Detecting arrivals within Room Impulse Responses using Matching Pursuit

G. Defrance, L. Daudet and J-D Polack *defrance@lam.jussieu.fr*
UPMC Univ. Paris 06 / CNRS / Ministère de la Culture et de la Communication

Institut Jean Le Rond d'Alemebrt, LAM team, 11 rue de Lourmel, 75015, Paris, FRANCE

DAFx08, 1-4TH SEPTEMBER 2008, ESPOO, FINLAND

Statistical room acoustics

Room Impulse Responses (RIRs) are composed of :

1. the direct sound
2. arrivals (or early reflections -set of discrete events) [ER]
3. the late reverberation (or diffuse sound field) [DSF]

An arrival = a sound ray emitted by the source which reaches the receiver after having undergone one or more reflections on the boundaries of the room.

The echo density (average number of arrivals per second) : $D_e(t) = 4\pi c_o^3 \frac{t^2}{V}$

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

For larger times $\longrightarrow$ echo density become large

$\Rightarrow$ Foundation of statistical models of room responses

Polack defines the mixing time as 10 reflections that overlap within a characteristic time resolution of the auditory system ($24ms$).

$$t_{mix} \approx \sqrt{V} \text{ (ms)}$$

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

Mixing time :

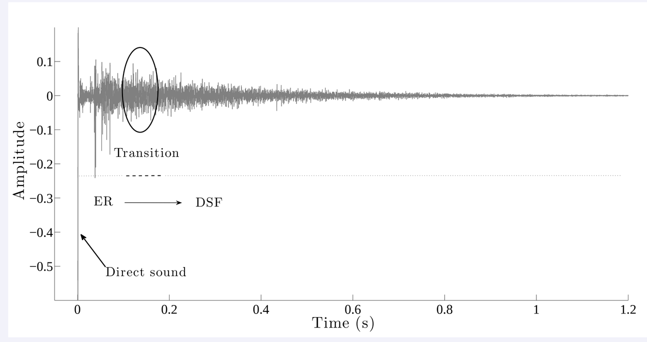
$\Rightarrow$ Transition time between early reflections and late reverberation. It is shown that the exponentially decaying stochastic model can be established within the framework of geometrical acoustics and billiard theory.

$\Rightarrow$ Defined as the time it takes for initially adjacent sound rays to spread uniformly across the room.

$t < t_{mix}$: arrivals can be predicted by the image-source theory.

$t > t_{mix}$: process has become diffuse $\Rightarrow$ 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 of the hall.



Matching Pursuit

A RIR can be seen as a linear set of occurence of the direct sound translated in time, and filtered by the surfaces of the hall. Assuming a high correlation between the RIR and the direct sound, with due consideration to the filtering of the room, Matching Pursuit (MP) can help for understanding more deeply the architecture of a RIR.

Algorithm

Input : $x \in \mathbb{R}^N, \mathcal{D} = \{\phi_\gamma, \gamma\}$

Output : $\gamma_{opt}^k, \alpha^k, k = 1 \ldots (n-1)$

$n = 0$

$R^0 x \leftarrow x$

Repeat (until a stopping criterium is reached) :

$\gamma_{opt}^{(n)} \leftarrow argmax_{\gamma \in \Gamma} |\langle R^n x, \phi_\gamma \rangle|$

$\alpha^{(n)} \leftarrow |\langle R^n x, \phi_{\gamma_{opt}^{(n)}} \rangle|$

$R^{n+1} x \leftarrow R^n x - \alpha^{(n)} \phi_{\gamma_{opt}^{(n)}}$

$n \leftarrow n+1$

$x^{(n)} = \sum_{k=0}^{n-1} \alpha^{(k)} \phi_{\gamma_{opt}^{(n)}}$

with x being the original RIR, R the residual, $x^{(n)}$ the synthesized signal, $\langle . \rangle$ the scalar product, and ϕ_γ the dictionary of atoms γ.

MP is run on 21 experimental RIRs (Salle Pleyel -pistol shots).

Dictionary

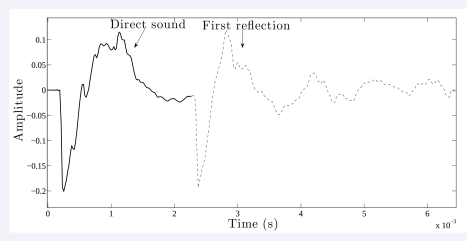
The dictionary is built of the direct sound (mother atom) translated in time.

An efficient decomposition of a signal x in a linear set of atoms can only be achieved if the mother atom has an exact temporal and spectral definition. Looking for the lowest number of iterations to reach the stopping criterium is equivalent to looking for the best temporal boundaries of the direct sound. This is achieved by using a grid search method on the first and the last indices of the atom.

