Using Matching Pursuit for estimating mixing time within Room Impulse Responses

G. Defrance, L. Daudet and J-D. Polack
UPMC Univ. Paris 06, IJLRDA LAM, CNRS UMR 7190
11 rue de Lourmel, 75015 Paris, France

March 29, 2009

Abstract

In Room Acoustics, the quantity that fully describes the hall is a set of room impulse responses (RIRs), which are composed of the succession of arrivals (i.e., some sound rays which have undergone one or more reflections on their way from the source to the receiver). The mixing time is defined as the time it takes for initially adjacent sound rays to spread uniformly across the room. This paper proposes to investigate the temporal distribution of arrivals and the estimation of mixing time. A method based on maxima of correlations (Matching Pursuit) between the source impulse and the RIR allows to estimate in practice arrivals. This paper compares the cumulative distribution function of arrivals of experimental and synthesized RIRs (using a stochastic model). The mixing time is estimated when the arrival density becomes constant. The dependance of mixing time upon the distance source/receiver is investigated with measured and synthesized RIRs.

Keywords
Room acoustics; stochastic processes; mixing; stochastic modelisation; statistical properties of signals

1 Introduction

In 1975, Joyce [1] revolutionized the approach to auditorium acoustics by introducing to the acoustical community the concept of ergodic dynamical systems. Indeed, this theory was able to justify some earlier spectral observations in the frequency domain, namely the existence of discrete modes at low frequencies and modal superposition at high frequency. The purpose of the present paper is to illustrate similar findings in the time domain, namely a domain where reflections are singled out, shortly after the direct sound, followed at longer times by superposed reflections that follow a different distribution in time (Fig.1). More specifically, we focus on the experimental evaluation of the transition time between the two regimes, which is called the *mixing time* throughout this paper, in reference to Krylov [2].
Indeed, a standard way to document the acoustics of a room is to measure a set of Room Impulse Responses (RIRs). The RIR is built by the superposition of arrivals, that is, modified versions of an original pulse emitted by the source and reaching the receiver after travelling through the room. Sound emitted by the source undergoes scattering and absorption when encountering boundaries of the room. The source pulse is therefore divided in many wavelets that each follows a different trajectory within the hall. The RIR is composed of the succession of all these trajectories, called *arrivals* [3], with due consideration for their respective amplitudes. In room acoustics, the time distribution of these arrivals plays an important role since it is directly linked to the acoustical quality of the room, as is demontrated in numerous books [4] [5]. But so far, no demonstrattion of the role of the mixing time has been given, since its estimation was lacking of consistency [6]. The goal of this paper is to study the relevance of using a tool from the signal processing domain as an estimator of mixing time on experimental room impulse responses, and to check its consistency on a model of RIRs. The comparison of our estimator to some others is no goal of the present paper.
This article is built as follows.
Section 2 reviews the ergodic theory of room acoustics and the idea of a transition time, insisting on the time domain.
The mixing time is estimated by studying the temporal distribution of the arrivals of RIRs. This is achieved using a well-known audio decomposition technique called Matching Pursuit (MP) [7] (Section 3).
Assuming the hall to be a system of time invariant linear impulse responses, or a bunch of filters with delay, the source is expected to be filtered and translated in time along the RIR. Therefore, supposing a high correlation between the RIR and the pulses emitted by the source (i.e., the direct sound), with due consideration to the filtering of the room, MP appears to be well suited for this purpose. Matching Pursuit (Section 3.1) decomposes any signal into a linear expansion of waveforms that belong to a dictionary. These waveforms are selected iteratively in order to best match the signal structures. Although Matching Pursuit is non-linear [7], it maintains an energy conservation which guarantees its convergence. In practice the number of iterations must be finite, leading to only an approximate decomposition of the signal (Section 3.2). In these experiments, the waveforms (or atoms) that constitute the dictionary is limited to the direct sound itself, translated in time. The determination of its exact temporal boundaries is of importance for a perfect match with the RIR, but is achieved within MP itself, as exposed in Section 3.3 and explored in [8].

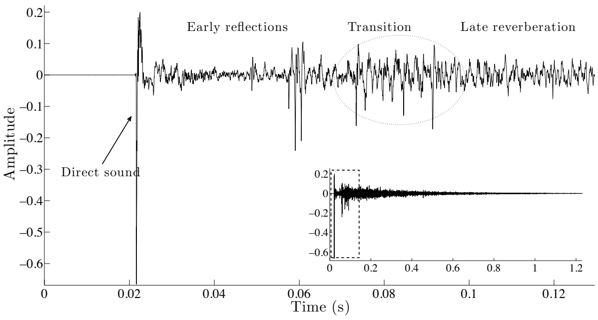

Figure 1: Early times of an experimental RIR (encapsuled graphic), measured in Salle Pleyel.

Matching Pursuit run on RIR provides a linear set of coefficients which can be seen as a temporal distribution of arrivals of the RIR, providing the knowledge of the cumulative distribution function (CDF) of arrivals. Section 4 compares the temporal density of arrival to the theory presented in Section 2 and in [9], using a stochastic model of RIRs.
Mixing times are estimated from experimental RIRs of a concert hall and are compared to those obtained with the stochastic model (Section 5). The relationship between mixing time and the distance source/receiver is investigated in both cases. Differences between data from the model and from the measurements are explained in Section 6. The issues of estimating mixing times using MP are discussed in Section 7.

2 Theoretical review of ergodic rooms

Since Weyl's path-breaking paper on eigenfrequency density in bounded rooms [10], it is well known that this density increases with the square of the frequency. However, when experimentally measuring this density in a box albeit with electromagnetic waves, that display much sharper resonances, Schroeder found out in 1954 [11] that it is not the case at high frequencies: the finite width of the modes, created by the losses, leads to a superposition of many modes at any frequency and to a constant density of peaks.

Schroeder carried out further his investigation in order to determine the transition frequency between the two regimes, and introduced what is now called the Schroeder frequency [11], linked to volume and decay time, i.e., to mode width.

Weyl's estimation of eigenfrequency density was based on the approximation of the general solution to the Helmholtz equation in a bounded space by the Green function of free space. Thus, the notion of trajectory, that is, sound travelling in a room, was implicitly introduced in mode-counting, albeit in a crude way as represented by the Green function that links a source position to a receiver position. In Weyl's free space approximation, no reflections on the boundaries were taken into account, but later refinements to his theory [12] took reflections into account.

