32 kHz.
(frequency resolution: 62.5 Hz; time resolution: 16 ms)

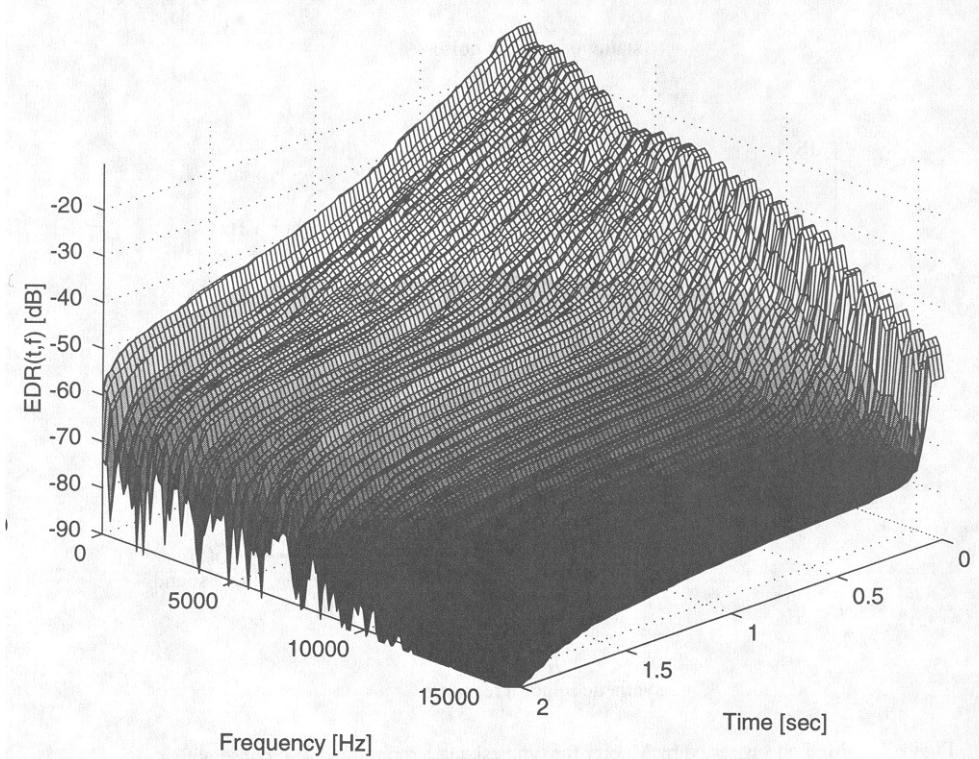


Figure 4: EDR computed by backward integration along time on the spectrogram of Fig. 3, without removing the effect of the measurement noise.

