



# Caractérisation du mélange dans les réponses impulsionales de salles. Application à la détermination expérimentale du temps de mélange.

Guillaume Defrance

## ► To cite this version:

Guillaume Defrance. Caractérisation du mélange dans les réponses impulsionales de salles. Application à la détermination expérimentale du temps de mélange.. Acoustique [physics.class-ph]. Université Pierre et Marie Curie - Paris VI, 2010. Français. NNT: . tel-00488969

**HAL Id: tel-00488969**

<https://theses.hal.science/tel-00488969v1>

Submitted on 3 Jun 2010

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



**THESE DE DOCTORAT DE  
L'UNIVERSITÉ PIERRE ET MARIE CURIE**

Spécialité :  
Mécanique, Acoustique

Présentée par :  
M. Guillaume DEFRAANCE

Pour obtenir le grade de  
**DOCTEUR DE L'UNIVERSITÉ PIERRE ET MARIE CURIE**

Sujet de la thèse :

CARACTÉRISATION DU MÉLANGE DANS LES RÉPONSES IMPULSIONNELLES DE SALLES.  
*Application à la détermination expérimentale du temps de mélange.*

Soutenue le : 30 Novembre 2009

Devant le jury composé de :

M. Denis Bosq, Professeur Emérite à l'Université Pierre et Marie Curie  
M. Dominique Delande, Directeur de recherche CNRS  
M. Finn Jacobsen, Associate Professor à l'Université Technologique du Danemark  
M. Fabrice Mortessagne, Professeur à l'Université de Nice  
M. Jean-Dominique Polack, Professeur à l'Université Pierre et Marie Curie  
M. Bob L. Sturm, Docteur de l'Université de Santa Barbara, CA, Etats-Unis

Examinateur  
Président du Jury  
Rapporteur  
Rapporteur  
Directeur de Thèse  
Examinateur



Version révisée le 04/03/2010 et approuvée par le jury.

*A Isabelle,  
A Jean-Dominique,  
A Aube, Eric et Emma.*

*A ceux du lointain et obscur passé :  
vos fautes et votre bêtise m'ont toujours  
poussé à vous dépasser.*

# Remerciements

Je remercie les membres du jury, rapporteurs et examinateurs, pour avoir bien voulu évaluer ce travail : MM. Denis Bosq et Dominique Delande qui ont bien voulu prendre de leur précieux temps, MM. Finn Jacobsen et Fabrice Mortessagne qui viennent du Danemark et de Nice respectivement, et enfin M. Bob L. Sturm, docteur de l'Université de Santa Barbara en post-doctorat au LAM.

Je tiens particulièrement à remercier Jean-Dominique Polack, qui a dirigé cette thèse durant trois années. J'imagine sans difficulté que la tâche n'a pas été simple de son côté. Jean-Dominique a pourtant réussi à me transmettre l'esprit de recherche, de rigueur et de critique. Je reste conscient néanmoins du long chemin qu'il me reste à parcourir avant d'atteindre son niveau de connaissance et de compréhension. Je garderai en mémoire ces moments de discussions, d'échange de points de vue qui ont fait avancer notre compréhension du problème. Je ne peux pas oublier de mentionner l'ouverture intellectuelle de Jean-Dominique qui a été un réel enrichissement : de l'acoustique des salles à Rabelais, de la physique à la perception, en passant par Wagner, les billards et Proust. Je tiens à clore ce paragraphe par le souvenir de ma première conférence qui se tint à Séville, en Espagne à l'automne 2007. A la fin d'un repas au restaurant avec les autres conférenciers, je réussis à convaincre Jean-Dominique (épuisé de la journée) de m'accompagner boire une bière dans un bar enfumé de la vieille ville. Même si ce moment fut de courte durée, j'espérais pouvoir m'en souvenir longtemps. Je ne me rappelle pas précisément de notre discussion (était-ce important ?), mais c'était un moment simple, très agréable, au cours duquel j'ai commencé à découvrir l'homme qui était mon directeur. Pour tout cela, je l'en remercie chaleureusement.

Il me faut également remercier deux autres acteurs de ces trois dernières années. Tout d'abord, Laurent Daudet que j'ai rencontré au LAM et qui, au moment où j'écris ce mémoire, devient Professeur de l'Université Paris 7. Je lui souhaite de réussir dans les nouvelles voies qui lui sont ouvertes (ce qui ne fait de toute manière aucun doute). Je dois souligner le rôle de Laurent, coordonné à celui de Jean-Dominique. Laurent m'a été d'une grande aide lors de la rédaction de mon premier article dont le cœur était l'application d'outils du traitement du signal à l'acoustique des salles. Ce premier passage obligé dans une thèse -la rédaction d'un article- fut difficile à de nombreux égards, mais tellement enrichissant !

J'ai rencontré Michel Retbi, architecte et enseignant à l'école d'architecture de Paris-Val de Seine, lors d'une conférence à Oslo à l'automne 2008. Cette semaine passée dans la capitale norvégienne fut primordiale à mes yeux : j'étais au contact d'un architecte sensible aux questions de l'acoustique. Une semaine à écouter, questionner, et apprendre du discours de Michel. Il me faut aussi avouer que nous n'avons pas parlé uniquement d'acoustique et d'architecture. Le Jazz a également occupé une grande place lors de cette semaine. A l'époque, je débutais la trompette et terminais la lecture de "Really the blues"<sup>1</sup> de Milton "Mezz" Mezzrow. Quant à Michel, il était depuis longtemps rompu à la pratique de la clarinette et ne tarissait pas d'anecdotes sur des grands ou petits jazzmen.

Il est impossible de remercier avec autant de détails que précédemment chaque membre de l'équipe LAM. Ce serait beaucoup trop long. C'est pourquoi j'ai une pensée particulière pour celles et ceux qui ont été présents ponctuellement ou quotidiennement : permanents, post-docs,

---

<sup>1</sup> *La rage de vivre*, dans sa version française.

et thésards. Je suis convaincu que l'ouverture intellectuelle de cette équipe a joué un rôle important dans mon travail et évolution au cours de ces trois années.

Même s'il s'agit aux yeux de certains d'une coutume un peu désuète, je tiens sincèrement à remercier mon entourage proche pour leurs questions, encouragements, et particulièrement Isabelle pour avoir supporté ces trois années parfois difficiles et avoir été à mes côtés.

Enfin, la dernière personne de cette liste que je ne pourrai oublier est M. Jean-Luc Buraud, Professeur à l'Institut Universitaire de Technologies du Mans. En effet, alors que je commençais -après un IUT Mesures Physiques- l'étude de l'acoustique et des vibrations à l'Université du Maine, je me suis tourné vers Jean-Luc afin de trouver quelques explications à mes questions. Un soir, au cours d'une conversation, Jean-Luc me dit sans détour que j'étais fait pour la recherche et devais préparer une thèse de doctorat. Pendant presque trois ans j'ai considéré cette idée saugrenue : je ne comptais pas faire de longues études, et la recherche me paraissait être un monde inaccessible. Cependant, je me suis régulièrement rappelé cette conversation durant laquelle je n'ai rien dit (il s'agissait plus d'un monologue) ; bien souvent son souvenir m'a aidé dans mon orientation universitaire. Aujourd'hui que j'écris les dernières lignes de ce mémoire, je prends pleinement conscience des paroles et de l'enseignement de Jean-Luc. Je l'en remercie sincèrement.