As a consequence, Weyl's theory and its successors naturally induce researchers to look at the temporal succession of arrivals from trajectories linking the source to the receiver and taking into account reflections at the boundaries. More specifically, the present paper investigates whether a transition time can experimentally be demonstrated between two domains: an initial time domain where many arrivals are discrete; and a late time domain where arrivals are superposed and therefore become statistically equidistributed. Indeed, Krylovs theory of mixing [2] teaches us that most dynamical systems gradually loose memory of their history with time. Thus, trajectories become independent of their origin, so that the probability of reaching any phase point, at any time, along the trajectory becomes equivalent (ergodicity). In such a case, the arrival density at any receiver position becomes constant and independent of the receiver position.

However, at short times, arrival density is not constant, but increases with time. A simple theory, derived for rectangular rooms [13], teaches us that the density $D_e(t)$ (Eq.(1) is proportional to the square of the time elapsed since the sound was emitted by the source (which is different from the time laps between the arrival and the direct sound).

$$D_e(t) = \frac{dN}{dt} = 4\pi c^3 \frac{t^2}{V} \tag{1}$$

where $D_e(t)$ is in number of arrivals per second, N is the number of arrivals, c is the speed of sound in $m.s^{-1}$, and V is the volume of the room in m^3.

Therefore, it should be experimentally possible to determine the transition time as the time after which arrivals are superposed so that their ditribution becomes constant. The present paper concentrates therefore on the experimental determination of this transition time, which we call *mixing time* in accordance with [2], and investigate the robustness of different estimators of it.

In [14], Polack proposes a somewhat different definition of mixing time, base on experiment, linked to the resolution of the auditory system instead of the characteristic duration of arrivals. Mixing time is reached when 10 reflections overlap within the characteristic time resolution of the auditory system, taken equal to $24ms$ [5]. Then Eq. (1) leads to:

$$t_m \approx \sqrt{V} \tag{2}$$

where t_m 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 (Fig.1). It is shown in [9] and [14] that the exponentially decaying stochastic model [15] can be established within the framework of geometrical acoustics and billiard theory. The mixing character of a room depends on its geometry and on the diffusing properties of the boundaries of 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 [9] [16].

3 Matching Pursuit applied to RIR

A RIR can be seen as a linear combination of occurence of the direct sound reproduced in time, and filtered by reflections on the surfaces of the hall. Figure 2 shows a first reflection which is similar to the direct sound, up to the filtering of the surfaces of the hall. For this latter reason, it is believed that a technique based on a correlation between the RIR and the direct sound is well indicated for detecting arrivals into the signal, as presented in the Introduction.

3.1 Theoretical reviews

Matching Pursuit can help understanding more deeply the architecture of a RIR, since this algorithm introduced by [7] provides information, which can be seen as maxima of correlation (Eq.4) [17] between two signals: the RIR (x) and the direct sound (the atom).

Matching Pursuit works as follows:

1. Initialization: $m = 0$, $x_m = x_0 = x$

2. Computation the correlations between the signal x_m and every atom γ of a dictionary ϕ, using inner products:

   $$\forall \gamma \in \phi \quad : \quad CORR(x_m, \gamma) = |\langle x_m, \gamma \rangle| \tag{3}$$

   The dictionary ϕ is a set of atoms γ, of the same length than x, constituted by the direct sound and translated in time, by step of one sample.

3. Search the most correlated atom, by searching for the maximum inner product:

   $$\tilde{\gamma}_m \quad = \quad argmax(CORR(x_m, \gamma))_{\gamma \in \phi} \tag{4}$$

4. Substracting the corresponding weighted atom $\alpha_m \tilde{\gamma}_m$ from the signal x_m:

   $$x_{m+1} \quad = \quad x_m - \alpha_m \tilde{\gamma}_m \tag{5}$$

   $$x_R^{(m)} \quad = \quad \sum_{k \leq m} \alpha_k . \tilde{\gamma}_k \tag{6}$$

   where $\alpha_m = \langle x_m, \tilde{\gamma}_m \rangle$;

5.
   - stops if the desired level of accuracy is reached: $R = x_{m+1}$.
   - otherwise, re-iterate the pursuit to step 2: $m \leftarrow m + 1$.

where x is the RIR, R the residual, γ the atom (here, the direct sound), ϕ the dictionary of atoms γ, and $x_R^{(m)}$ the reconstructed signal.

In theory, any signal x can be perfectly decomposed in a set of atoms for an infinity of iterations. In practice, this number must be finite and a stopping criterium has to be set. The authors propose to use the signal/residual ratio (SRR) in dB of the norm of x over the residual (R).

3.2 Stopping criterium -Finding an appropriate value

The quality of the decomposition of x in atoms depends on the value of SRR, that is, the stopping criterium. On the one hand, for a too low SRR, the residual has an energy level too high and the rebuilt signal $x_R^{(m)}$ is an impoverished approximation of x. On the other hand, a too high SRR leads to a high number of iterations. In that case, it is not necessary to perform more iterations beyond a certain threshold of accuracy.
Room acoustics quality is evaluated using a set of objective parameters. Among all these acoustical indices proposed in standards [18], we refer to only four, selected because of their wide use in room acoustics. Let $x(t)$ be the impulse response of the hall.

- Reverberation time (RT) mesures the energy decay, and is probably the most widely used index. It is measured using the Schroeder integrated impulse response technique [19], and linear regression between -5 and $-25dB$, and between -5 and $-35dB$, for RT_{20} and RT_{30} respectively.

- Early Decay Time (EDT_{10}), proposed by Jordan [20], is measured by the same method as RT, between 0 and $10dB$.

- Central Time, proposed by [5] and [21], expressed in seconds, is a measure of the centre of gravity of the impulse response energy:

$$T_C = \frac{\int_0^\infty t.x^2(t)dt}{\int_0^\infty x^2(t)dt} \qquad (7)$$

Comparing previous acoustical indices calculated on x to those calculated on $x_R^{(m)}$ for different values of a SRR allows to set the stopping criterium on a physical background. Figure 3 shows the variations in percent between indices calculated on x and those calculated on $x_R^{(m)}$ for different values of SRR, for an impulse response, in which the visual identification of the direct sound is obvious (Figure 2). This RIR presents the particularity that the direct sound (pistol shot) is immediatly followed by the first reflection. This way, the identification of temporal boundaries of the direct sound; in order to consitute the dictionary ϕ for that particular RIR, is facilitated.
According to [22], acceptable variations of acoustical indices are 5% and below. Thus, Figure 3 indicates that a convenient SRR would be $20dB$. Reasons of variations of acoustical indices can depend upon several factors [22]: the lack of reproducibility of the sound source, the microphone directivity and positioning, software parameters, and also the estimation of the onset of the RIR [23]. Figure 4 shows a RIR and the linear set of coefficients.