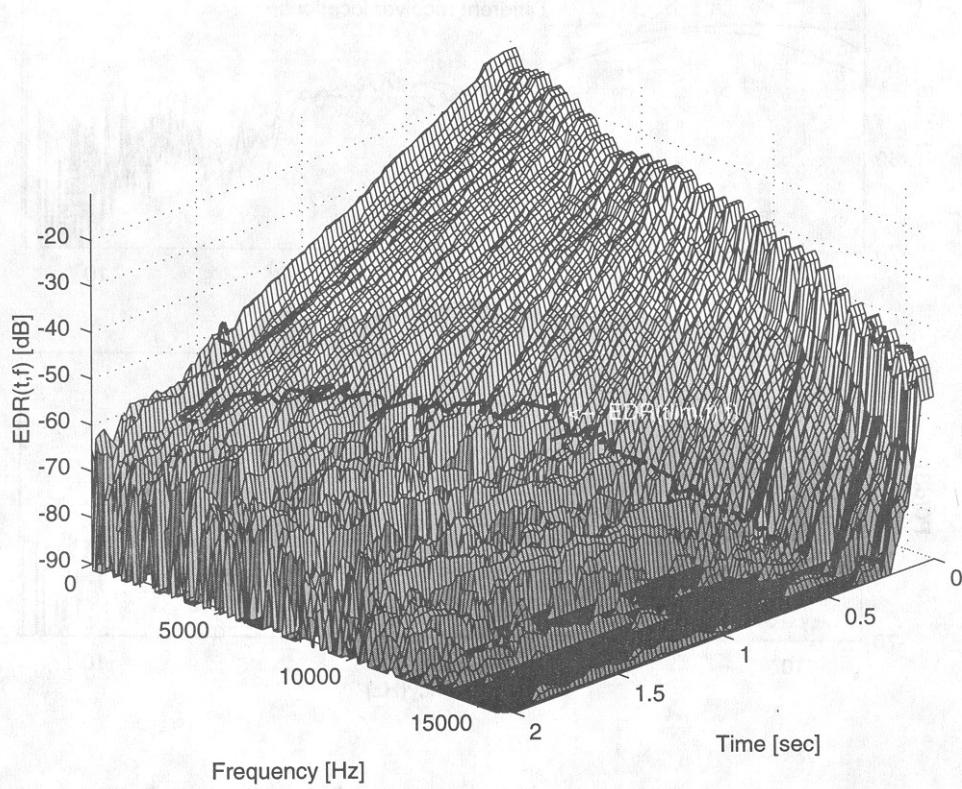


Figure 5: EDR computed from the spectrogram of Fig. 3, with prior removal of the estimated noise spectrum from each instantaneous spectrum.

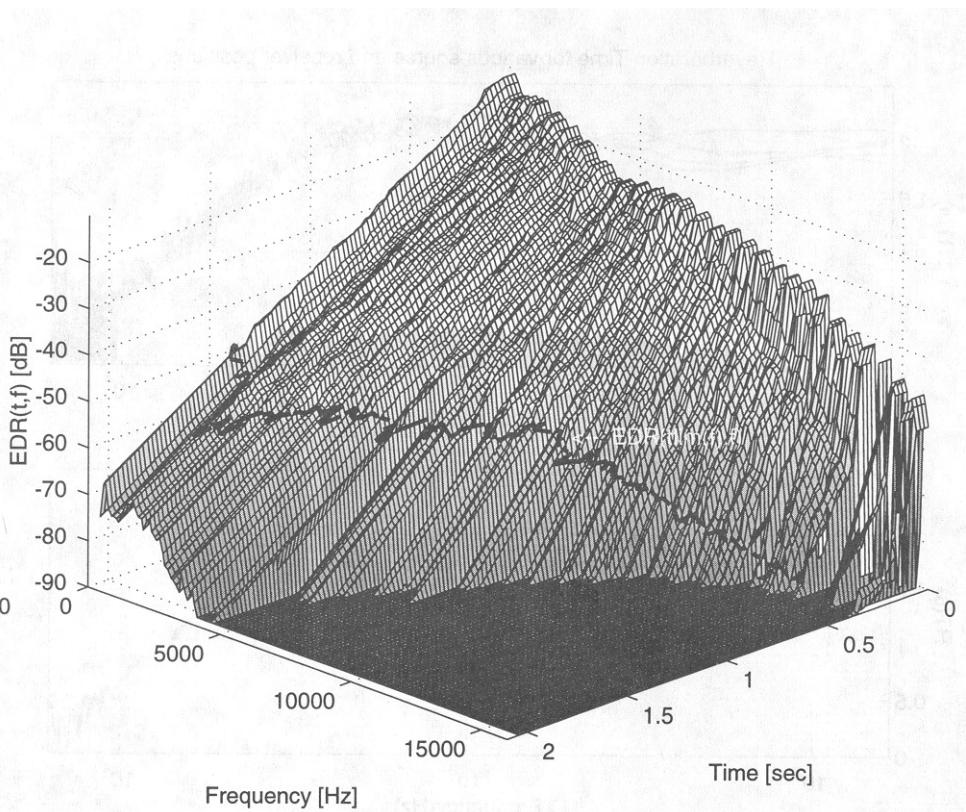


Figure 6: EDR computed as in Fig. 5, with ideal extrapolation of the estimated exponential decay in each frequency channel.

