DETECTING ARRIVALS WITHIN ROOM IMPULSE RESPONSES USING MATCHING PURSUIT

Guillaume Defrance, Laurent Daudet, Jean-Dominique Polack
IJLRDA, LAM team
UPMC Univ. Paris 06, CNRS, Minsitère de la Culture et de la Communication
Paris, France
`defrance@lam.jussieu.fr`

ABSTRACT

This paper proposes to use Matching Pursuit, in order to investigate some statistical foundations of Room Acoustics, such as the temporal distribution of arrivals, and the estimation of mixing time. As this has never been experimentally explored, this study is a first step towards a validation of the ergodic theory of reverberation. The use of Matching Pursuit is implicit, since correlation between the impulse response and the direct sound is assumed. The compensation for the energy decay is necessary to obtain stationnary signals. Methods for determining the best the temporal boundaries of the direct sound, for choosing an appropriate stopping criteria based on the similarity between acoustical indices of the original RIR and those of the synthesized signal, and for experimentally defining the mixing time constitute the scope of this study.

1. INTRODUCTION

This paper proposes a method for studying the temporal distribution of the arrivals of Room Impulse Responses (RIRs). The theory, developed and presented in [1] [2] has never been experimentally validated yet. This aim of this work is to lead toward a better understanding of the physical and statistical behaviour of the propagation of energy in halls.

Assuming a high correlation between the RIR and the direct sound, with due consideration to the filtering of the room, Matching Pursuit [3] appears to be well suited for this purpose. As a first step, a dictionnary of atoms has to be defined that matches exactly the signal. The atoms are represented, in this case, by the direct sound itself. The determination of its exact temporal boundaries is of importance for a perfect match. 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 sound sources in room acoustics measurements [4], such as balloon bursts or pistol shots. A basic approach of this problem would consist in computing the Short Time Fourier Transform of a RIR, and to detect temporal or spectral changes for the beginning and the end of the motive. Nevertheless, attention must be paid to the drawback of this method, that is, its resolution. Indeed, the time resolution is proportional to the effective duration of the analysis window. Similarly, the effective frequency resolution is proportional to the effective band-width of the analysis window. Consequently, a *trade-off* between time and frequency resolution has to be made. On the one hand, a good time resolution requires a short analysis window, while on the other hand a good frequency resolution requires a long window. This limitation leads to choose another approach, based on the inter-correlation between the RIR and the direct sound itself (section 2). This section details the choice of a stopping criteria, and presents a method which can be useful for detecting the direct sound of RIRs (without any knowledge of its properties). For RIR measurements carried out with balloons bursts or pistol shots, it becomes difficult to identify clearly temporal boundaries when the sound source is not recorded in the near field, or in an anechoic chamber. Moreover, the visual identification by hand of these boundaries may vary from expert to expert, since the background noise often disturbs the readability of the signal. The Matching Pursuit run on a RIR provides original results. The base of coefficients obtained using Matching Pursuit can be seen as a weigth distribution of arrivals of the RIR (an arrival being a sound ray which has undergone one or more reflections on its way from the source to the receiver [1]). Section 3 exposes theoretical reviews of some statistical concepts of room acoustics, and compares the temporal density of arrival, derived from experiments using Matching Pursuit, to the theory [2].

The presented results are derived from one hundred measurements of RIRs in salle Pleyel carried out with pistol shots [5].

2. MATCHING PURSUIT APPLIED TO RIR

2.1. Theoretical reviews

A RIR can be seen as a linear base of occurence of the direct sound reproduced in time, and filtered by the surfaces of the hall, as seen in Figure 1. For this latter reason, it is believed that Matching Pursuit can help for understanding more deeply the architecture of a RIR, since this algorithm introduced by [3] provides information, which can be seen as maxima of correlation (Eq.1) [6] between two signals: the RIR and the direct sound (the atom).

In theory, any signal x can be perfectly decomposed in a linear base 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/noise ratio (SNR) in dB of x over the residual (R).