# Table des matières

<b>Remerciements</b>	<b>7</b>
<b>1 Introduction</b>	<b>13</b>
<b>2 Acoustique des salles</b>	<b>17</b>
2.1 Introduction . . . . .	17
2.2 Généralités . . . . .	17
2.2.1 Les phénomènes physiques mis en jeu . . . . .	17
2.2.2 Le champ rayonné dans une salle . . . . .	18
2.3 Théorie statistique de l'acoustique des salles . . . . .	24
2.3.1 Domaine fréquentiel et temporel . . . . .	25
2.3.2 Réflexion diffuse . . . . .	29
2.4 Méthodes de synthèse de réponses impulsionnelles de salles . . . . .	30
2.4.1 Introduction . . . . .	30
2.4.2 Remarque générale -Vers une méthode probabiliste . . . . .	32
2.5 Modélisation stochastique de l'acoustique des salles de spectacles . . . . .	33
2.5.1 Introduction . . . . .	33
2.5.2 Le modèle stochastique . . . . .	33
2.5.3 Améliorations proposées . . . . .	34
2.5.4 Conclusion . . . . .	39
2.6 Qualité acoustique d'une salle . . . . .	39
2.6.1 Rappel des phénomènes physiques mis en jeu . . . . .	39
2.6.2 Quelques indices acoustiques . . . . .	44
2.6.3 Mesure de réponses impulsionnelles de salles . . . . .	45
<b>3 Mesure et analyse</b>	<b>49</b>
3.1 Introduction . . . . .	49
3.2 From MIDAS to OpenMIDAS . . . . .	50
3.2.1 Introduction . . . . .	50
3.2.2 Description of module P . . . . .	50
3.3 Détection de l'onset . . . . .	52
3.3.1 Préambule . . . . .	52
3.3.2 Abstract . . . . .	52
3.3.3 Introduction . . . . .	52
3.3.4 Precision of onset timing vs precision of the indices . . . . .	53
3.3.5 Proposed onset detection methods . . . . .	53
3.3.6 Comparison of the 12 automatic onset detection methods . . . . .	56
3.3.7 Conclusion . . . . .	58
3.4 Caractérisation de sources sonores . . . . .	59

3.4.1	Préambule . . . . .	59
3.4.2	Abstract . . . . .	59
3.4.3	Introduction . . . . .	59
3.4.4	Visualizing the direct sound using the Wigner-Ville distribution . . . . .	60
3.4.5	Methods for estimating the direct sound duration . . . . .	61
3.4.6	Reference impulses . . . . .	63
3.4.7	Results . . . . .	63
3.4.8	Discussion . . . . .	63
3.4.9	Conclusions . . . . .	64
<b>4</b>	<b>Détection du temps de mélange à l'aide de tests de Gaussianité</b>	<b>69</b>
4.1	Introduction . . . . .	69
4.2	Les cumulants d'une loi normale . . . . .	70
4.2.1	L'écart-type $\sigma$ . . . . .	71
4.2.2	Le kurtosis $k$ . . . . .	71
4.3	Tests statistiques . . . . .	72
4.3.1	Tests d'ajustement . . . . .	73
4.3.2	Test de Kolmogorov-Smirnov . . . . .	76
4.4	Discussion et conclusion . . . . .	78
<b>5</b>	<b>Estimation du temps de mélange à l'aide de Matching Pursuit</b>	<b>81</b>
5.1	Préambule . . . . .	81
5.2	Using Matching Pursuit for estimating mixing time within Room Impulse Responses	81
5.2.1	Abstract . . . . .	81
5.2.2	Introduction . . . . .	82
5.2.3	Theoretical review of ergodic rooms . . . . .	83
5.2.4	Matching Pursuit applied to RIR . . . . .	85
5.2.5	Detection of arrivals . . . . .	87
5.2.6	Estimation of the mixing time . . . . .	89
5.2.7	Differences between data from the experiment and the model . . . . .	95
5.2.8	Discussion . . . . .	96
5.2.9	Conclusions . . . . .	97
5.3	Développements ultérieurs . . . . .	99
5.3.1	Taux d'accroissement du nombre de retours . . . . .	99
5.3.2	Améliorations du modèle - Impacts . . . . .	101
5.3.3	Les retours dans l'espace . . . . .	105
5.3.4	Estimation perceptive du critère d'arrêt de Matching Pursuit . . . . .	110
5.3.5	Ouvertures possibles . . . . .	114
5.3.6	Conclusion . . . . .	116
<b>6</b>	<b>Estimation du temps de mélange à l'aide de l'eXtensible Fourier Transfrom</b>	<b>125</b>
6.1	Préambule . . . . .	125
6.2	Estimating the mixing time of concert halls using the eXtensible Fourier Transform	125
6.2.1	Abstract . . . . .	125
6.3	Introduction . . . . .	126
6.4	Statistical properties of room acoustics . . . . .	127
6.5	The eXtensible Fourier Transform . . . . .	129
6.5.1	Principle and definition . . . . .	129
6.5.2	Properties of the XFT . . . . .	129
6.6	Application to an audio signal . . . . .	131

6.7	Estimating mixing time of RIRs by calculating phase differences . . . . .	133
6.7.1	Phase of the XFT . . . . .	133
6.7.2	Phase differences . . . . .	135
6.7.3	Estimating of the mixing time in practice . . . . .	135
6.8	Estimating the mixing time of a real concert hall . . . . .	136
6.9	Estimating the mixing time of synthesized concert halls . . . . .	137
6.9.1	Stochastic model of RIRs . . . . .	138
6.9.2	Estimation of the mixing time on synthesized RIRs . . . . .	138
6.10	Variability of the XFT . . . . .	139
6.10.1	Variability with the background noise level . . . . .	139
6.10.2	Variability with the analysis window width . . . . .	140
6.11	Discussion . . . . .	140
6.12	Conclusion . . . . .	142
6.13	Acknowledgments . . . . .	142
6.14	Développements ultérieurs . . . . .	142
6.14.1	Comparaison avec Matching Pursuit . . . . .	142
6.14.2	Ouvertures . . . . .	143
<b>7</b>	<b>Discussion des résultats</b>	<b>155</b>
7.1	Principaux résultats . . . . .	155
7.2	Les définitions du temps de mélange . . . . .	156
7.3	Alternatives . . . . .	157
<b>8</b>	<b>Conclusion</b>	<b>161</b>
<b>Bibliographie</b>		<b>169</b>
<b>A Annexe : Méthodes de synthèse de réponses de salles</b>		<b>171</b>
A.1	Méthode des sources-images . . . . .	171
A.1.1	Principe . . . . .	171
A.1.2	Limites de la méthode . . . . .	172
A.2	Méthode du lancer de rayons . . . . .	173
A.2.1	Principe . . . . .	173
A.2.2	Limites de la méthode . . . . .	174
A.3	Méthode hybride . . . . .	174
A.3.1	Principe . . . . .	174
A.3.2	Limites de la méthode . . . . .	175
<b>B Annexe : OpenMIDAS</b>		<b>177</b>
B.1	OpenMIDAS - Quick start of Processing Module . . . . .	177
B.1.1	Installing OpenMIDAS package . . . . .	177
B.1.2	Running OpenMIDAS . . . . .	178
B.1.3	Set MatLab . . . . .	178
B.1.4	Possible errors . . . . .	179
B.2	Detailed description of the Processing module of OpenMIDAS . . . . .	179
B.3	Action of each brick . . . . .	181
B.3.1	IndexComputation.m . . . . .	181
B.3.2	SetConstants.m . . . . .	184
B.3.3	SetBandfilter.m . . . . .	185
B.3.4	ReadCritDescriptor.m . . . . .	186