3.3 Detection of the direct sound

The determination of the exact temporal boundaries of the direct sound is of importance for a perfect match, as seen in the Introduction. Moreover, the knowledge of the direct sound provides useful information on the sound source itself, and allows to whiten the RIR. This study, by presenting a method for detecting the direct sound, contributes to the characterization of some frequently used

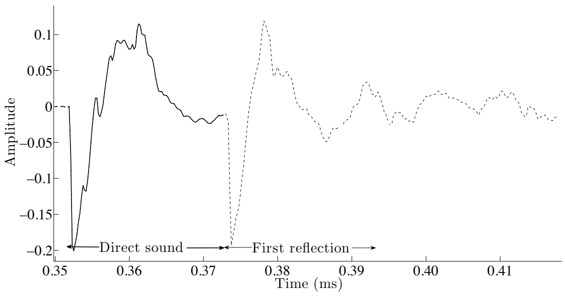

Figure 2: Experimental RIR for which determining the direct sound is obvious. Plain line: direct sound.

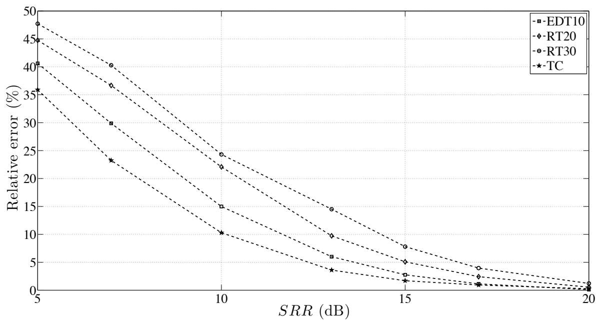

Figure 3: Variations in % of EDT_{10}, RT_{20}, RT_{30} and T_C versus SRR in dB.

sound sources in room acoustics measurements [8] [24] [25], such as balloon bursts or pistol shots. For RIR measurements carried out with these latter sources, it becomes difficult to clearly identify temporal boundaries when the sound source is not recorded in the near field, or in an anechoic chamber. Moreover, the visual identification of these boundaries may vary from expert to expert [26] (Fig.5).
The impulse duration is assumed to be inferior to $5ms$, defining the atom γ. The time index t_0 f the maximum of the atom is detected. A dictionary of atoms ϕ is constituted of atoms with temporal boundaries that are varying (with step of $0.1ms$) from $0ms$ to t_0 and from t_0 to $5.0ms$, for the first and last indices respectively. For each couple of [first index:last index], MP is ran. The temopral boundaries that are thought to be the best correspond to the lowest number of iterations. Impulses durations are estimated running MP onto experimental RIRs, with a stopping criteria of $SRR = 20dB$ to reach. For further details, please refer to [8].

4 Detection of arrivals

4.1 Validity of Matching Pursuit on a model of RIRs

This section aims at testing Matching Pursuit on a given reference set of arrivals synthesized by a model of RIR presented and detailed in [9]. This model has been validated on a set of 8 americain concert halls for which a set of room acoustical indices were recovered [27]. In this model, arrivals are distributed in time according to a Poisson process, with a parameter which is time dependent. The cumulative number of arrivals(CNA) is a cubic function of time (Fig.6). Using input parameters, such as the reverberation time (RT_{30}), the mean absorption ($\bar{\alpha}$), and

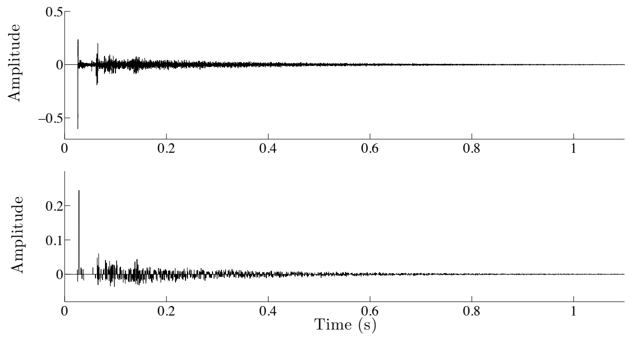

Figure 4: Matching Pursuit ran on an experimental RIR ($SRR = 20dB$). -(top): Experimental RIR -(bottom): Linear set of coefficients, which correspond to the arrivals.

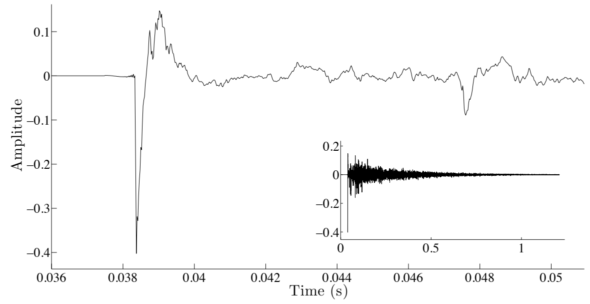

Figure 5: Detail of an experimental RIR for which determining the direct sound is not obvious.

the volume (V) of the hall, the set of time arrivals, and their respective amplitude, are generated. Room impulse responses of a considered hall (Salle Pleyel: $V = 19000m^3$, $RT_{30} = 1.9s$, $\bar{\alpha} = 30\%$) are synthesized and convolved by a pistol shot. Then, the linear set of arrivals is estimated by MP.

4.2 Towards a validation of the statistical theory

The presented results are derived from RIRs measured in Salle Pleyel carried out with pistol shots [28], for 21 different source-receiver positions. Therefore 21 experimental RIRs are under consideration. Based on section 4.1, the linear set of coefficients (Fig.4), derived from Matching Pursuit run on experimental RIRs, are assumed to represent the temporal distribution of arrivals.

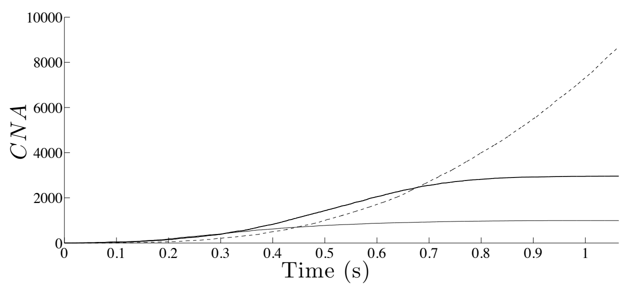

Figure 6: Cumulative numbers of the arrivals (CNA) generated by the stochastical model of RIR (without compensation of energy decay). -dashed: model -plain: the atom is a pistol shot -bold: the atom is a dirac.

