The goal of prediction, in Room Acoustics, is to synthesize the impulse responses (IRs) of a hall, in order to derive acoustic indices or to allow auralization. The process assumes the hall to be a time invariant linear system. Furthermore, the IR is known to behave stochastically when the sound field becomes diffuse, that is, after a certain time called mixing time. This study aims at characterizing the IR mixing time. Three methods are presented for visualizing and detecting the time evolution of the IR behaviour. The first one highlights the transition from early reflections to diffuse sound field by monitoring the phase evolution versus time. The two others exploit the gaussian distribution of pressure in a diffuse sound field, when the IR becomes statistical. These methods are applied to measurements, carried out in Salle Pleyel, and confirm the simple relationship found earlier between mixing time and volume.

1 Introduction

In room acoustics, impulse responses (IRs) are composed of the direct sound, arrivals (an arrival is a sound ray which reaches the listener ears after having undergone one or more reflections on the boundaries of the room), and the diffuse sound field, more often called the late reverberation (Fig.1). The arrivals, commonly named early reflections, are a set of discrete events. IRs can be synthesized using different sets of techniques, such as image-sources [1], sound rays [2], etc. Knowing the geometry of the room and the distance source-receiver, the direct sound and early reflections can easily be synthesized. The diffuse sound field is assumed to be a gaussian process [3], and is synthesized as such [4]. The transition from early reflections to late reverberation, that occurs after a certain time called the mixing time, is supposed to be related to the volume of the considered hall [5].

In order to investigate the previous relationship, this paper proposes three methods for estimating the mixing time of experimental IRs, with knowledge of the volume of the hall. After reviewing statistical properties of room acoustics in Section 2, Section 3 introduces an original way of visualizing the phaseshift evolution of the signal vs. time. Then, Section 4 presents two simple estimators of the mixing time based on the statistical properties of the diffuse sound field. In Section 5, the phaseshift evolution and the gaussianity estimators are used to estimate the mixing time of 21 experimental room impulse responses, measured in Salle Pleyel [6], with pistol shots as sound sources. A comparison between the three methods is carried out with focus on the spreading of the results. Further, the statistical relevance of each method is discussed. Finally, this paper conludes with a link bewteen the mixing time and the mean free path [5].

2 Some statistical properties of room acoutics

According to [3] and [7], the modal density (D_m -average number of modes per Hz) of an IR is proportionnal to the square of frequency f, while the echo density (D_e - average number of reflections per second) is proportional to the square of time t, such as:

$$D_m(f) = 4\pi V \frac{f^2}{c_0^3} \quad (1)$$

$$D_e(t) = 4\pi c_0^3 \frac{t^2}{V} \quad (2)$$

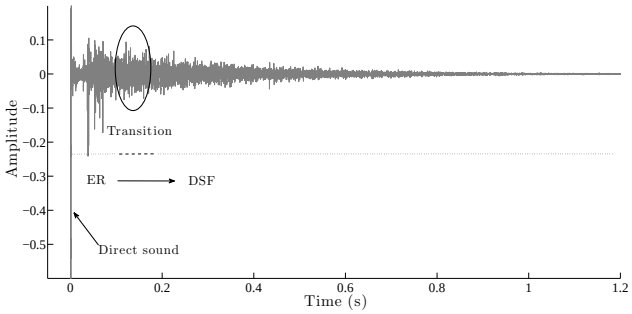

Figure 1: Experimental IR. Note that the transition from the early reflections (noted ER on the graph) to the diffuse sound field (noted DSF on the graph) is not well-defined.

where c_0 is the speed of sound in $m.s^{-1}$, and V is the volume of the room in m^3.