B.3.5	SetCalibration.m . . . . .	187
B.3.6	ComputeLateralEnergy.m . . . . .	188
B.3.7	SetAirComp.m . . . . .	189
B.3.8	ComputeDirectEnergy.m . . . . .	190
B.3.9	SignalDecay.m . . . . .	191
B.3.10	NoiseDetectionCompensation.m . . . . .	192
B.3.11	ComputeTransIndex.m . . . . .	193
B.3.12	ComputeDecayCurve.m . . . . .	194
B.3.13	ComputeIIA.m . . . . .	194
B.4	Functions needed for computation . . . . .	196
<b>C</b>	<b>Annexe : Estimation des retours et du temps de mélange à partir de deux corpus de réponses de salles</b>	<b>199</b>
<b>D</b>	<b>Annexe : Tests de Gaussianité appliqués à des réponses impulsionales Gaus-siennes</b>	<b>203</b>
<b>E</b>	<b>Annexe : Evaluation perceptive du modèle stochastique</b>	<b>207</b>
E.1	Introduction . . . . .	207
E.2	Choix du test . . . . .	207
E.3	Choix des paramètres du modèle . . . . .	208
E.3.1	Test de catégorisation libre . . . . .	209
E.3.2	Résultats du test de catégorisation libre . . . . .	210
E.4	Protocole expérimental . . . . .	211
E.5	Discussion . . . . .	211
E.6	Résultats du test de catégorisation libre . . . . .	211

# Chapitre 1

## Introduction

La compréhension des lois de propagation du son dans un espace clos est une préoccupation qui date au moins de l'Antiquité. En effet, les Grecs travaillaient déjà à améliorer l'intelligibilité dans leurs amphithéâtres [1]. L'acoustique des salles est un sujet qui aujourd'hui concerne tant les chercheurs que les consultants en acoustique du bâtiment ou les architectes. Cette discipline est présente dans de nombreux domaines : salles de spectacles, salles de classes, halls de gare, bibliothèques, ou encore les rues (acoustique urbaine).

Au XIXème siècle, Sabine fait un pas décisif en proposant une formalisation statistique de l'acoustique des salles [2]. Cela aboutit, entre autres choses, à la fameuse loi de la réverbération dite de Sabine. Alors que la conception classique de la propagation de l'énergie sonore dans une cavité est basée sur une approche ondulatoire, l'approche géométrique, uniquement valable dans la limite des hautes fréquences, propose de conceptualiser une onde par son front et donc d'assimiler la marche de ce front d'onde à celui d'un rayon sonore. En 1975, Joyce [3] s'appuie sur près de 70 ans de travaux sur les systèmes dynamiques et propose à la communauté scientifique une vision particulière : il ne s'agit plus d'ondes, ou de rayons sonores, mais de phonons se mouvant dans une cavité. Si le milieu de propagation est isotrope et l'absorption est faible et uniformément répartie (homogène), Joyce montre que l'on retrouve la loi de Sabine. Ces travaux importants établissent théoriquement que les grandes salles sont ergodiques et mélangeantes, comme le sont certains systèmes dynamiques<sup>1</sup>.

Parallèlement, de nombreux travaux expérimentaux ont validé certains aspects de la théorie statistique de l'acoustique des salles. Dans les années 1930, nous assistons aux prémisses des mesures dans les maquettes de salles, marquant ainsi les débuts de la prédiction de l'acoustique d'un lieu avant qu'il ne soit construit. Le développement des technologies de mesure et d'enregistrement ont permis d'affiner la précision de la prédiction, mais aussi la recherche en acoustique des salles. Dans les années 1970, l'avènement des ordinateurs donne naissance aux modèles numériques [4, 5]. Ceux-ci, bien que lourds en temps de calculs et en préparation (la salle et ses détails doivent être modélisés en 2 ou 3 dimensions), prédisent précisément les temps d'arrivée des rayons sonores -et leur amplitude- en un point de la salle. Ces modèles admettent quelques approximations<sup>2</sup> et ne synthétisent en réalité que les premiers instants de la réponse impulsionale des grandes salles. Ainsi la réverbération tardive est modélisée d'après des observations expérimentales (expliquées par la théorie) : cette partie de la réponse impulsionale est assimilée à bruit Gaussien à partir du temps de mélange (propre aux systèmes dynamiques).

Dans les années 1990, Jean-Dominique Polack revisite les travaux de Joyce et montre mathémati-

---

<sup>1</sup>Le billard de Sinaï, par exemple.

<sup>2</sup>C'est le propre de la modélisation des phénomènes de la nature : elle est possible dans de nombreux cas, mais au prix de certaines approximations.

quement en s'appuyant sur la théorie des billards que la réverbération peut-être modélisée par un processus stochastique [6].

Le présent manuscrit synthétise le travail effectué dans le cadre d'une thèse de doctorat dont le thème est la modélisation stochastique de l'acoustique des salles de spectacles. En d'autres termes, il s'agit de travailler à décrire de manière probabiliste la propagation du son dans une grande salle à la géométrie complexe. Il y a 15 ans environ, J-D. Polack proposait le modèle stochastique dont il est ici question [6]. Après avoir vérifié que ce modèle, simple et léger d'utilisation car ne prenant pas en compte la géométrie de la salle<sup>3</sup>, permet de retrouver les indices acoustiques calculés sur une large base de réponses impulsionales mesurées dans de nombreuses salles de spectacles [7], la question concerne maintenant la validité statistique et écologique du modèle. C'est que nous allons étudier dans les pages suivantes.

Avant de rentrer dans le vif du sujet, nous nous proposons de passer en revue les phénomènes rencontrés lors de la propagation du son dans une salle et l'approche statistique employée pour décrire de tels phénomènes. Nous serons amenés à présenter les bases du modèle stochastique développé par Polack. Comme l'acoustique des salles est un domaine pluridisciplinaire, nous étudierons également les indices acoustiques utilisés pour décrire objectivement l'acoustique d'une salle et les aspects perceptifs, ainsi que les principales méthodes de synthèse de réponses de salles.

Ces présentations théoriques sont suivies des travaux réalisés au cours de cette thèse.