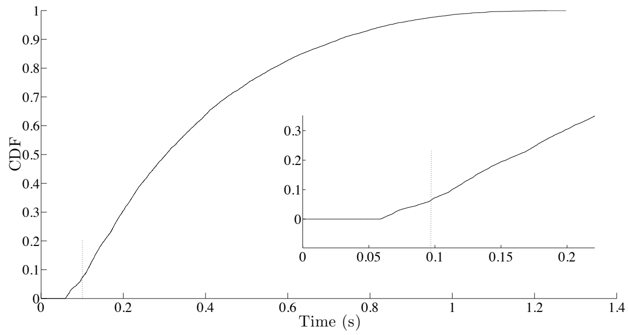

Figure 7: CDF, for an experimental RIR (without compensation of the energy decay). Dashed line: a kink in the distribution of arrivals is noticeable. The encapsuled grpahic is a zoom on the early times of the CDF.

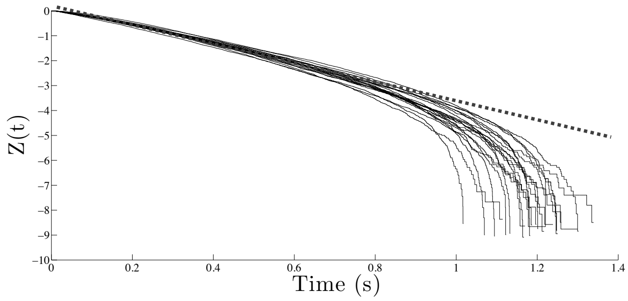

Figure 8: plain: Logarithm of $(1 - CDF)$ of the arrivals of 21 experimental RIRs (without compensation of the energy decay) -dashed: mean logarithmic decay of the RIRs.

4.2.1 Without compensating for the energy decay

The CNA normalized by the total number of arrivals represents an estimate of the Cumulative Distribution Function (CDF), plotted in Fig. 7. In other words, the CDF describes the time evolution of the probability to detect arrivals in the RIR.

Figure 7 underlines the decreasing of probability to detect arrivals at the end of the RIR. Mallat et al. [7] states that Matching Pursuit is not suitable for non-stationary signals. Indeed, as MP selects the maximum of correlation at each iteration, it becomes obvious that it has a high probability to be found at the beginning of the RIR (Fig.4). Thus, the probability to detect arrivals is directly linked to the local energy of the signal. As this latter decreases exponentially, one can expect the probability to decrease exponentially too.
By calculating the logarithm of $1 - CDF$ for the 21 RIRs (Fig.8), the mean reverberation decay of the room ($RT \approx 2.0s$) is recovered in comparison to what has been calculated in [28]. This observation highlights that the CDF is linked to the energy of the signal.

4.2.2 Compensating for the energy decay

As seen in section 4.2.1, arrivals have a higher probability to be found in the beginning of the RIR, than in the tail. Energy compensation, by making the signal stationary and ergodic, ensures equal weight to all parts of the RIR and thus equiprobability of detecting arrivals. This is observed in Figure 9, where it corresponds to the almost constant slope of the CDF. Note that the energy com-

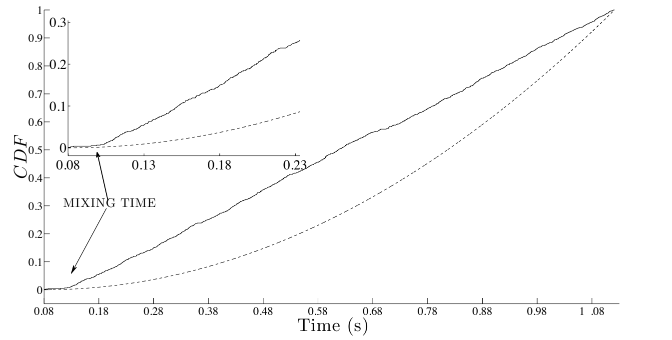

Figure 9: $CDFs$. -dash: theoretical CDF ($\approx t^3$) -plain: Average of CDF for 21 RIRs (with compensation of the energy decay). The encapsuled graphic is a detail of the early part of the CDF.

pensation is achieved by applying an inverse exponential, whose argument is proportional to the reverberation time and to the mean absorption [5].

However, the beginning of the RIR presents a different behaviour, in agreement with theory, which predicts a lower number of arrivals after the direct sound than for the diffuse sound field. This difference of behaviour allows to define the mixing time (T_M) as the time where this difference occurs. Indeed, mixing precisely expresses the equiprobability of arrivals, as defined by Krylov [2]. The mixing time is then defined as the time at which the process becomes ergodic, taking into account the time propagation from the source position to the receiver position.

Moreover, the estimation of arrivals, and thus of mixing time, depends on the atom, that is, on the temporal and spectral properties of the direct sound (Fig.10), as it is explained in Section 5.

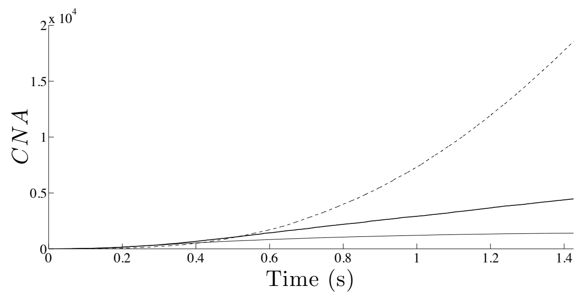

Figure 10: CNA of a theoretical RIR. -dash: theoretical CNA -plain: estimated CNA (the atom is a pistol shot) -bold: estimated CNA (the atom is a dirac) (with compensation of the energy decay).

Matching Pursuit is ran on 21 RIRs measured in Salle Pleyel [28], compensating the energy decay. $CDFs$ all show a breaking point (Fig.9) which divides the curve into two parts:

- $t < T_M$: a few arrivals are detected after the direct sound. A cubic fit (Fig.11) of this part of the curve permits to verify that the number of arrivals is a function of t^3, as seen in Eq. (1). The goodness of fit is attested by the correlation coefficient $\bar{r} > 0.90$.
- $t \geq T_M$: the number of arrivals increases at a constant rate with time. Furthermore, the evolution of the number of arrivals as a function of time is almost constant all over the room (Fig.12).