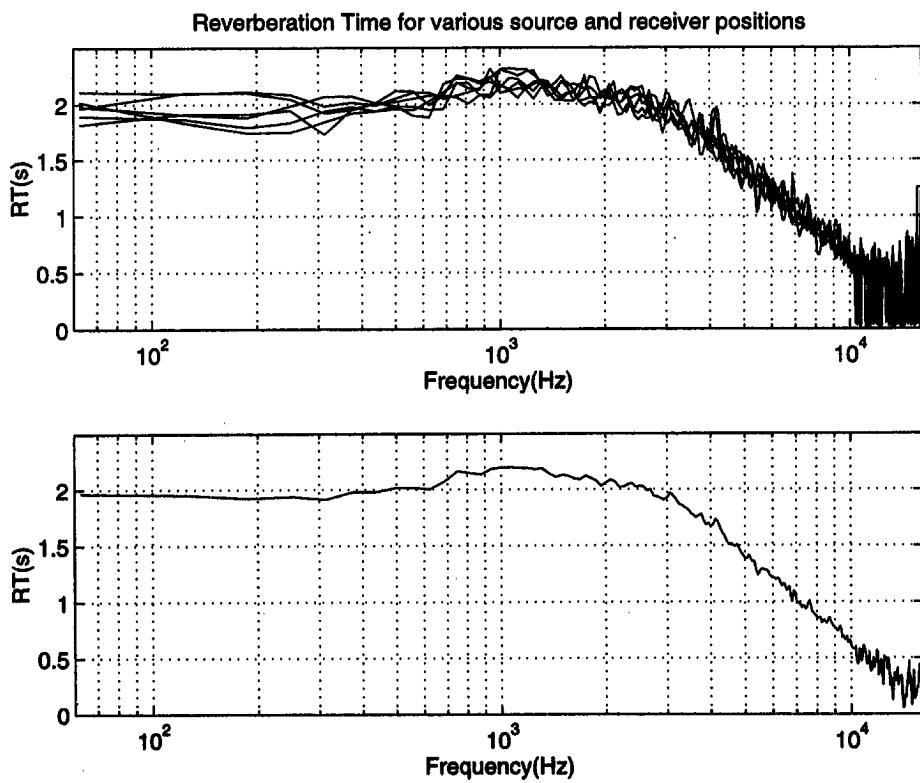


Figure 7: (top) Reverberation time vs. frequency estimated for 6 different receiver positions in the Berlin Philharmonic hall. (bottom) Estimated reverberation time curve computed by averaging the above reverberation time estimates.