Dans un premier temps, nous nous concentrerons sur la mesure de réponse de salles et de leur analyse. Nous présentons la rénovation d'un logiciel de mesure MIDAS très employé dans les années 1990 par la communauté des chercheurs et des acousticiens des salles. Nous nous focalisons ensuite sur l'estimation du premier (*onset*) et dernier instant du son direct dans le signal de la réponse de la salle. Nous montrons d'une part qu'une estimation fine de l'onset, et dont l'erreur est connue, est nécessaire, et d'autre part, que la détection de ces deux instants permet la caractérisation de sources sonores couramment employées lors des mesures de réponses de salles.

Dans un second temps, nous portons notre attention sur ce qui constitue le cœur de cet ouvrage : l'estimation du temps de mélange des réponses impulsionales de salles de spectacles. Tout d'abord, les tests de Gaussianité sont utilisés afin de détecter le début de la réverbération tardive. Nous montrons en quoi ils ne constituent pas des outils robustes à cette fin. En revanche, une analyse fine des résultats obtenus pose quelques questions quant à l'évolution du système dynamique qu'est une salle de spectacle. Ainsi, nous mettons en évidence grâce aux estimateurs de Gaussianité (Chapitre 4) ce que nous appelons des *bouffées de cohérence*. Ces dernières sont visibles à des temps supérieurs au temps de mélange, alors que le système est censé être diffus.

Nous verrons comment l'emploi d'une technique basée sur la corrélation entre une forme d'onde et le reste de la réponse permet d'estimer les temps d'arrivée des retours. La statistique de ces retours permet également de définir un instant auquel le système change de comportement : de déterministe, il passe à un comportement statistique. La détection de ce temps nous permet de définir le temps de mélange, en accord avec la définition donnée par Krylov [8], et de discuter la validité de la formulation heuristique de celui-ci proposée par Polack [9]. Nous passerons en revue les différentes pistes ouvertes grâce à cet estimateur de retours : estimation du filtrage d'une salle, de son entropie, des directions d'arrivées des retours comme des fonctions du temps. Les résultats présentés sont confrontés à ceux tirés du modèle stochastique.

La même analyse est effectuée à l'aide d'un outil temps-fréquence (l'*eXtensible Fourier Transform -XFT*) que nous présentons, et qui permet de visualiser l'évolution des composantes

---

<sup>3</sup>En effet, le modèle génère des salles aléatoires.

fréquentielles comme une fonction du temps au regard d'une origine fixe du temps et de l'état initial du système.

Pour chaque estimateur (Matching Pursuit et XFT), nous vérifions l'adéquation entre la statistique des retours prédit par le modèle et celle mesurée dans des salles de spectacles.

L'investigation (outils et résultats) de la propriété de mélange (et donc d'ergodicité) des grandes salles est discutée dans un Chapitre à part (Chapitre 7). Il nous paraît en effet important de souligner que les résultats présentés sont une interprétation de la théorie des systèmes dynamiques, et non une démonstration expérimentale. Nous revenons aux différentes définitions du temps de mélange que nous pouvons rencontrer dans la littérature, et mettons en regard nos résultats aux définitions du temps de mélange. Nous concluons sur le fait que le temps que nous estimons n'est pas celui dit de mélange. A la place, nous sommes amenés à proposer un nom différent : temps de transition. Enfin, nous identifions quelques moyens expérimentaux qui pourraient être mis en oeuvre afin de d'observer le caractère mélangeant d'une salle de spectacle.



# Chapitre 2

## Acoustique des salles

### 2.1 Introduction

Ce chapitre, nous le devinons à son titre, traite de l'acoustique des salles. Nous ne traiterons pas ici de l'acoustique générale telle que l'équation d'ondes [10], car cela ne ferait qu'alourdir le propos et serait sans doute hors-sujet. Dans la suite, nous nous attachons à présenter la notion de réponse de salle et les différents phénomènes physiques mis en jeu lors de l'excitation de la salle par un signal. Seront abordés plus particulièrement la notion de rayon sonore, le phénomène de réflexion d'un rayon, l'absorption des parois, ainsi que l'évolution temporelle de la densité des retours, et la densité des modes d'une cavité parallélépipède. Nous serons amenés à définir le temps de mélange tel que nous le trouvons décrit dans le domaine de l'acoustique des salles. Puis, nous passons en revue différentes méthodes de synthèse de réponses de salles et concluons sur la pertinence (*a priori*) d'un modèle stochastique. Enfin, nous terminons par la présentation des indices acoustiques qui servent à décrire objectivement l'acoustique d'une salle.

### 2.2 Généralités

#### 2.2.1 Les phénomènes physiques mis en jeu

Considérons une source sonore placée dans un espace fermé qui rayonne dans toutes les directions. Les ondes sonores se propagent vers les limites du volume et se heurtent aux parois, ou aux obstacles, de la cavité. Le comportement de ces ondes obéissent à des lois qui dépendent principalement des caractéristiques du signal émis et de l'impédance des matériaux rencontrés.

Nous identifions plusieurs facteurs qui interviennent dans la structure du champ rayonné :

##### 1. La source sonore

On s'intéresse à sa répartition temporelle (signaux continus, intermittents, impulsifs, etc.), à ses propriétés spectrales (bande passante, spectre de raies, etc.), à sa puissance ou à son énergie rayonnée, et à sa directivité.

##### 2. Le milieu de propagation

En général, il s'agit de l'air, que nous supposons homogène et isotrope. En d'autres termes, nous négligeons la présence de perturbations susceptibles de créer des variations locales de pression (température, hygrométrie, courants d'air) ou des dispersions (humidité, poussières, etc.). Notons que cette stabilité est rarement observée dans les conditions expérimentales de mesure de réponses impulsives de salle. En effet, dans les salles de concert la présence du public et les spots d'éclairage crée un gradient de température.

### 3. La nature des parois et les obstacles

La nature et la forme des obstacles et parois d'une salle (parois, balcons, tentures, auditoire, etc.) influent sur le comportement des ondes sonores lors de leur propagation dans la salle. Nous distinguons le phénomène de diffusion et d'absorption.

#### 4. La diffusion

Elle est principalement due à la superposition des phénomènes de réflexion, réfraction et diffraction. La diffusion du son est le résultat de tous les changements de direction de propagation des ondes sonores incidentes.

- *Réflexion (spéculaire)* : le changement de direction de l'onde sonore incidente sur une paroi avec un angle donné est régi par les lois de Snell-Descartes. Une analogie avec la réflexion d'un rayon lumineux facilite la compréhension de ce phénomène. Les lois de Snell-Descartes permettent d'établir des relations géométriques ou statistiques sur la répartition du son dans une cavité. Elles sont à l'origine des méthodes de synthèses des sources-images et du tir de rayons sonores (Chapitre 2.4).
- *Diffraction* : l'onde sonore incidente voit sa direction de propagation changée par la présence et la nature des obstacles qu'elle rencontre sur son chemin (ou les inhomogénéités du milieu).

#### 5. L'absorption

L'absorption résulte de la réfraction/transmission d'une onde à l'interface de deux milieux (l'air et une paroi par exemple) et des phénomènes de dissipation.