These expressions (Eq.(1)-(2)) can easily be demonstrated for simple shape rooms (rectangular for instance), but only Eq.(1) can be generalized to rooms of any geometry [8]. For larger times and higher frequencies, both densities become large. This provides the foundation of statistical models of room responses, as developped in the frequency domain [9], and more recently in the time domain [3]. The resulting time-frequency model is valid for the later reverberation decay at frequencies above the "Schroeder frequency", and is essential basis for artificial reverberation techniques [10] [11]. Equation (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 bandwith Δf of a mode. Thus, at high frequencies, any source signal will simultaneously excite several modes. The complex frequency response can be considered as a space frequency dependent stochastic process whose real and imaginary parts are some independent gaussian processes having the same variance [7] [9] [12] [13]. These properties are valid irrespective of the listener position and of the room (above a limit frequency - "Schroeder frequency" [14]- which depends on the room). In the time domain too, there is also a time (the mixing time) after which the stochastic model becomes valid [15]. The time domain response can only be gaussian if a sufficient number of reflections overlap at any time along the response. Since the reflection density increases with time, according to Eq.(2), the situation is similar to that found in the frequency domain. 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 bandwidth of modes), but with

reference to the bandwidth of the source, which determines the spreading of the source.
Polack [3] proposes as a criterium that 10 reflections overlap within a characteristic time resolution of the auditory system, taken equal to $24ms$ [7]. Then Eq. (2) leads to:

$$t_{mixing} \approx \sqrt{V} \tag{3}$$

where t_{mixing} is the mixing time, expressed in ms, and V is the volume of the room in m^3.

This value was proposed as a reasonnable approximation for the transition time between early reflections and late reverberation. It is shown in [3] and [16] that the exponentially decaying stochastic model can be established within the framework of geometrical acoustics and billiard theory [17] [18]. The mixing time is defined as the time it takes for initially adjacent sound rays to spread uniformly across the room. By that time (the origin is taken as the time of emission of the impulse by the source), the process has become diffuse, i.e. the acoustical energy density and the direction of the intensity are uniformly distributed across the room. The mixing character of a room depends on its geometry and on the diffusing properties of the boundaries or the hall. Consequently, the value $\sqrt{V}$ can be considered as an upper limit for the mixing time in mixing rooms, as it has been discussed in [16] [19].

In the following, three different methods are used to estimate the mixing time in experimental IRs, and compared to the theoretical value given by Eq.(3).

3 Phaseshift time evolution

3.1 The eXtensive Fourier Transform: XFT

When the source emits in the room, sound rays travel into the air and hit walls once or several times before reaching the ears of the listener. Each time a sound ray hits a surface, it is reflected according to the intrinsic properties of the materials. A part of the incident energy is reflected specularly, another is absorbed and the last part is scattered in the space. Depending on the geometrical configuration of the surfaces (diffusers) and on the properties of the materials, sound rays are filtered out (in modulus and in phase) so that phaseshifts and scattering are assumed to be an increasing function of time [7]. Hence, the phase of the IR is expected to behave more and more randomly as time goes by, until being definitely random after the mixing time.
The authors propose an original approach for discriminating this phenomenum. Basically, it is a Fourier Transform on an extensible window of variable width, instead of a sliding window of fixed width, as in the Short Time Fourier Transform ($STFT$). The idea is to document the evolution of the system at each time step, according to the past of the signal, by conserving the time origin, instead of using a snapshot as in the $STFT$. This new transformation is called XFT, as eXtensive Fourier Transform, and is defined by:

$$Y(w,\nu) = \sum_{n=0}^{wN-1} y(n).e^{-2j\pi n\nu} \tag{4}$$

where $y(n)$ is the signal to analyze (of N samples length), and w is a hop size window.

One could interpret the XFT as a kind of cumulative integration of $y(n)$ in time and in frequency domains. At each time step, the XFT is calculated for the same total number of samples (using zero padding), in order to keep the same frequency resolution for each time window. Obviously, the smaller the hop size window w, the higher the time resolution, but the longer the computational time is.

3.2 Visualization of the phaseshift

Figure 2 shows several graphs of the unwrapped phases of the XFT, according to Eq.(4), at several time steps. As expected from Section 3.1 and [7], the phase wraps faster between $\pm\pi$ as time progresses, denoting the increasing randomness of the signal. Moreover, while the unwrapped phase of the first milliseconds (and especially of the direct sound) is a linear function of frequency, the phase of late reverberation becomes rugged. This gradually change of behaviour is assumed to be an efficient way to visualize the randomness of the signal, and thus to estimate the mixing time. This is achieved by calculating, at each time step, the mean regression error, called $D(t)$), between a linear regression made on the unwrapped phase and the unwrapped phase itself (Fig.3). $D(t)$ increases with time, showing abrupt changes of behaviour, materialized by several inflexion points. The mixing time is choosen as the first inflexion point. It is noticeable that $D(t)$ equals zero for the direct sound, and increases between reflections, underlying a diffusion process occuring at each reflection. Further, $D(t)$ decreases for each arrival, since in the early part of the IR, arrivals are still correlated with the direct sound. Figure 3 (bottom) presents a scattering effect of $35ms$ approximately, caused by a balcony. During this time, the diffusion makes the phase behaviouring randomly.