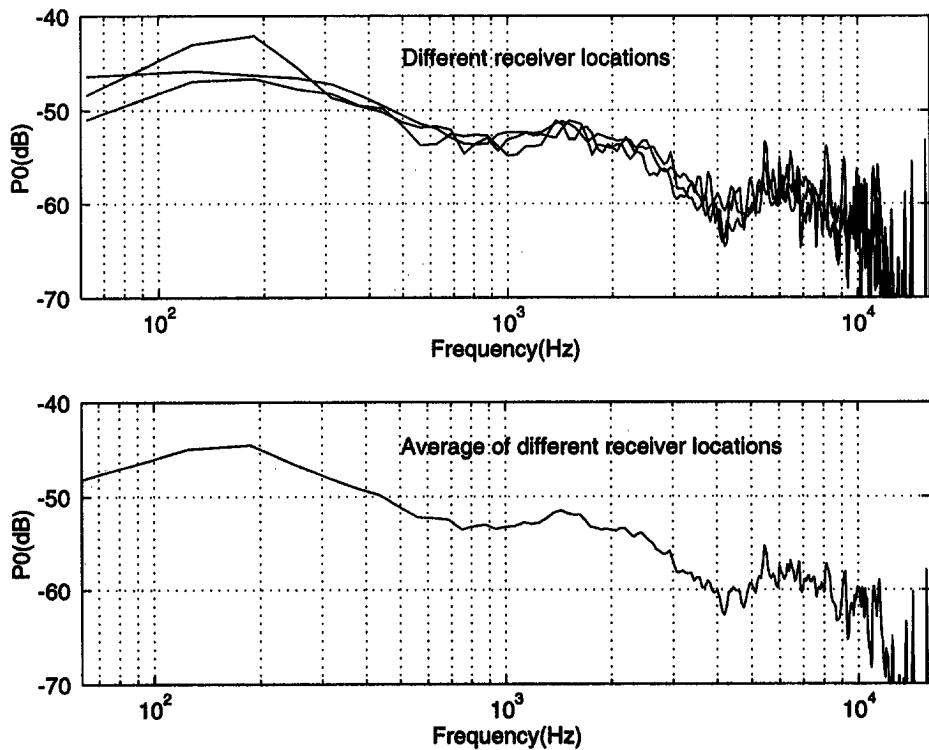


Figure 8: (top) Initial spectrum estimated for 3 different receiver positions in the Berlin Philharmonie hall. (bottom) Initial spectrum estimated by power-averaging from the above.

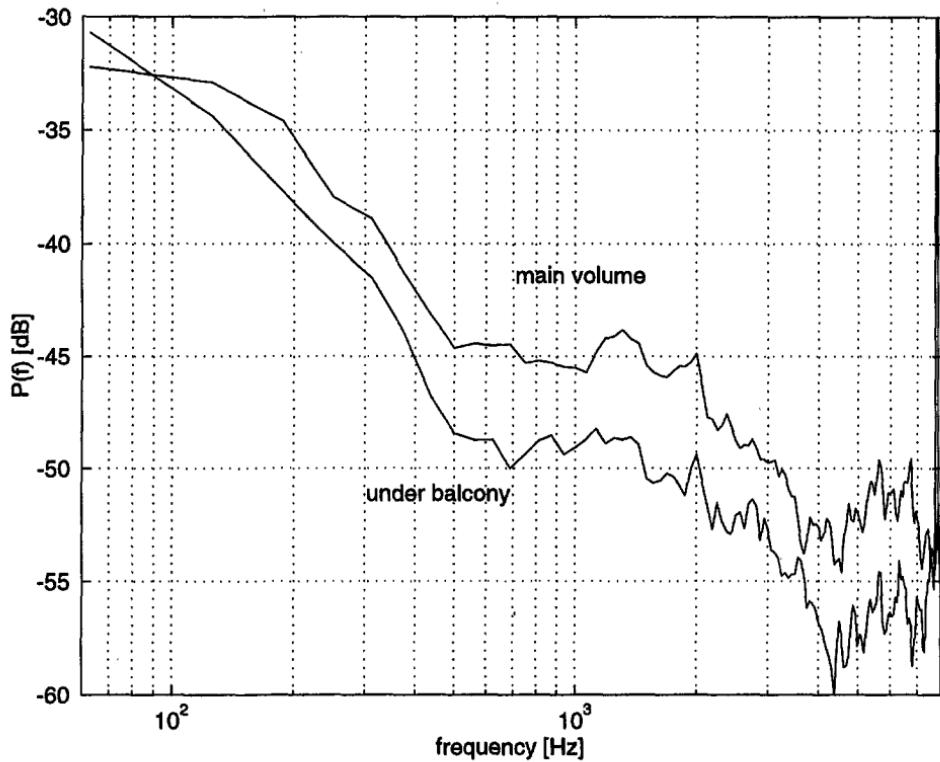


Figure 9: Comparison of initial spectrum in the main volume and under the first balcony in the Salle Pleyel (Paris). Initial spectrum for main volume computed by averaging over different receiver positions. Initial spectrum under balcony computed for one receiver position, by averaging over different source positions on the stage.