- *Réfraction et transmission* : en acoustique, les lois de réfraction, due au passage d'un milieu homogène à un autre, sont analogues à celles de l'optique. En effet, le changement d'impédance du milieu de propagation provoque un changement de célérité du son et une variation de direction du front d'onde. Une fraction de l'énergie peut être dans le matériau et s'y trouver en partie absorbée et en partie transmise vers l'extérieur.
- *Dissipation* : ce phénomène est prépondérant à la frontière de deux milieux de nature différente. L'énergie acoustique se dissipe sous forme de chaleur.

Dans la pratique, diffusion et absorption interviennent simultanément de telle sorte que la complexité des lois mises en jeu au cours des multiples réflexions qui se produisent après l'émission d'un son ne permettent pas de traiter l'acoustique des salles autrement qu'à l'aide d'une approche statistique (Section 2.3). Cette approche s'appuie sur une vision géométrique, ondulatoire, statistique et aussi perceptive des lois fondamentales de l'acoustique des salles.

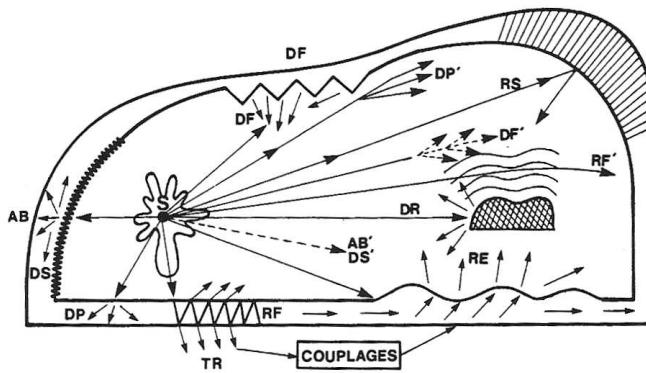
La Fig. 2.1 illustre les précédents phénomènes mis en jeu dans une salle lors de l'émission d'un son.

## 2.2.2 Le champ rayonné dans une salle

### 2.2.2.1 Propagation d'une onde. Les rayons sonores

L'approche géométrique de l'acoustique des salles repose sur l'équivalence des lois de la réflexion des rayons sonores avec celles des rayons lumineux. Cette approche permet une représentation graphique pratique de la propagation des ondes sonores. Considérons l'exemple d'une source omnidirectionnelle émettant dans une salle parallélépipédique. Le tracé des front d'ondes devient complexe après seulement quelques réflexions (Fig. 2.2). Ce tracé est calculé à partir des images de la source engendrées par chaque réflexion spéculaire sur les parois de la cavité. C'est la méthode des sources-images (Chapitre 2.4).

La méthode du tracé de rayons (Chapitre 2.4) permet de rendre compte, aisément, de la propagation d'une onde dans une cavité au-delà des premières réflexions. Chaque rayon est perpendiculaire au front d'onde, et peut être considéré comme l'image d'un pinceau infiniment fin



Phénomènes de diffusion et d'absorption des ondes sonores dans un espace clos :

1) <i>des aux obstacles</i>	<i>réflexion</i>	<i>RS</i>
	<i>réfraction</i>	<i>RF</i>
	<i>diffraction</i>	<i>DR</i>
	<i>diffusion</i>	<i>DF</i>
	<i>réémission</i>	<i>RE</i>
	<i>absorption</i>	<i>AB</i>
	<i>transmission</i>	<i>TR</i>
2) <i>des au milieu de propagation</i>	<i>dissipation</i>	<i>DS</i>
	<i>absorption</i>	<i>AB'</i>
	<i>réfraction</i>	<i>RF'</i>
	<i>diffusion</i>	<i>DF'</i>
	<i>dispersion</i>	<i>DP'</i>
	<i>dissipation</i>	<i>DS'</i>

FIG. 2.1 – Phénomènes de diffusion et d'absorption des ondes sonores dans un espace clos [11].

(Fig. 2.3). Le trajet du rayon peut être suivi pendant un temps assez long, selon les moyens informatiques mis en œuvre. Il est alors possible de créer des modèles de propagation de l'énergie sonore dans une salle dans lesquels des rayons sonores partent de la source sonore avec une certaine énergie. A chaque réflexion sur une paroi, le rayon perd une fraction de son énergie.

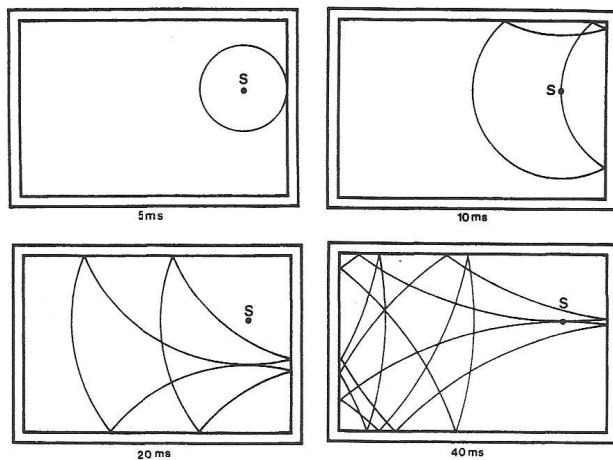


FIG. 2.2 – Représentation bidimensionnelle de l'évolution du front d'onde créé par une source omnidirectionnelle dans une salle parallélépipédique [11].

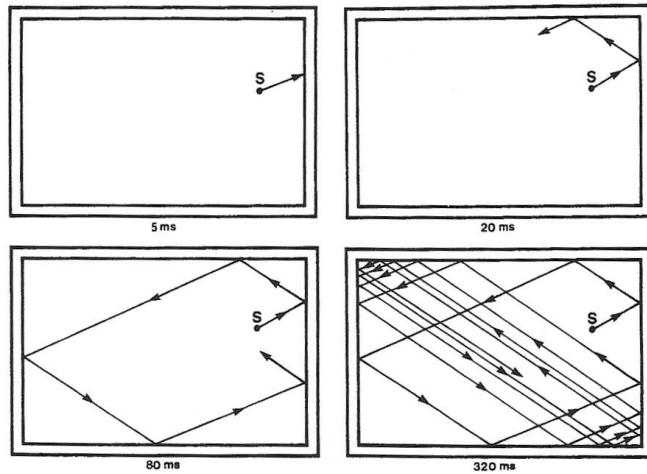


FIG. 2.3 – Représentation bidimensionnelle de l'évolution du trajet d'un rayon sonore dans la salle parallélépipédique de la Fig. 2.2 [11].