4 Gaussianity estimators

The transition from early reflections to late reverberation occurs gradually, as mentionned in Section 2, so that the temporal distribution of the signal tends toward a gaussian distribution. The following two sections present two frequently used statistic tools (cumulants, also used in [20]), that help defining a gaussian distribution: the standard deviation and the kurtosis. Higher order cumulants contain amplitude and phase information, unlike second order statistics (such as correlation), which are phase blind [21]. Further, cumulants of order greater than two measure the non-gaussian nature of a time series.

4.1 Standard deviation

The standard deviation (σ, second order cumulant) of a group of samples is a measure of the spread of the

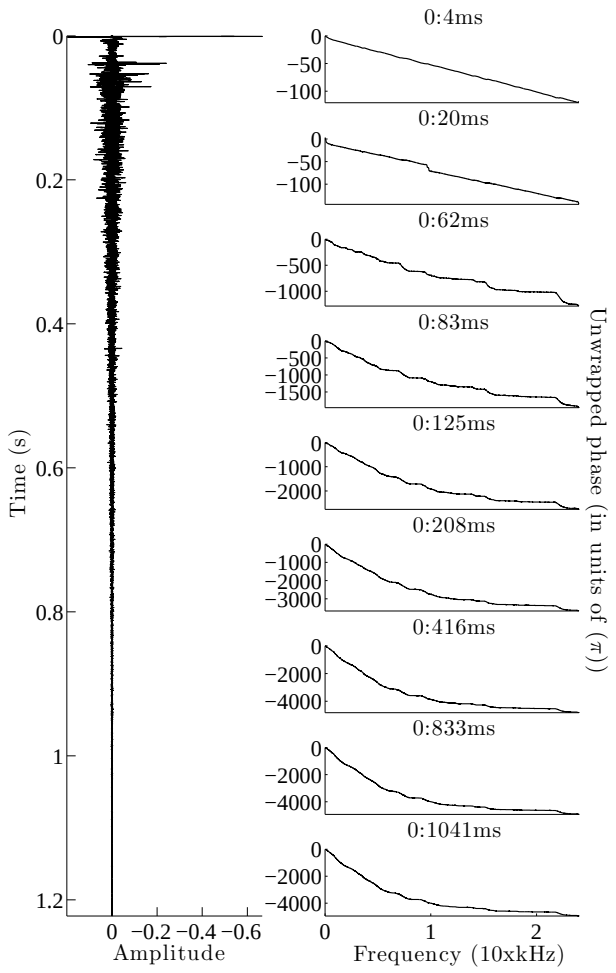

Figure 2: Unwrapped phase of the XFT calculated, for several time steps (right), on an experimental IR (left).

samples around the average, and is defined as:

$$\sigma = \sqrt{E(x^2) - E^2(x)} \tag{5}$$

where $E(x)$ is the expected value of x.

In a normal and centered distribution, two third of the samples lie between $[-\sigma; +\sigma]$, while a third of the samples lie between $]-\infty; -\sigma[$ and $]\sigma; +\infty[$. The mixing time is estimated as the time when the ratio $R(n)$, defined in Eq.(6), reaches 2.

$$R(n.h) = \sum_{n=0}^{N-1} \frac{N_{x(n.h)}[-\sigma; \sigma]}{N_{x(n.h)}[-\infty; -\sigma] + N_{x(n.h)}[\sigma; +\infty]} \tag{6}$$

where $N_{x(n.h)}(a; b)$ is the number of samples of $x(n.h)$ within a and b, and h is a hop size window.