——— *Algorithm of Matching Pursuit* ———

```
Input : $x \in \mathbb{R}^N, \mathcal{D} = \{\phi_\gamma, \gamma\}$
Output : $\gamma^k_{opt}, \alpha^k, k = 1 \ldots (n-1)$

    $n = 0$
$R^0 x \leftarrow x$
```

Repeat:

$$\gamma_{opt}^{(n)} \leftarrow argmax_{\gamma\in\Gamma}|\langle R^n x, \phi_\gamma\rangle| \qquad (1)$$

$$\alpha(n) \leftarrow |\langle R^n x, \phi_{\gamma_{opt}^{(n)}}\rangle| \qquad (2)$$

$$R^{n+1}x \leftarrow R^n x - \alpha^{(n)}\phi_{\gamma_{opt}^{(n)}} \qquad (3)$$

$$n \leftarrow n+1 \qquad (4)$$

$$x^{(n)} = \sum_{k=0}^{n-1}\alpha^{(k)}\phi_{\gamma_{opt}^{(n)}} \qquad (5)$$

---

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.

2.2. Stopping criterium

2.2.1. Finding an appropriate value

The quality of the decomposition of x in atoms depends on the value of SNR, that is, the stopping criteria. On the one hand, for a too low SNR, the residual has a too high energy level and the rebuilt signal $x^{(n)}$ is an impoverished approximation of x. On the other hand, a too high SNR leads to a high number of iterations, that becomes useless over a certain value, since $x^{(n)}$ is almost identical to x.

Comparing the acoustical indices, used in Room Acoustics [7], calculated on x to those calculated on $x^{(n)}$ for different values of a SNR allows to set the stopping criterium on a solid physical background. The acoustical indices presented here are the reverberation times at $20dB$ (RT_{20}), $30dB$ (RT_{30}), the Early Decay Time at $10dB$ (EDT_{10}), and the Central Time (T_C). Figure 2 shows the variations in percent between indices calculated on x and those calculated on $x^{(n)}$ for different values of SNR, for an impulse response, in which the visual identification of the direct sound is obvious (Figure 1). 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 is facilitated.

According to Figure 2, a convenient SNR would be $20dB$, since variations of acoustical indices lie under 5%. Figure 3 shows a RIR and the linear base of coefficients.

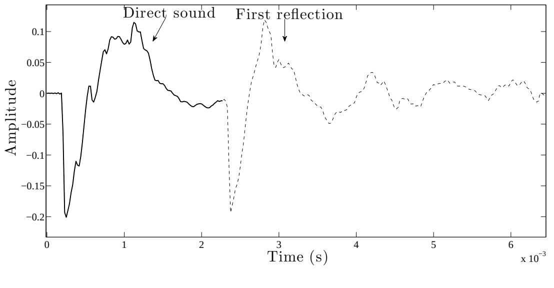

Figure 1: *Experimental RIR for which determining the direct sound is obvious.*

2.2.2. Detection of the direct sound

Matching Pursuit can also be used for determining the time duration of impulses of experimental RIRs, with stopping criterium of $SNR = 20dB$ (as seen in section 2.2.1). An efficient decomposition of a signal x in a linear base of atoms (i.e the direct sound) can only be achieved if the 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 searching method on the first and the last indices of the atom. This method is thought to provide a higher precision in time than one based on $STFT$ (section 1).

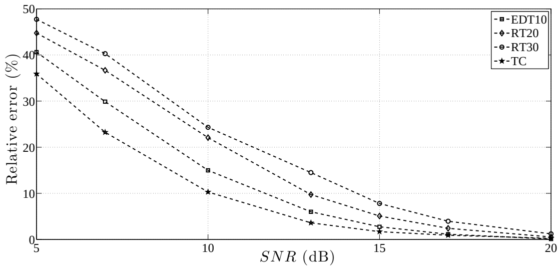

Figure 2: Variations in % of RT_{20}, RT_{30}, EDT and T_C versus SNR in dB.

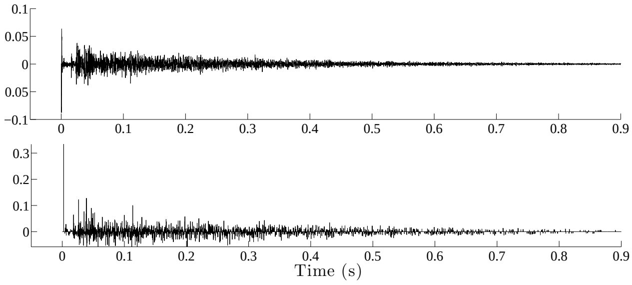

Figure 3: *Matching Pursuit ran on an experimental RIR ($SNR = 20dB$). -(top): Experimental RIR -(bottom): Linear base of coefficients, which correspond to the arrivals.*