Un rayon est défini comme un pinceau infiniment fin transportant l'énergie sonore. Le concept de rayon sonore n'est valable que si l'on considère des réflexions spéculaires (on pourra se référer à l'optique géométrique), et que si nous ne tenons pas compte des phénomènes de diffraction et de diffusion. Bien qu'en principe, il soit possible (à l'aide de la méthode des rayons -Chapitre 2.4) de suivre le trajet de rayons sonores se diffractant sur des objets, il n'est pratiquement pas possible de prédire comment l'énergie du rayon incident est distribuée dans l'espace à chaque réflexion. L'acoustique géométrique se limite donc à la propagation en ligne droite du son et aux réflexions spéculaires, comme en optique. Cette approche ne vaut que si les surfaces de réflexions sont suffisamment grandes et que les dimensions des aspérités de ces surfaces sont petites comparativement aux longueurs d'ondes du son incident. Ces conditions sont difficilement valables dans la réalité, et certainement pas pour toutes les fréquences. Bien que l'approche géométrique de l'acoustique des salles puisse être mise en doute, elle n'en reste pas moins très utile. En effet, cette approche est surtout appliquée lorsque l'on ne s'intéresse qu'aux trajectoires des rayons ou aux premières réflexions.

### 2.2.2.2 Propagation d'une impulsion sonore

Considérons l'exemple d'une salle parallélépipédique dans laquelle est placée une source sonore. Le son issu de la source émettrice se propage uniformément dans toutes les directions de l'espace. Avant d'atteindre un point de réception, le son emprunte une infinité des trajets. Dans le cas où l'on admet uniquement l'existence de réflexions spéculaires, il est possible d'identifier un certain nombre de ces trajets et d'établir les temps d'arrivée de *retours* (Fig. 2.4). Un retour est défini comme un rayon sonore qui a subi au moins une réflexion sur son trajet entre le point source et le point de réception.

La source  $S$  émet au temps  $t = 0$  une impulsion sonore dans une salle de volume  $V$ . Au point de réception  $M$  est placé un microphone omnidirectionnel qui capte les rayons incidents. Arrivent successivement :

- l'onde directe atténuee par la divergence sphérique et l'absorption de l'air. Ce signal atteint le point  $M$  au temps  $t_D = SM/c_0$  ( $c_0 \approx 340m.s^{-1}$  étant le célérité du son dans l'air) ;
- le premier retour, issu du rayon qui a suivi le chemin le plus rapide entre les points  $S$

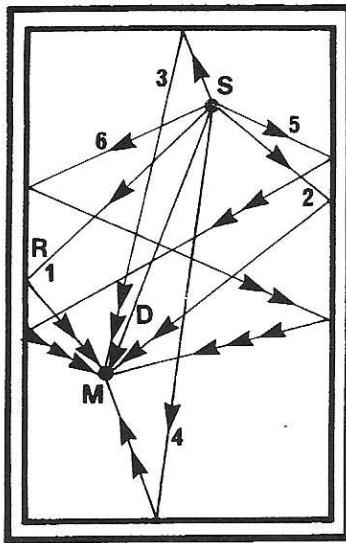


FIG. 2.4 – Représentation bidimensionnelle du trajet de l'onde directe et des premiers retours dans la salle parallélépipédique de la Fig. 2.2 dont les parois ont une impédance infinie [11].  $S$  est la source sonore,  $M$  le point de réception,  $D$  le trajet direct de l'onde entre  $S$  et  $M$ ,  $R_1$  est un rayon sonore subissant une réflexion.

et  $M$ , et ayant subi une réflexion. Ce signal est atténué par la divergence sphérique et l'absorption de l'air et de la paroi rencontrée par le rayon sonore. Il atteint le point  $M$  au temps  $t_1 = (SR + RM)/c_0$ . Rappelons ici que d'après Fermat, toute onde qui se propage entre un point source et un point de réception emprunte le chemin le plus rapide, et non le plus court [12] (p.9) ;

- les retours fournis par les réflexions d'ordre 1 (réflexion sur une seule paroi), d'ordre 2 (réflexions sur deux parois), etc. Ces signaux constituent une série d'impulsions dont la densité croît avec le temps, tandis que leur amplitude décroît avec la distance parcourue, le nombre de réflexions et la dissipation. Ces retours arrivent successivement aux temps  $t_2, t_3, \dots, t_n$ . Lorsque la densité de retours devient importante, il n'est plus possible de différencier les différentes composantes du signal. Le signal obtenu à partir de cet instant au point  $M$  est appelé champ diffus.

La Fig. 2.5 illustre la distribution temporelle des impulsions arrivant au point de réception  $M$ . Nous distinguons :

- le son direct (ou champ direct) dont l'amplitude et le temps d'arrivée dépendent des caractéristiques de la source (niveau de puissance et directivité) et la distance source/récepteur ;
- le champ réverbéré qui comprend par définition [11] :
  - le champ des retours successifs donc l'ensemble forme la texture du son ;
  - le champ diffus qui, en première approximation, est isotrope et homogène dans tout le volume de la cavité à un instant donné. C'est-à-dire qu'il est constitué d'une infinité d'ondes planes se propageant dans toutes les directions de l'espace.

Les temps d'arrivée  $t_1, t_2, \dots, t_n$  dépendent de la distance source/récepteur et de la position relative de la source et du récepteur par rapport aux parois, donc du volume de la salle.

**Remarque.** En acoustique architecturale, des phénomènes de focalisation et des échos sont parfois observés. Les réflexions sur des surfaces courbes n'engendrent pas des images ponctuelles,

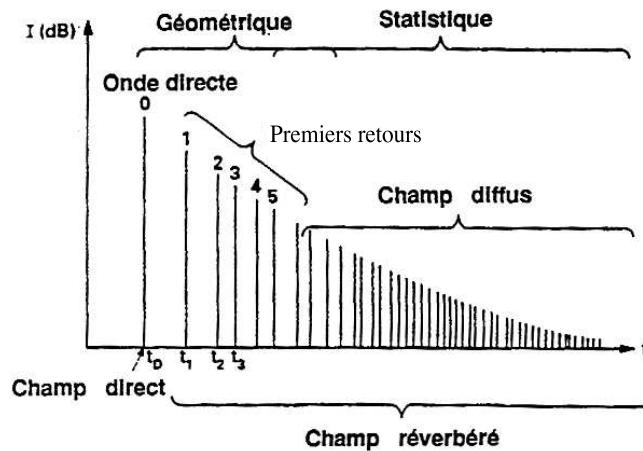


FIG. 2.5 – Représentation schématique de la distribution temporelle des retours captés en un point  $M$  dans une salle [11]. La source émet au temps  $t = 0$  une impulsion.

mais des images étendues ou diffuses (caustiques réelles ou virtuelles). Les phénomènes de focalisation correspondent à une concentration de l'énergie sonore dans une zone limitée de la salle. Les traitements acoustiques les plus couramment employés sont basés sur l'absorption, la diffusion et la transmission. Quant aux échos, ils sont perçus lorsque deux sons identiques se succèdent dans un intervalle de temps supérieur à environ 50ms. Dans le cas contraire, le système auditif humain ne perçoit qu'une seule image sonore au lieu de deux. C'est le phénomène de fusion. L'écho est le plus souvent induit par le son direct et le premier retour, mais il peut également intervenir en présence de deux retours de niveaux intenses (intensité supérieure de 6 à 8dB par rapport aux autres retours). Certains échos peuvent donner lieu à une perception plus ou moins accentuée de hauteur (écho tonal et flutter echo) [11].