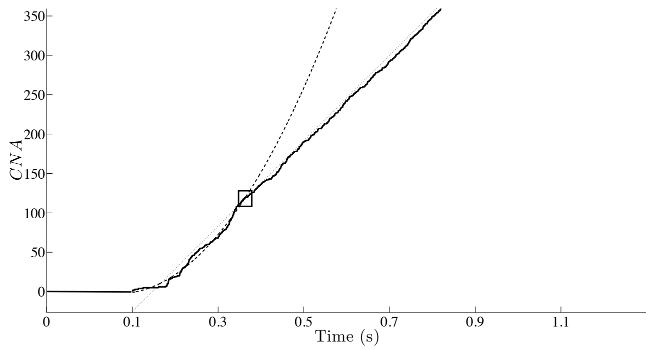

Figure 11: Detail of a cubic ($t < T_M$) and a linear ($t > T_M$) fits made on a CNA. Plain bold line: CNA. Dashed bold line: cubic fit. Dashed line: linear fit. Square: mixing time.

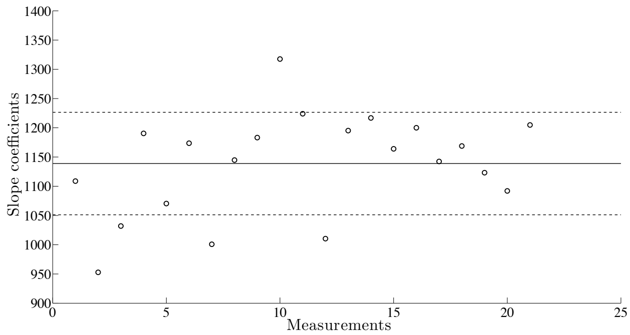

Figure 12: Circles: Slope coefficients of the linear fits made on the 21 $CNAs$ ($mean$ = 1140 arrivals/second, std = 7.7%). Plain line: mean slope coefficients. Dashed lines: standard deviation.

5 Estimation of the mixing time

According to the theory of room acoustics, for $t \geq T_M$ arrivals overlap. We observe that MP detects only one arrival instead of two when the time delay between them is inferior or equal to the equivalent duration $\tilde{d}$ of the impulse (i.e. the direct sound). The equivalent duration is related to the equivalent statisitical bandwidth of the impulse, defined by [29].

The satistical time is then defined as the time at which two successive arrivals are delayed one from the other of the equivalent duration of the impulse $\Delta t \leq \tilde{d}$. Figures 13-14 show statisitcal times detected in the set of arrivals estimated by MP from an experimental and a synthesized RIR, respectively.

Statistical times do mark the beginning of the linear part of the $CNAs$ (Fig.11-14), where the RIR can be defined with a statistical behavior, rather than with a deterministic behavior. In that sense, statistical times are then related to Krylov's definition of mixing time (Section 4.2.2). Consequently, in the following, the estimation of the mixing time is carried out by detecting the statistical time of RIRs from the set of arrivals.

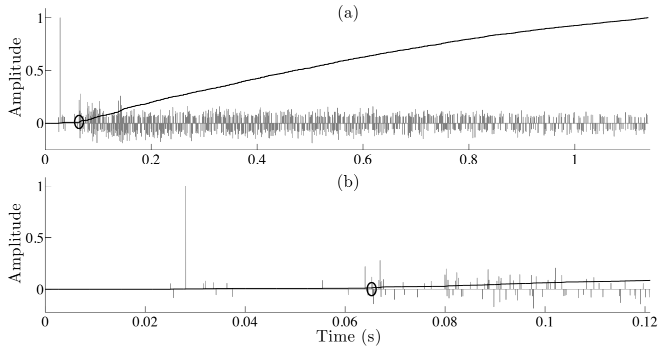

Figure 13: a)): Mixing time (circle) detected on a CDF (bold line) for an experimental RIR and the set of estimated arrivals. ($SRR = 5dB$). Plot (b) details the beginning of plot (a).

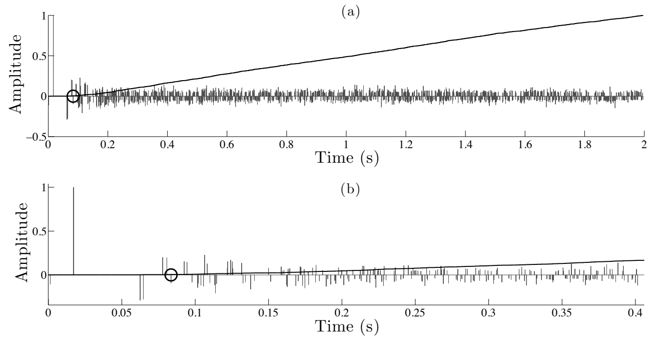

Figure 14: a)): Mixing time (circle) detected on a CDF (bold line) for a synthesized RIR convolved with a pisol shot and the set of estimated arrivals ($SRR = 5dB$). Plot (b) details the beginning of plot (a).

5.1 Dependence upon the stopping criterium value (SRR)

The energy compensation leads to set the SRR differently. First, as the exponential decay is compensated, the decrease of energy along the signal is approximated more accuratly by MP than without compensation (see Section 4.2.2). As a consequence, reverberation times and energetical indices of the synthesized signal ($x_R^{(m)}$) are expected to be close to the original signal for a smaller SRR than in Section 3.2. Second, as studied in [30] some selected atoms may exist in some regions where the original signal does not possess any energy. These terms are part of the estimation of the original signal by MP and interfere with each other, creating before and after arrivals small coefficients for which the interpretation is not obvious. This point is discussed in Section 6. Therefore, one may expect a strong dependency of the accuracy of the estimation of the mixing time upon the SRR, since the bigger SRR the larger the number of arrivals is.

Figure 15 shows mixing times estimated from 21 experimental RIRs for different values of SRR. On the one hand, for a too low SRR ($SRR < 4dB$), mixing times do not always exist, and present a large spreading. On the other hand, mixing times present smaller variations for $4 \leq SRR < 5dB$. Futhermore, one may notice that mixing times are consistent for $5 \leq SRR < 7dB$. Finally, mixing times are close to zero for $7 \leq SRR < 10dB$ and null for $SRR \geq 10dB$.

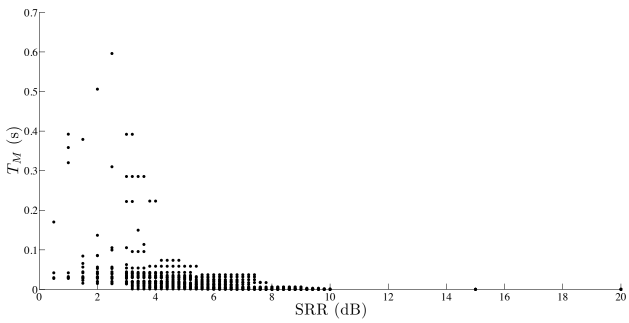

Figure 15: Mixing times estimated on experimental RIRs using different values of SRR (dB) (compensating the energy decay).