Initial Power Spectra : different microphones

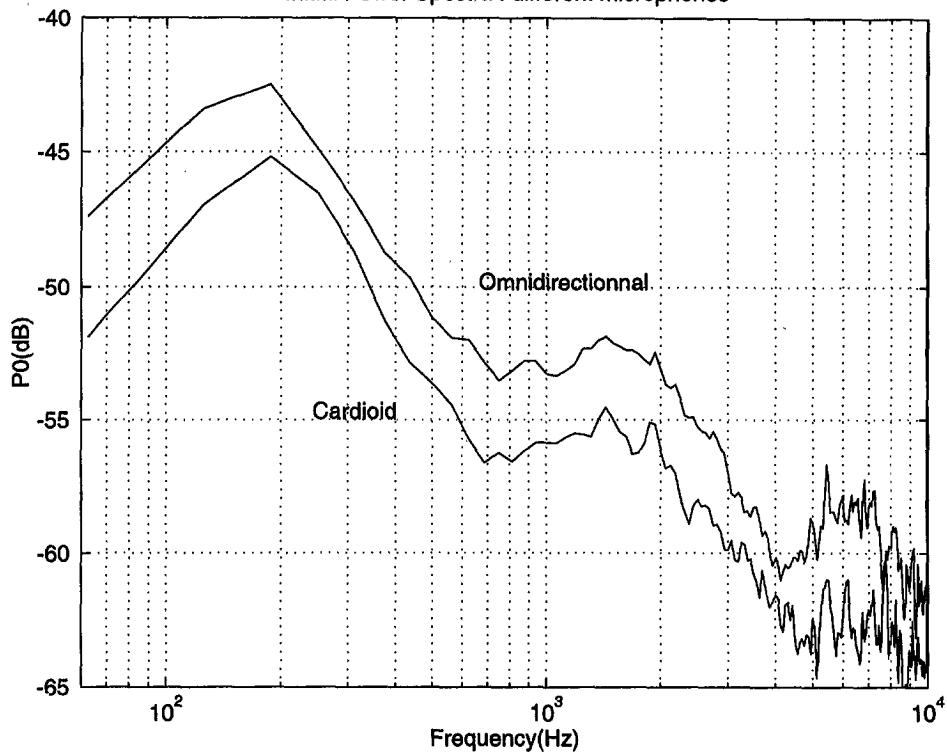


Figure 10: Comparison of initial spectrum obtained with two different microphone sensitivities in the Berlin Philharmonie hall (computed by averaging over several receiver positions in each case).

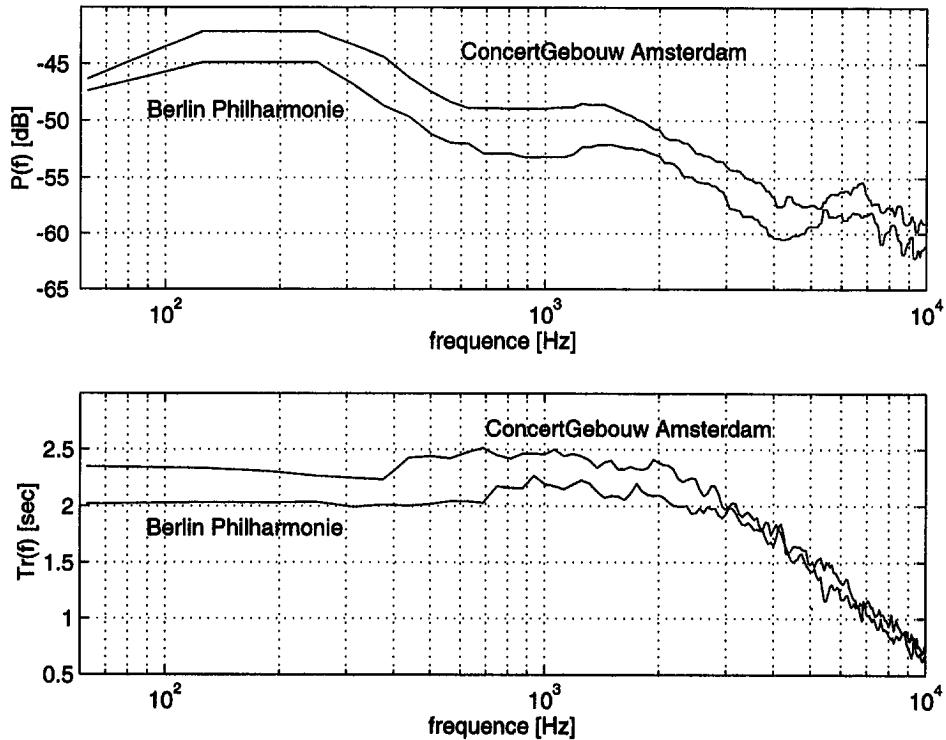


Figure 11: (top) Initial spectrum estimated for two different halls, using the same microphone and the same loudspeaker. (bottom) Estimated reverberation time vs. frequency, in each hall.