3. DETECTION OF ARRIVALS

The rest of the paper focuses on all trajectories emitted from a given fixed source position and reaching a given fixed receiver position after some given elapsed time. These trajectories are called arrivals [1], because they correspond to the different arrivals of sound at the receiver position when an impulse is emitted from the source. In room acoustics, the time distribution of these arrivals plays an important role since it is directly linked to the impulse response. Indeed, the RIR is built by the superposition of all the pulses emitted by the source at the same time, and reaching the receiver after travelling through the room. In other words, the RIR is composed of the succession of all arrivals, with due consideration for their respective amplitudes. In a mixing room, that means with an irregular shape, the energy is equidistributed in the hall after sufficient time has elapsed, i.e after the mixing time [2]. Since adjacent trajectories in a mixing room diverge at an exponential rate, intensity reduces exponentially along any trajectory. Therefore, to conserve the energy, the number of arrivals must compensate the divergence. Hence, the number of arrivals is assumed to increase exponentially with time in any mixing room. However, at the early stage, the image sources are regularly distributed in space and one can assume a power low distribution [2].

3.1. Robustness of Matching Pursuit on a model of RIRs

This section aims at testing the robustness of Matching Pursuit on IRs using a model of IR presented in [2] [8]. This model assumes that the arrivals are distributed in time according to a Poisson process, with a parameter which is time dependent. The cumulative number of arrivals is a cubic function of time (Figure 4). Using simple input parameters, such as the reverberation time, the mean absorption, and the volume of the hall, the time arrivals base is generated. The IR of the considered hall is synthesized by convolving a pistol shot with the linear base of arrivals calculated previously. Surprisingly, applying Matching Pursuit (with knowledge of the atom) to the synthesized IR does not return the exact linear base of arrivals derived from the model (Figure 4). Explanations are given in the following.

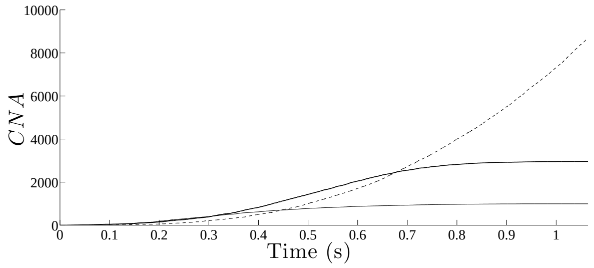

Figure 4: *Cumulative numbers of the arrivals (CNA) generated by the stochastical model of IR (without compensation of energy decay). -dashed: model -plain: the atom is a pistol shot -bold: the atom is a dirac.*

3.2. Towards a validation of the statistical theory

Based on section 3.1, the linear base of coefficients (Figure 3), derived from Matching Pursuit run on experimental RIRs, are assumed to represent the temporal distribution of arrivals.

3.2.1. Without compensating for the energy decay

The cumulative number of arrivals (CNA) normalized by the total number of arrivals, plotted in Figure 5, represents an estimate of the Cumulative Distribution Function (CDF). In other words, the CNA describes the time evolution of the probability to detect arrivals in the RIR.

Figure 5 underlines the decreasing of probability to detect arrivals at the end of the RIR. This is an artefact of Matching Pursuit. Indeed, as Matching Pursuit selects the maximum of correlation at each iteration, it is obvious that it has a high probability to be found at the beginning of the RIR (Figure 3). 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.

Measurements in salle Pleyel have been carried out for 21 different source-receiver positions, hence for 21 RIRs. By calculating the logarithm of $1 - CDF$ for the 21 RIRs (Figure 6), the mean reverberation decay of the room is recovered ($RT \approx 2.0s$). This observation highlights that the CDF is linked to the energy of the signal.

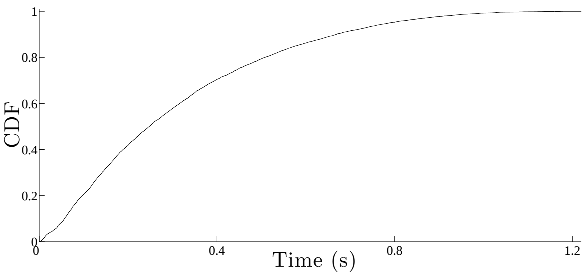

Figure 5: *CDF, for an experimental RIR (without compensation of the energy decay).*

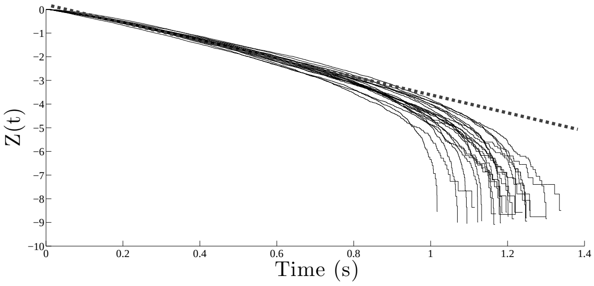

Figure 6: *plain: Logarithm of ($1 - CDF$) of the arrivals of 21 experimental RIRs (without compensation of the energy decay) -dashed: mean logarithmic decay decay of the RIRs.*

3.2.2. Compensating for the energy decay

As seen in section 3.2.1, 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 7, where it corresponds to the almost constant slope of the CDF.