In order to make a figure readable, the ratio $R(n)$ is normalized according to:

$$\tilde{R}(n) = \frac{R(n.h) - 2}{max(R(n.h))} \tag{7}$$

As seen in Fig.4, which presents the values taken by $\tilde{R}(n)$ computed on an experimental IR, early reflections present extreme values ($\tilde{R}(n) > 0$), while the progression of the signal toward a gaussian distribution is

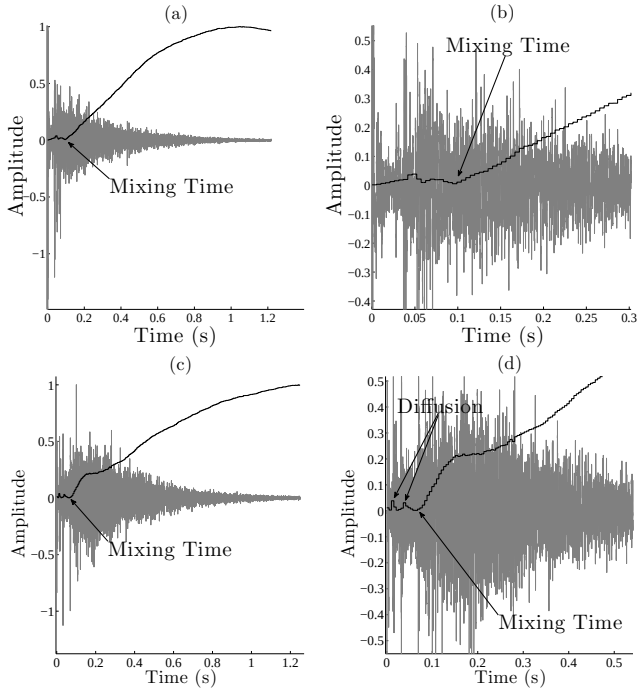

Figure 3: Estimation of the mixing time with the unwrapped phase on an experimental IR. Note that curves are normalized for an accurate readability. *Top*: the IR begins with the direct sound. *Bottom*: the IR begins with a scattering effect during the first $35ms$ (the diffusion is noticeable). Plots (b) and (d) are details of plots (a) and (c).

obvious around $100ms$, since $\tilde{R}(n) = 0$. The hop size window (h) is taken equal to $24ms$, since it is the time integration of the human ear [3].

4.2 Kurtosis

The fourth order zero lag cumulant of a process is called kurtosis. It is defined as (with normalization):

$$k = \frac{E(x-\mu)^4}{\sigma^4} - 3 \tag{8}$$

where E, μ and σ are the expected value, the mean and the standard deviation of x, respectively.

For a gaussian distribution, $k = 0$. Figure 4 also shows the values of k of calculated on an experimental IR, using a hop size window of $24ms$ width. As noticed previously, early reflections are easily recognizable, while the mixing regime is achieved when $k = 0$.

Both cumulants show that the IR evolves gradually from a less diffuse regime toward a more diffuse state of the system (Fig.4).

5 Mixing time estimation

The mixing times are estimated over 21 IRs measured in Salle Pleyel ($V = 19000m^3$), with pistol shots as sound sources [6]. The time elapsed between the emission of the impulse by the source and the reception at the measuring point is added to the estimated times. Results are presented in Table 1.

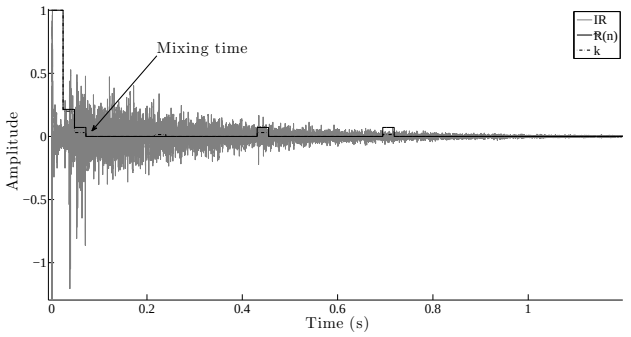

Figure 4: Estimation of the mixing time using the standard deviation and the kurtosis methods on an experimental IR.

The theoretical value of the mixing time (according to Eq.(3)) for the Salle Pleyel is $137ms$, which is in agreement with the results found by the three methods. Differences of variations between methods should be noticed. The estimation of the mixing time by computing the phase of the XFT provides the smallest variations. Moreover, the results confirm on the one hand that the mixing time is linked to the volume of the hall, and on the other hand, that its theoretical value is an upper limit.