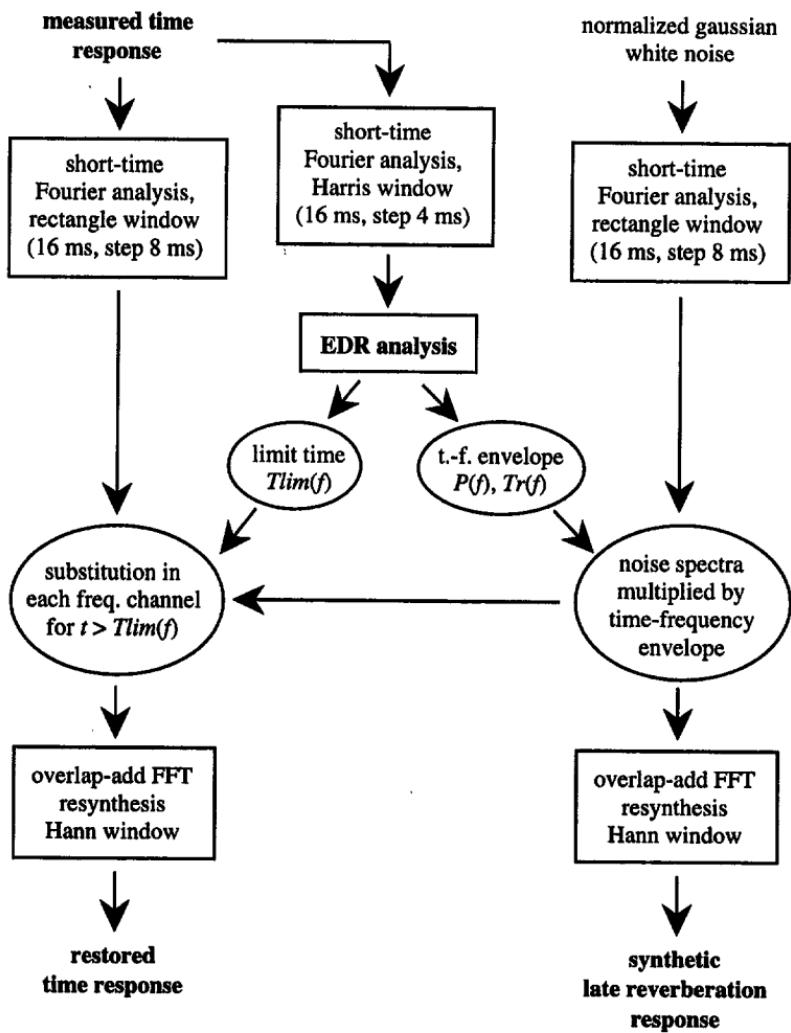


Figure 12: Diagram of analysis / synthesis procedure for restoring impulse responses corrupted by measurement noise or synthesizing artificial late reverberation responses. The limit time $T_{lim}(f)$, the initial spectrum $P(f)$ and the reverberation time $Tr(f)$ are estimated by the EDR analysis procedure described in section 3.2.