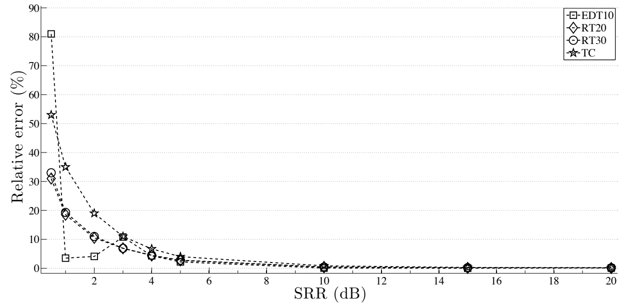

Figure 16: Variations in % of EDT_{10}, RT_{20}, RT_{30}, T_C versus the SRR in dB (compensating the energy decay).

The choice of a SRR is guided by the mean variations of the usual acoustical indices, which are calculated between the original signal x and the synthesized one $x_R^{(m)}$, as in Section 3.2, for the 21 RIRs. From these results (Fig.16), and according to [22], an acceptable SRR is of $5dB$, since variations lie below 5%.

In the following, mixing times of experimental and synthesized RIRs (from the stochastical model presented in Section 4.1) are estimated by detecting statistical times for a $SRR = 5dB$. The relationship between mixing time and the distance source/receiver is investigated in each case.

5.2 Experimental RIRs

For each of the 21 experimental RIRs, the equivalent duration of the impulse ($\tilde{d}$) is calculated, according to [29], since the direct sound is estimated by learning the dictionary of atoms (Section 3.3). Mixing times are estimated from the sets of arrivals (Fig.13), as described in Section 5.1. Results are given as a function of the distance with and without the time propagation between the source and the receiver positions taken into account (Fig.17).

The mean value of estimated mixing times (with time propagation) is about $85ms$ (the median value is $92ms$), while the standard deviation is of 30%. Large variations of mixing times are in contradiction with the theory of ergodic rooms, which predicts that the mixing occurs at the same time in the whole room. Surprisingly, mixing time can rather be well described by an increasing linear function of the distance source/receiver (Fig.18), which is not predicted by the theory (Eq.2). The relationship is given by:

$$T_M = 0.0026.d + 0.026 \quad (8)$$
$$r = 0.82 \quad (9)$$

where T_M is the mixing time in second, d is the distance source/receiver in meter, and r is the correlation coefficient.

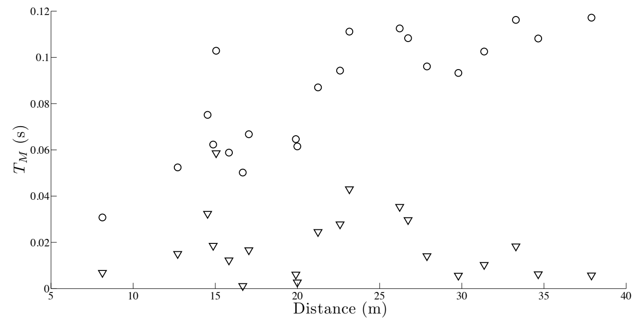

Figure 17: Estimated mixing times as a function of distance (without (triangles) and with (circles) time propagation between the source and the receiver).

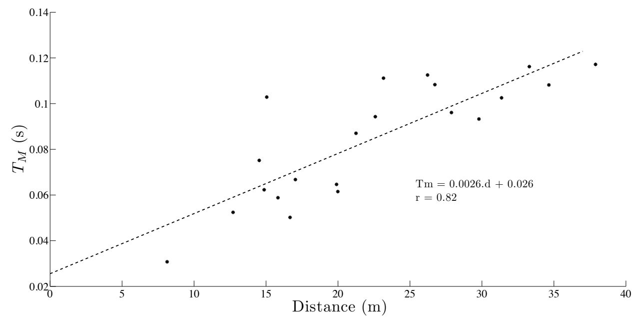

Figure 18: Estimated mixing times (dots) as a function of distance (with time propagation between the source and the receiver). Dashed line: linear regression.

5.3 Synthesized RIRs

This section aims at estimating statisitical times on the set of arrivals, obtained with MP, from RIRs synthesized with the model presented in Section 4.1. This is achieved by synthesizing a large number of times RIRs for the same input parameters. This is inspired by the Monte Carlo methods, oftenly used when simulating physical and mathematical systems [31].
In practice, twenty RIRs are synthesized (in one dimension) at 23 different distances (from $1m$ to $45m$, by step of $2m$). Therefore, a total of 460 RIRs is under consideration. Parameters of the model are still those of Salle Pleyel (Section 4.1). MP is run on the model of RIRs convolved by a pistol shot (compensating the energy decay, and using $SRR = 5dB$). Estimated arrivals are identical to those of the model until a certain time, the statistical time. Figure 14 shows the statistical time detected into a set of coefficients of MP. Figure 19 shows statistical times estimated for each synthesized RIR, taking into account the distance source/receiver.
On the one hand, considering the mean statistical times, one may notice that they are approximately constant for $d < 25m$ (the average equals $101ms$; the standard deviation equals 20%) and are an increasing linear function of distance, for $d \geq 25m$, according to:

$$T_S = 0.0017.d + 0.057 \quad (10)$$
$$r = 0.97 \quad (11)$$

where T_S is the statistical time in second, d is the distance source/receiver in meter, and r is the correlation coefficient.

On the other hand, considering the minima statistical times, it is noticeable that they are approximately constant for $d < 19m$ and are an increasing linear function of distance for $d \geq 19m$. The relationship between the statistical time of the model and the distance source/receiver is given by:

$$T_S = 0.0029.d \quad (12)$$
$$r = 0.98 \quad (13)$$

where T_S is the statistical time in second, d is the distance source/receiver in meter, and r is the correlation coefficient.

Furthermore, if $a = 0.0029$ is the slope (Eq.12), then $1/a = 1/0.0029 = 340m.s^{-1} \approx c_0$. Hence, statistical times and distance are clearly linked. This is obvious, since the mixing can only occur once the direct sound has reached the receiver position.