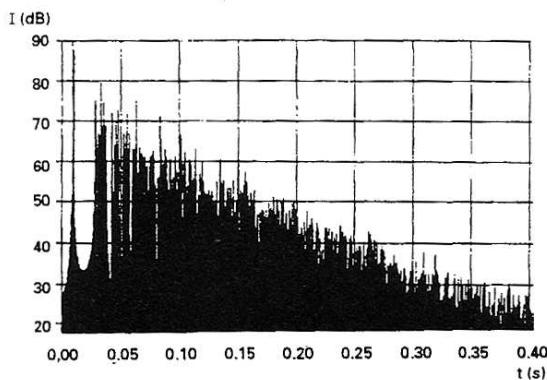


FIG. 2.6 – Croissance et décroissance de l'énergie sonore dans une salle [11]. La source émet au temps  $t = 0$  un signal stationnaire de niveau constant.

### 2.2.2.3 Impédance d'une paroi

Les lois de la réflexion spéculaire ne sont applicables que dans certains cas simplificateurs : sources parfaitement omnidirectionnelles, salles rectangulaires pavant l'espace euclidien et dont les parois sont parfaitement réfléchissantes, salles dont les parois ont une impédance infinie. Dans la réalité, ces cas simples ne sont que très rarement rencontrés. En effet, les parois présentent toujours une impédance complexe. Nous donnons ici un aperçu des limites simplificatrices que l'on utilise.

Rappelons que l'impédance acoustique  $Z$  d'une paroi est donnée par  $Z = \rho_0 c_0 z$ , où  $z$  est l'impédance acoustique réduite de la paroi, et  $\rho_0$  est la masse volumique de l'air ( $\rho = 1.2 \text{ kg.m}^{-3}$  à  $20^\circ\text{C}$ ). Pour une onde arrivant en incidence normale sur une paroi, le coefficient de réflexion  $R$  est donné par :

$$R = (Z - \rho_0 c_0)/(Z + \rho_0 c_0) = (z - 1)/(z + 1), \quad (2.1)$$

alors que le coefficient d'absorption  $\alpha$  en incidence normale (indiqué  $n$ ) est donné par :

$$\alpha_n = 1 - |R|^2. \quad (2.2)$$

En posant  $z = r + jx$  dans les Eq. 2.1 et 2.2, nous obtenons l'expression suivante :

$$\alpha_n = \frac{4r}{x^2 + (r + 1)^2}. \quad (2.3)$$

Lorsque l'impédance d'une paroi est infinie ( $r \rightarrow 0$ ) ou nulle ( $r, x = 0$ ), l'absorption (Eq. 2.3) est nulle. De plus, lorsque l'angle d'incidence n'est pas nul, mais égal à  $\theta$  (incidence oblique), l'Eq. 2.1 devient :

$$R(\theta) = (z \cos(\theta) - 1)/(z \cos(\theta) + 1), \quad (2.4)$$

et l'Eq.(2.3) devient :

$$\alpha(\theta) = \frac{4r}{(x \cos(\theta))^2 + (r \cos(\theta) + 1)^2}. \quad (2.5)$$

L'Eq. 2.5 montre que le coefficient d'absorption est une fonction de l'angle d'incidence du rayon sonore.

Le cas de l'incidence rasante est intéressant car il se rencontre fréquemment. L'une des conséquences de cette configuration est l'effet d'inversion produit dans le plan d'écoute. La théorie (Eq. 2.1) et l'expérience [12] montrent que lorsque que l'angle d'incidence  $\theta$  tend vers  $90^\circ$ , le coefficient d'absorption tend vers 0. Il y a alors réflexion totale de l'onde incidente, avec un déphasage de  $180^\circ$ . L'auditeur se trouve alors en présence de deux ondes en opposition de phase (onde incidente et onde réfléchie), dont la conséquence est un affaiblissement important du niveau sonore.

Notons qu'une paroi d'impédance nulle (ou quasi nulle) correspond à une fenêtre ouverte ou un tissu léger (ou acoustiquement transparent) présents dans une salle (bien que l'on ne rencontre que très rarement ce genre d'éléments).

Lorsque les ondes arrivent en incidence normale sur une paroi, nous observons l'apparition d'ondes stationnaires. La vitesse et la pression deviennent nulle et maximum respectivement au niveau des parois d'impédance infinie. Cependant, les matériaux utilisés dans les salles ont une impédance de module fini. Le système d'ondes stationnaires est alors différent de celui décrit

précédemment. Lorsque l'impédance est réelle et finie, la pression est maximum au niveau des parois, mais l'amplitude des ondes décroît avec la distance à ces parois. En revanche, pour une impédance imaginaire, les parois sont mises en vibration par la pression acoustique et la vitesse devient en quadrature de phase avec la pression. Le maximum de pression se trouve alors à une certaine distance des parois.

#### 2.2.2.4 Modes propres d'une cavité

Considérons à présent une salle parallélépipédique de dimension  $L_x$ ,  $L_y$  et  $L_z$ . On montre aisément que les fréquences propres de la salles sont données par la relation [10] :

$$f_{lmn} = \frac{c_0}{2} \times \sqrt{\frac{l^2}{L_x^2} + \frac{m^2}{L_y^2} + \frac{n^2}{L_z^2}} \quad (2.6)$$

Nous distinguons :

- les modes propres fondamentaux des trois systèmes formés par les parois parallèles deux à deux ( $f_{100}$ ,  $f_{010}$ ,  $f_{001}$ ) et leurs harmoniques ( $f_{l00}$ ,  $f_{0m0}$ ,  $f_{00n}$ ) (modes axiaux) ;
- les modes combinés correspondant aux systèmes d'ondes stationnaires perpendiculaires à l'une des parois ( $f_{lmo}$ ,  $f_{l0n}$ ,  $f_{0mn}$ ) (modes tangentiels) ;
- les modes obliques tels que  $l$ ,  $m$ ,  $n$  sont tous différents de 0.

Signalons que ces modes sont perceptivement importants aux basses fréquences, car ils constituent un réseau de raies discrètes qui donnent à la salle une coloration particulière qui peut se révéler gênante dans certains cas. On assiste alors à un déséquilibre tonal, à un bourdonnement et à un masquage de certaines régions du spectre. Weyl [13] montre que la densité des modes d'une salles est une fonction quadratique de la fréquence. On distingue plus précisément deux régions :

- aux basses fréquences, les modes sont discrets ;
- au-delà d'une certain fréquence (fréquence de Schroeder -Section 2.3.1.1), la densité de mode devient apparemment constante<sup>1</sup> et les modes statistiquement distribués [14].

L'inconvénient des ondes stationnaires est une inégalité de la répartition de l'énergie dans la salle. Afin de parer à cet inconvénient, les architectes de salles de spectacles ont pour habitude de concevoir des salles de forme irrégulière.