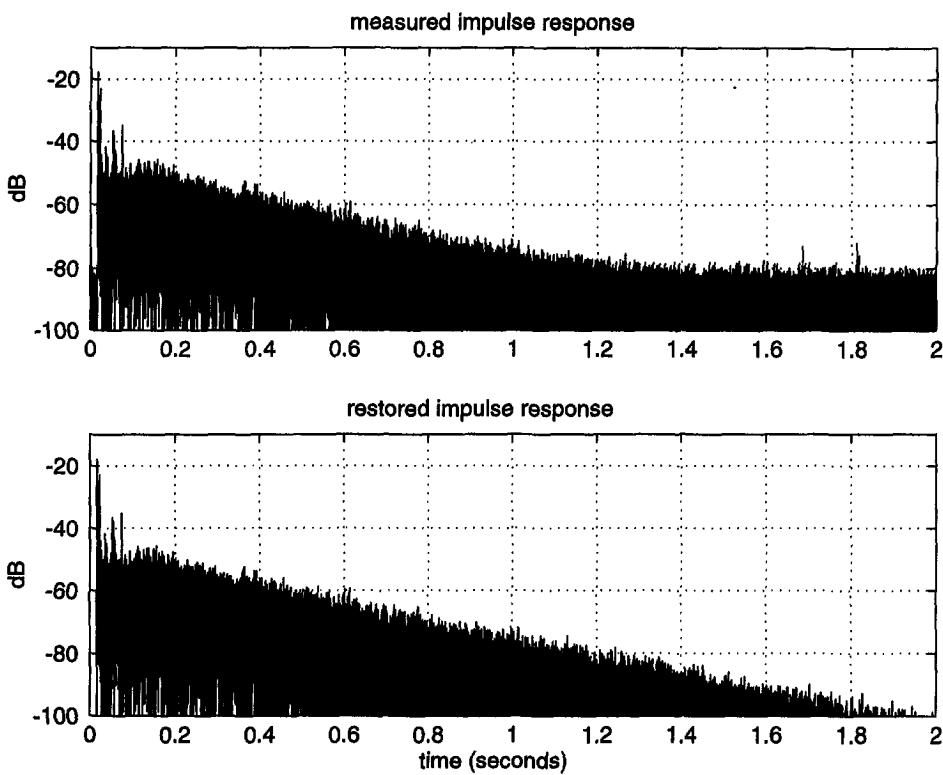


Figure 13: Echograms of original impulse response and restored impulse response, where the measurement noise has been replaced by a synthetic exponential reverberation decay prolonging the natural decay at each frequency, by the method of Fig. 12.

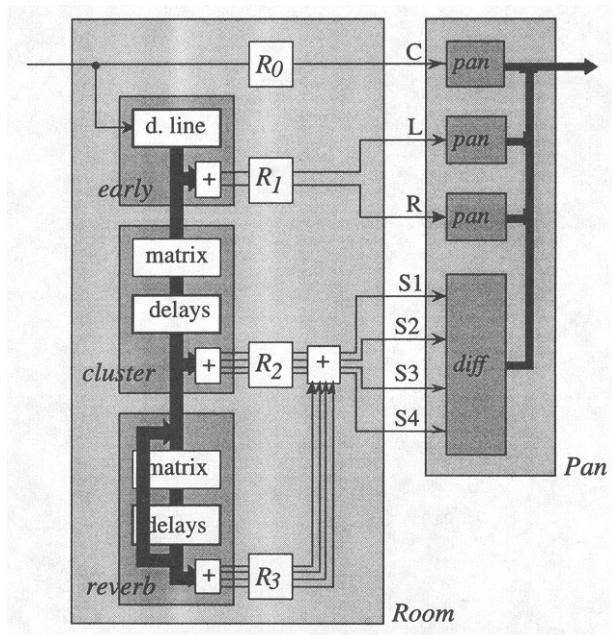
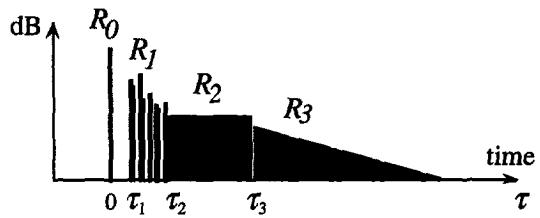


Figure 14: (top) Generic impulse response model. (bottom) Structure of a spatial sound processor adaptable to various reproduction formats over loudspeakers or headphones, associating an artificial reverberation module (*Room*) and a directional rendering module (*Pan*), as implemented in the IRCAM *Spat* software.