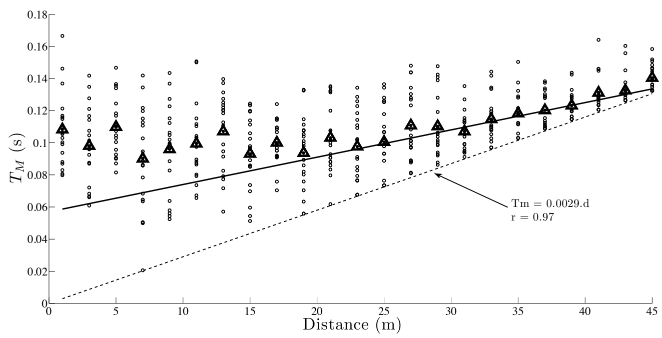

Figure 19: Mixing times (dots) detected for 20 RIRs synthesized at different distances (the time of propagation is taken into account) ($SRR = 5dB$). Triangles: mean times. Dashed line: linear fit on the minima mixing times (for $d \geq 19m$). Plain line: linear fit on the mean mixing times (for $d \geq 25m$).

6 Difference between data from the experiment and the model

In Salle Pleyel, the volume is $V = 19000m^3$ and the theoretical value of the mixing time equals $T_{M_{th}} = 137ms$, according to Eq.(2.) Mean mixing times estimated from experimental RIRs ($\tilde{T}_M = 85ms$) on the one hand, and from synthesized RIRs ($\tilde{T}_S = 101ms$) on the other hand, recover the theoretical assumption that states that $T_{M_{th}} \approx \sqrt{V}(ms)$ is an upper value of mixing time (Section 2).

If the mean mixing times of experimental and synthesized RIRs are not equal strictly, one may notice that the relationships that link the mixing time to the distance are very similar (Fig.20-21). Actually, the diffusion phenomenum (which is related to mixing -see Section 2) is assumed to explain the difference between the model and the experiments (Fig.21).
On the one hand, as seen previously in Section 5, mixing time is a function of distance, as is the Initial Time Delay Gap (ITDG: time delay between the direct sound and the first arrival of the RIR) [4]. The ITDG is a decreasing linear function of distance, that is, for short distances, the first arrival occurs later after the direct sound than for long distances. Diffusion occurs after the first reflection, which can in practice arrive at a much shorter time, after the direct sound than the mean statistical delay of the model. Remember that all measurements are taken near a boundary, a particular situation that may significantly differ from the statistical average of the model.

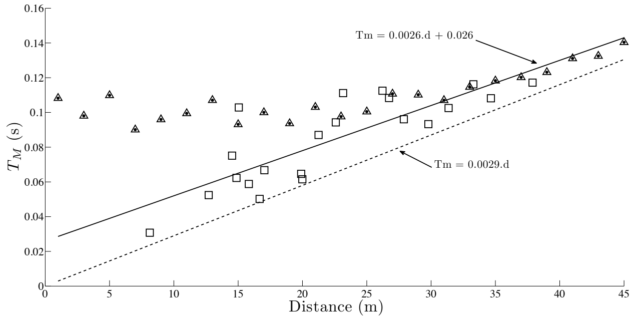

Figure 20: Mean mixing times of the model (triangles) and of the experiments (squares). Dashed line: linear regression made on the model. Plain line: linear regression of experimental RIRs.

On the other hand, the number of arrivals estimated by MP is a function of the SRR (Section 5.1). As a consequence, MP estimates arrivals and also diffusion in experimental RIRs. Sturm et al. present, in [30], an artefact of MP: constructive and destructive interferences. When MP finds a correlation between an atom and the signal, it substracts the contribution of this atom from the signal. But in the case of experimental RIRs, the mother atom (the direct sound) can only accuratly be found at the beginning of the RIR and not after, since the RIR is a succession of delayed and filtered versions of the mother atom. Hence, MP, by substracting the contribution of the mother atom to the signal, creates residuals -or interferences- that may be compensated (by finding some other correlations) in further iterations. In other words, MP creates more coefficients than arrivals.
As the model does not take into account either diffusion or the filtering of the walls of the room, the number of arrivals is larger in the experiment than in the model. Consequently, the probability that 2 successive arrivals are separated by $\Delta t \leq \tilde{d}$ is larger for short times with experimental RIRs than with synthesized RIRs. This is why experimental mixing times are smaller than those of the model for small distances, specially for $d < 27m$, which is the mean free path of Salle Pleyel.
Differences between the model and the experiment correspond to the hatched area of the Figure 21.

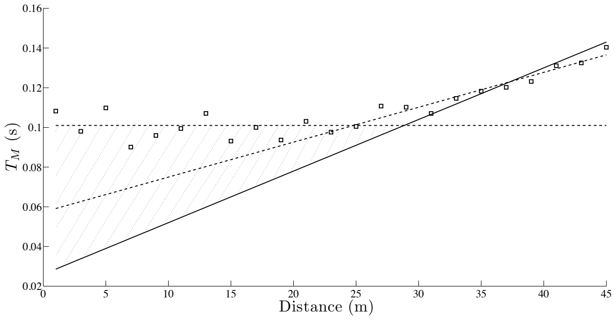

Figure 21: Mean mixing times from the model (triangles) and from the experiments (squares). Dashed lines: linear regressions made on the model (for $d \geq 25m$). Plain line: linear regression of experimental RIRs. Hatched part: missing contribution of diffusion and room filtering in the model.

7 Discussion

Studying the distribution of arrivals is not a new idea in room acoustics. As early as 1958, Schodder [32] published an extensive survey of more than 1000 RIRs measured in more than 15 different halls. Schodder study presents two shortcomings that impair direct comparison: his photographic techniques did not allow for decay compensation; he did not include time propagation between the source and the receiver. As a consequence, his mean distribution of arrival times cannot directly be compared to our results. Yet, on some of them, a change of slope on the distribution of arrival times can be observed, similar to what can be observed around $50ms$ in Figure 7.

As seen previously, the stopping criterium is important, since its value specifies the approximation of the original signal and the accuracy of the estimation of the mixing time. However, the interference issue described in the last section raises the question of the physical interpretation of small coefficients. Further, future work should investigate, on other concert halls, the relevance of setting the $SRR = 5dB$.
Furthermore, an analysis on octave bands is thought to be an original manner to discriminate the phenomenum of diffusion at high frequencies, on the one hand, and also to give information about the filtering of the room, on the other hand.
Finally, first tests show that the mixing time depends on the atom, that is, on the temporal and spectral properties of the direct sound (Fig.10). But this goes beyond the scope of the present paper.

The good agreement between mixing times of the model and of the experiment is in favor of this mixing time estimator based on MP, and of the model, which has already been through a first validation based on comparisons of acoustical indices between real and synthesized RIRs [27]. However, a future work would consist in adding diffusion to the model and in comparing the obtained mixing times to the experimental ones.