6 Discussion

Attention must be paid to the drawback of using the standard deviation and the kurtosis as gaussianity estimators, because they both constitute necessary but not sufficient conditions to characterize a normal distribution. Moreover, as mentionned in [20], the identification of the mixing time becomes difficult for different rooms using standard deviation and kurtosis, since the values $R(n)$ and k are not constant along the curve. For instance, $\tilde{R}(n)$ and k equal zero in presence of diffusion between arrivals, even if it is not the mixing time. Further, the high sensitivity to the window width of analysis (h) of these two cumulants (Fig.5) makes these tools hardly reliable. On the other hand, the XFT does not suffer from the same limitation, as explained in Section 3.
One may notice that differences between the kurtosis and the XFT methods are smaller than between the XFT and the standard deviation methods, since the kurtosis is a local measure of the phase of the signal.

Sabine [7] defined the mean free path as the average distance covered by a sound ray between subsequent reflections. Blesser [5] assumes that the mixing time is approximately equals to three times the mean free path ($< Lm >$), which is given by:

$$< Lm >= \frac{4V}{S}(m) \qquad (9)$$

where V is the volume and S is the total surface of the hall.

In Salle Pleyel, we estimate: $S = 4530m^2$, $V = 19000m^3$ and the mixing time $\tilde{T}m = 130ms$. From Eq.(9), we obtain: $< Lm >= 17.3m$. And the ratio between the mixing time and the mean free path is:

$$\frac{\tilde{T}m}{< Lm > /c_0} = 2.7$$

where c_0 is the sound velocity. This confirms Blesser's assumption.

Distances (m)	$\tilde{T}_m$ (Φ) (ms)	$\tilde{T}_m$ (σ) (ms)	$\tilde{T}_m$ (k) (ms)
8.13	117	101	96
12.71	132	71	92
14.53	145	126	182
14.87	165	151	197
15.05	105	152	85
15.83	201	92	116
16.67	85	84	145
17.04	157	140	115
19.90	160	98	98
19.99	141	259	150
21.25	170	153	184
22.59	89	112	82
23.16	175	233	245
26.22	94	136	116
26.73	64	120	91
27.88	124	148	90
29.79	139	139	130
31.36	155	162	170
33.29	67	61	95
34.65	158	222	160
37.89	105	122	107
Average(ms)	131	140	130
σ(%)	28	38	34

Table 1: Estimated mixing times, where $\tilde{T}_m$ (Φ), $\tilde{T}_m$ (σ) and $\tilde{T}_m$ (k) are the mixing times estimated respectively by the phase of the XFT ($w = 5ms$), the standard deviation and the kurtosis methods ($h = 24ms$).

7 Conclusion

Impulse response synthesis is based on theroretical considerations, such as the time at which the system becomes stochastical. This paper proposes three methods for estimating the mixing time of room impulse responses. First, an original tool is introduced (XFT), which allows for the visualization and the computation of the Fourier Transform of the transient signal step by step, while keeping fixed time origin and time length. Cumulants such as the standard deviation and the kurtosis are not as relevant since results have a strong dependence on the window width of analysis.
This study permits to confirm the simple relationship that links the mixing time to the volume of a hall. Furthermore, it is also shown that diffuse reflections randomize the phase of the signal so that it gradually increases with time.
Further investigations should be made on other concert halls in order to test the robustness of the XFT, as well as to test if the XFT is perceptually relevant, and if

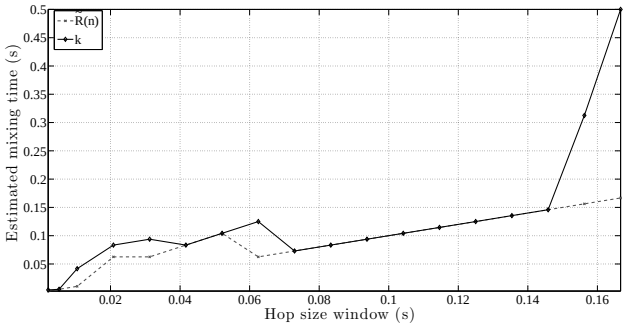

Figure 5: Estimation of the mixing times using different sizes of the window analysis, h.

the randomness of the phase can be linked to any global properties of a room.