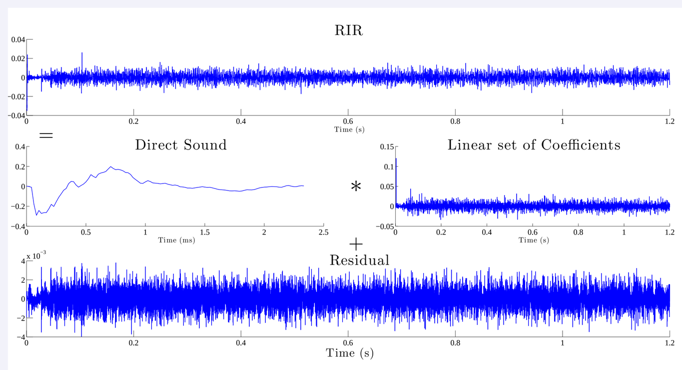
Stopping criterium

In theory, any signal x can be perfectly decomposed in a linear 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 x over the residual (R).

The SRR is set by comparing the acoustical indices computed on the signal x and on $x^{(n)}$. A convenient SRR would be $20dB$, since variations of acoustical indices lie under 5%.



The direct sound and the first reflection of an exerimental RIR.



Matching Pursuit ran on an experimental RIR ($SRR = 20dB$, compensation of the energy decay).

The linear set of coefficients, derived from MP run on experimental RIRs, is assumed to represent the temporal distribution of arrivals.

Detecting arrivals

Validation using a theoretical model of RIR

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.

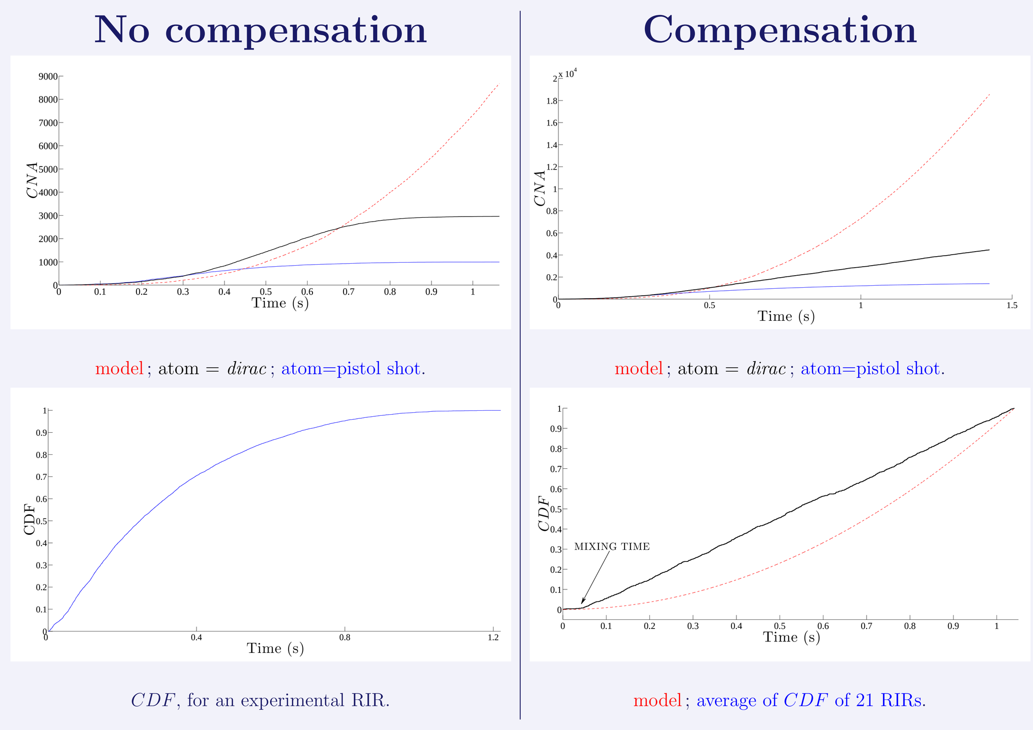
The RIR of the considered hall is synthesized by convolving a pistol shot with the linear set of arrivals calculated previously.

Without compensating the energy decay

As MP selects the maximum of correlation at each iteration, it is obvious that it has a higher probability to be found at the beginning of the RIR.

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.

No compensation | **Compensation**



CDF, for an experimental RIR. | model ; average of CDF of 21 RIRs.

CNA : time evolution of the probability to detect arrivals in the RIR.

CDF (Cumulative Distribution Function) : CNA normalized by the total number of arrivals.

Compensating the energy decay

Energy compensation :

- makes the signal stationary and ergodic
- ensures equal weight to all parts of the RIR
- ensures equiprobability of detecting arrivals

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
- allows to define the mixing time as the time where this difference occurs
- mixing precisely expresses the equiprobability of arrivals, as defined by Krylov

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.

Further work & References

$\rightarrow$Small coefficients of MP = diffusion ?

$\rightarrow$Effect of varying values of the SRR ? atoms size ? sound source ?

$\rightarrow$Further studies should :

- be carried out on the robustness of such an estimator : measurements in other rooms
- perceptively evaluate, using listening tests, the relevance of such a stopping criterium

$\rightarrow$Filtering atoms $\Rightarrow$ filter bank of the room ?

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

J-D Polack, *Playing billiards in the concert hall : The mathematical foundations of geometrical room acoustics*, Applied Acoustics, vol. 38, pp.235-244, 1993

S. Mallat and Z. Zhang, *Matching Pursuit with time frequencies dictionaries*, IEEE. Trans. Signal Process., vol. 40, no. 12, pp. 3397-3415, 1993.

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

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