8 Conclusions

This study uses a well documented technique, Matching Pursuit, to determine the time of arrivals in RIRs. This leads to first set an appropriate stopping criteria, and second to define as precisely as possible the temporal boundaries of the direct sound, which is used as the mother atom of the dictionary. The stopping criterium is choosen by minimizing the difference between the acoustical indices of the original RIR and the synthesized one. Further studies should evaluate, using listening tests, the relevance of such a stopping criteria, but also by estimating the mixing time of other concert halls. This would lead to choose either another value, or another stopping criteria.
Temporal boundaries of the direct sound are estimated by looking at the speed of convergence of Matching Pursuit. In other words, the lowest the number of iterations, the best are the temporal boundaries of the direct sound. This seems to be an efficient method to characterize frequently used sound sources in Room Acoustics, and has been used in [8].
Matching Pursuit provides another vision of RIRs. Indeed, the linear set of coefficients obtained are seen as the arrivals of the RIR. The exponential decrease of energy necessitates a compensation, in order to obtain a stationary signal. The mixing time is then defined as the time at which the signal becomes stationary. The number of arrivals is a cubic function of time, before the mixing time. After the mixing time, the number of arrivals grows with a constant rate in the whole room. However, mixing times are found to be a function of the distance source/receiver. The stochastical model used in this paper confirms this latter point. This constitutes a hint in the favor of the robustness of the model, which needs to be supplemented by integrating diffusion. The relationship between mixing times and distance should be investigated in a future work that would consist in measuring RIRs in a mixing hall, with very close receiver positions.
It remains to generalize this estimator to other rooms, using some different atoms. Moreover, more investigation should be made with filtering the RIR and using threshold on the linear set of coefficients, derived from Matching Pursuit.

9 Acknowledgments

This work was partly supported by grants from Région Ile-de-France, France. The LAM team of IJLRDA is partly supported by the French Ministère de la Culture et de la Communication.

References

[1] W. B. Joyce, "Sabine's reverberation time and ergodic auditoriums," *J. Acoust. Soc. Am.*, vol. 58, no. 3, pp. 643–655, 1975.

[2] N. S. Krylov (translated by A.B. Migdal, Ya. G. Sinai, and Yu. L. Zeeman), *Works on Foundations of Statistical Physics*, Princeton University Press, 1979.

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

[4] L. Beranek, *Concert and Opera Halls, How They Sound ?*, Acoust. Soc. Am., 1996.

[5] L. Cremer and H. Müller, *Principles and Applications of Room Acoustics*, vol. 1, pp. 472, Applied Science Publishers Ltd, 1982.

[6] 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*, 2007, pp. 59–62.

[7] S. Mallat and Z. Zhang, "Matching pursuit with time-frequency dictionnaries," *IEEE Trans. Signal Process.*, vol. 40, no. 12, pp. 3397 3415, 1993.

[8] G. Defrance, L. Daudet, and Polack. J-D, "Characterizing sound sources for room acoustical measurements," in *Proceedings of the Institute of Acoustics*, 2008, vol. 30.

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

[10] H. Weyl, "Das asymptotische Verteilungsgesetz der Eigenwerte linearer partieller Differentialgleichungen," *Math. Ann.*, vol. 71, pp. 441–479, 1912.

[11] M.R. Schroeder, "Statistical parameters of the frequency response curves of large rooms," *J. Acoust. Eng. Soc.*, vol. 35, no. 5, pp. 307–316, 1987.

[12] R. Balian and C. Bloch, "Distribution of eigenfrequencies for the wave equation in a finite domain: I. Three-dimensional problem with smooth boundary surfaces," *Annn. Phys.*, vol. 60, pp. 401–447, 1970.

[13] R. H. Bolt, P. E. Doak, and P. J. Westervelt, "Pulse Statistics Analysis of Room Acoustics," *J. Acoust. Soc. Am.*, vol. 22, no. 3, pp. 328–340, 1950.

[14] 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, 1988.

[15] A. Baskind and J-D. Polack, "Sound Power Radiated by Sources in Diffuse Sound Field," in *Proceedings of the 108th AES Convention*, 2000, Paris.

[16] 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*, 1997.

[17] S. Krstulovic and R. Gribonval, "Mptk: Matching pursuit made tractable," in *ICASSP'06*, Toulouse, France, 2006.

[18] ISO 3382, *Acoustics-measurements of the reverberation time of rooms with reference to other acoustical parameters*, 1997.

[19] M. R. Schroeder, "New method of measuring reverberation time," *J. Acoust. Soc. Am.*, vol. 37, pp. 409–412, 1965.

[20] V. L. Jordan, "Room acoustics and architectural development in recent years," *App. Acoustics*, vol. 2, pp. 59–81, 1969.

[21] R. Kürer, "Zur Gewinnung von Einzahlkriterien bei Impulsmessung in der Raumakustik," *Acustica*, vol. 21, pp. 370–372, 1969.

[22] X. Meynial, Polack. J-D, and G. Dodd, "Comparison between full-scale and 1:50 scale model measurements in Théâtre Municipal, Le Mans," *Acta Acustica*, vol. 1, pp. 199–212, 1993.

[23] G. Defrance, L. Daudet, and J-D. Polack, "Finding the onset of a room impulse response: straightforward?," *J. Acoust. Soc. Am.*, vol. 124, no. EL248, 2008.

[24] D. Griesinger, "Beyond MLS- Occupied hall measurement with FFT techniques," 2008 March, http://world.std.com/ griesngr/sweep.pdf.

[25] P. Fausti and A. Farina, "Acoustic measurements in opera houses: Comparison between different techniques and equipment," *J. Sound and Vibration*, vol. 232, no. 1, pp. 213–229, 2000.

[26] G. Leveau, L. Daudet, and G. Richard, "Methodology and Tools for the evaluation of automatic onset detection algorithms in musicluation of automatic onset detection algorithms in music," in *Proceedings of the International Symposium on Music Information Retrieval*. ISMIR'04, 2004.

[27] J-D. Polack, "Reverberation time and mean absorption in concert halls," in *Proceedings of the Institute of Acoustics*, 2006, vol. 28, p. 2.

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

[29] J. S. Bendat and A. G. Piersol, *Random Data: Analysis and Measurements Procedures*, New York: Wiley, 1971.

[30] B.L. Sturm, J.J. Shynk, L. Daudet, and C. Roads, "Dark energy in Sparse Atomic Estimations," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 16, no. 3, pp. 671–676, 2008.

[31] W. T. Vetterling W. H. Press, S. A. Teukolsky and B. P. Flannery, *Numerical Recipes in C*, Cambridge University Press, second edition, 1992.

[32] G.R. Schodder, "Uber Die Verteilung Der Energiereicheren Schallruckwurfe In Salen," *Acustica*, vol. 6, pp. 445–465, 1956.