8 Acknowledgments

This work was partly supported by grants from Région Ile-de-France, France.

References

[1] J. B. Allen and D. A. Berkely, "Image method for efficiently simulating small-room acoustics," *Journal of the Acoustical Society of America*, vol. 65, no. 4, pp. 943–950, April 1979.

[2] *Computer modelling and auralisation of sound fields in rooms*, vol. 38, Elsevier Applied Acoustics, 1993.

[3] J-D. Polack, *La transmission de l'énergie sonore dans les salles*, Ph.D. thesis, Thèse de doctorat d'Etat, Université du Maine, Le Mans, France, 1993.

[4] James A. Moorer, *About This Reverberation Business*, Ircam, 1978.

[5] B. Blesser, "An interdisciplinary synthesis of reverberation viewpoints," *Journal of the Acoustical Engineering Society*, vol. 49, no. 10, pp. 867–903, October 2001.

[6] G. Defrance, J-D. Polack, and B-FG. Katz, "Measurements in the new Salle Pleyel," in *Proc. Int. Symp. Room Ac.*, Sevilla, September 2007.

[7] L. Cremer and H. Müller, *Principles and Applications of Room Acoustics*, vol. 1, Applied Science Publishers Ltd, 1982.

[8] H. Weyl, "Das asymptotische Verteilungsgesetz der Eigenwerte linearer partieller Differentialgleichungen (the asymptotic distribution law for the eigenvalues of linear partial differential equations)," *Math. Ann.*, vol. 71, pp. 441–479, 1912.

[9] M.R. Schroeder, "Statistical parameters of the frequency response curves of large rooms," *Journal of the Acoustical Engineering Society*, vol. 35, no. 5, pp. 307–316, May 1987.

[10] J-M. Jot, *Etude et réalisation d'un spatialisateur de sons par modèles physiques et perceptifs*, Ph.D. thesis, Doctoral Dissertation, Télécom Paris, France, 1992.

[11] M. R. Schroeder, "Natural-sounding artificial reverberation," *Journal of the Acoustical Engineering Society*, vol. 10, no. 3, 1962.

[12] M. R. Schroeder, "Frequency-correlation functions of frequency responses in rooms," *Journal of the Acoustical Society of America*, vol. 34, no. 12, pp. 1819–1823, 1962.

[13] K. J. Ebeling, K. Freudenstein, and H. Alrutz, "Experimental investigations of statistical properties of diffuse sound fields in reverberation rooms," *Acustica*, vol. 51, no. 3, pp. 145–153, 1982.

[14] P.M. Morse and K.U. Ingard, *Theoretical Acoustics*, McGraw-Hill Book Company, 1968.

[15] J-D Polack, "Modifying chambers to play billiards: the foundations of reverberation theory," *Acta Acustica*, vol. 76, pp. 257–272, February 1992.

[16] J-D Polack, "Playing Billiards in the Concert Hall: The Mathematical Foundations of Geometrical Room Acoustics," *Applied Acoustics*, vol. 38, pp. 235–244, 1993.

[17] W. B. Joyce, "Sabine's reverberation time and ergodic auditoriums," *Journal of the Acoustical Society of America*, vol. 58, no. 3, pp. 643–655, September 1975.

[18] F. Mortessagne, O. Legrand, and D. Sornette, "Transient chaos in room acoustics," *CHAOS, American Institute of Physics*, vol. 3, no. 4, pp. 529–541, 1993.

[19] J-M. Jot, L. Cerveau, and O. Warusfel, "Analysis and synthesis of room reverberation based on a statistical time-frequency model," in *103rd AES Convention, New York, NY*, September 1997.

[20] R. Stewart and M. Sandler, "Statistical measures of early reflections of room impulse responses," in *Proc. of the 10th Int. Conference on Digital Audio Effects (DAFx-07), Bordeaux, France*, September 2007, pp. 59–62.

[21] J. N. Chrysostomos and P. P. Athina, *Higher-order spectra analysis : a nonlinear signal processing framework*, Prentice Hall, 1993.