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 as the time where this difference occurs. Indeed, mixing precisely expresses the equiprobability of arrivals, as defined by Krylov [9]. 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.

First tests show that the mixing time depends on the atom, that is, on the temporal and spectral properties of the direct sound (Figure 8). But this goes beyond the scope of the present paper.

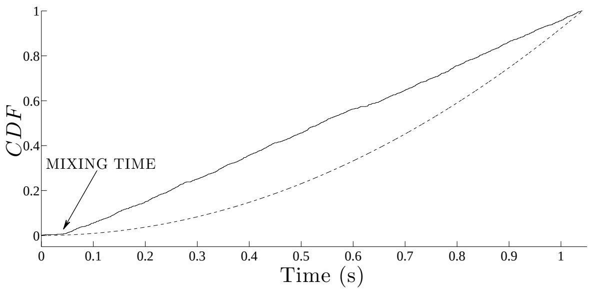

Figure 7: $CDFs$. *-dash: theoretical* CDF *(*$\approx t^3$*) -plain: Average of* CDF *for* 21 *RIRs (with compensation of the energy decay).*

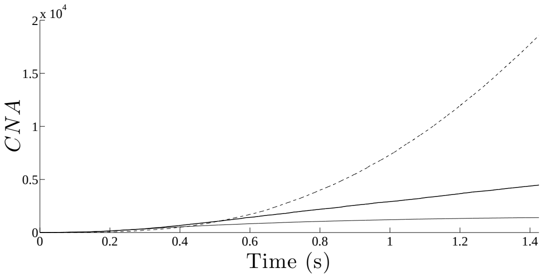

Figure 8: CNA *of a theoretical RIR. -dash: theoretical* CNA *-plain: estimated* CNA *(the atom is a pistol shot) -bold: estimated* CNA *(the atom is a dirac) (with compensation of the energy decay).*

4. DISCUSSION

As seen previously, the stopping criterium is important, since its value specifies the approximation of the original signal: the higher the SNR, the larger is the number of coefficients. For instance, coefficients of small amplitudes are thought to represent the scattering occuring between arrivals, specially after the first reflection. This latter point raises the question of using a threshold applied on the linear base of coefficients, so that small coefficients are not taken into account. Moreover, as the CDF defines the mixing time, further studies should be carried out on the robustness of such an estimator, specially for varying values of the SNR, and for different atoms size. It remains to link this experimental estimation of the mixing time to the theory of arrivals, which will be the topic of a future paper. Finally, 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.

5. CONCLUSIONS

This study uses a well known 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 atom of the dictionnary. The stopping criterium is choosen by minimizing the difference between the acoustical indices of the original RIR and the synthesized one. Further studies should perceptively evaluate, using listening tests, the relevance of such a stopping criteria. 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, that should be achieved in a future work.

On the other hand, Matching Pursuit provides another vision of RIRs. Indeed, the linear base of coefficients obtained are seen as the arrivals of the RIR. The cumulative number of arrivals is confrontated to the theory. 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. It remains to generalize this estimator to other rooms, using some different atoms, and to finally confrontate the results to the theory. Moreover, more investigation should be made with filtering the RIR and using threshold on the linear base of coefficients, derived from Matching Pursuit.

6. ACKNOWLEDGMENTS

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

7. REFERENCES

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

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

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

[4] P. Fausti and A. Farina, “Acoustic measurements in opera houses: Comparison between different techniques and equipment,” *Journal of Sound and Vibration*, vol. 232, no. 1, pp. 213–229, June 2000.

[5] G. Defrance, J-D. Polack, and B-FG. Katz, “Measurements in the new salle pleyel,” in *Proc. Int. Symp. Room Ac.*, Sevilla, September 2007.

[6] S. Krstulovic and R. Gribonval, “Mptk: Matching pursuit made tractable,” in *ICASSP’06*, Toulouse, France, May 2006.

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

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

[9] 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.