### 2.3 Théorie statistique de l'acoustique des salles

La Section 2.2 décrit les phénomènes physiques à partir desquels il serait possible, en principe, de prédire en utilisant la méthode des rayons (voir aussi Chapitre 2.4) le chemin suivi par chaque rayon sonore émis depuis un point source, ainsi que leur amplitude et leur temps d'arrivée en un point de la salle, connaissant la géométrie de la salle et l'emplacement depuis lequel ils ont été émis. On imagine aisément que ce calcul serait fastidieux, étant donné qu'il faudrait prendre en compte un grand nombre de rayons, et approximatif sachant qu'il faudrait discrétiliser l'espace et le temps. De plus, une description totale d'une salle demande de calculer la marche des rayons pour chaque emplacement de source et de récepteur, puisque dans chaque salle l'acoustique est différente d'un siège à un autre. Le spectateur que l'on voit, avant une représentation, choisir avec attention son siège -et le prix- le sait bien.

En dépit des variations de la qualité acoustique qui existe d'une place à une autre dans une salle, il est néanmoins possible d'estimer et/ou de prédire en moyenne l'acoustique d'un lieu clos.

---

<sup>1</sup>En réalité, cette densité croît en  $f^2$ .

Cette partie s'intéresse donc au moyen d'estimer de manière statistique l'acoustique d'une salle, en faisant abstraction des emplacements particuliers.

Il revient au physicien Wallace Clement Sabine [2] d'avoir établi que la *réverbération* (c'est-à-dire la persistance plus ou moins longue d'un son dans une salle après l'extinction de la source sonore) est une quantité importante caractérisant l'acoustique d'une salle.

### 2.3.1 Domaine fréquentiel et temporel

Cette partie vise à mettre en évidence le fait qu'au-delà d'un certaine densité de retours (ou dualement dans l'espace fréquentiel, au-delà d'une certaine densité modale), une approche statistique de la réponse de la salle est plus précise pour décrire le champ réverbéré. En-deçà de ces valeurs de transitions, l'approche statistique n'est pas ou peu pertinente (Fig. 2.5).

#### 2.3.1.1 Densité modale. Fréquence de Schroeder [14, 15, 16]

Considérons un parallélépipède rectangle de dimension  $L_x$ ,  $L_y$  et  $L_z$ , de volume  $V = L_x \times L_y \times L_z$ , dont les parois sont parfaitement réfléchissantes. L'origine du repère cartésien est choisie dans un des angles de la cavité. On montre que les fonctions propres  $\Phi_n^0(\vec{r})$  du système sont données par :

$$\Phi_n^0(\vec{r}) = \cos\left(\frac{n_x \pi x}{L_x}\right) \times \cos\left(\frac{n_y \pi y}{L_y}\right) \times \cos\left(\frac{n_z \pi z}{L_z}\right), \quad (2.7)$$

$$T(t) = \cos(\omega t + \phi_t), \quad (2.8)$$

avec les valeurs propres données par :

$$k_n^0 = \left(\frac{n_x \pi x}{L_x}\right)^2 \times \left(\frac{n_y \pi y}{L_y}\right)^2 \times \left(\frac{n_z \pi z}{L_z}\right)^2, \quad (2.9)$$

et les fréquences propres :

$$f_n^2 = \left(\frac{k_n^0 c_0}{2\pi}\right)^2 = \left(\frac{n_x c_0}{2L_x}\right)^2 + \left(\frac{n_y c_0}{2L_y}\right)^2 + \left(\frac{n_z c_0}{2L_z}\right)^2. \quad (2.10)$$

Le nombre de fréquences propres  $N_m(f)$  inférieures à une fréquence donnée est [?]:

$$N_m(f) = \frac{4\pi V}{3c_0^3} f^3 + \frac{\pi S}{4c_0^2} f^2 + \frac{L}{8c_0} f. \quad (2.11)$$

La densité modale  $D_m(f)$  est quant à elle donnée par :

$$D_m(f) = \frac{\Delta N_m(f)}{\Delta f} = \frac{4\pi V}{c_0^3} f^2 + \frac{\pi S}{2c_0^2} f + \frac{L}{8c_0}. \quad (2.12)$$

L'Eq. 2.12 peut être approximée au premier terme pour des valeurs de  $f$  suffisamment grandes, soit :

$$D_m(f) = \frac{4\pi V}{c_0^3} f^2. \quad (2.13)$$

On montre que les précédentes expressions se généralisent pour des salles aux parois non rigides et de forme quelconque [17]. La quantité  $D_m(f)$  est donc proportionnelle au carré de la fréquence. Cependant, sa valeur devient trop faible aux basses fréquences pour que l'on puisse

admettre la validité de l'hypothèse d'un champ diffus homogène dans tout l'espace. On peut alors calculer la fréquence à partir de laquelle une approche statistique de l'acoustique des salles devient valable : c'est la fréquence de Schroeder ( $f_{Sch}$ ) [14, 18]. Au-delà de cette fréquence, la distribution des modes devient statistique et en apparence constante<sup>2</sup>. Pour donner une évaluation quantitative de cette limite, nous introduisons l'indice de recouvrement modal  $M$ , tel que :

$$M = B_{-3} D_m(f), \quad (2.14)$$

où  $B_{-3}$  est la largeur de bande d'un mode à  $-3dB$  au-dessous du maximum.

Si  $\tau$  est la constante de la salle ( $\tau = T_R/13.8$ ), la largeur de bande à  $-3dB$  est l'inverse de  $\tau$  :  $\Delta f_{-3} = \frac{1}{\tau}$ . On en déduit donc que :

$$B_{-3} = \frac{13.8}{T_R}. \quad (2.15)$$

Schroeder propose de prendre  $M = 3$  [18], c'est-à-dire qu'il propose d'admettre que la largeur moyenne entre deux pics voisins doit être inférieure à  $B_{-3}/3$ . Par conséquent, la fréquence à partir de laquelle le spectre de la réponse impulsionale de la salle devient distribué statistiquement est donnée par :

$$f_{Sch} = \sqrt{\frac{\pi T_R c_0^3}{4 \times \ln(10) \times \pi V}} \approx 2000 \times \sqrt{\frac{T_R}{V}}, \quad (2.16)$$

où  $f_{Sch}$  est exprimé en  $Hz$ ,  $c_0$  est la célérité du son en  $m.s^{-1}$ ,  $T_R$  est le temps de réverbération en  $s$ , et  $V$  est le volume de la salle en  $m^3$ .

L'ordre de grandeur de la fréquence de Schroeder pour une salle de  $1000m^3$  ayant un temps de réverbération de  $1s$  est de  $63Hz$ . Le nombre de fréquence propres inférieures à  $20kHz$  est de l'ordre de  $10^9$ , selon l'Eq. 2.11. Cette dernière valeur montre qu'il est beaucoup trop long de calculer la réponse de la salle à chaque mode pour  $f \leq 20kHz$ . Ces pourquoi les modèles numériques de synthèse de réponse de salle basés sur une approche modale ne sont applicables qu'aux basses fréquences. Aux hautes fréquences, une approche statistique de la réponse et des phénomènes mis en jeu est non seulement plus rapide, mais aussi mathématiquement justifiée.

**Remarque** Notons que les modes d'une salle se recouvrent à condition qu'ils possèdent une certaine largeur, ou encore, si les parois sont absorbantes. On remarque ici une incompatibilité entre la définition du champ diffus, valable pour des parois parfaitement réfléchissantes, et la propriété de diffusion en fréquence. Cependant, l'hypothèse du champ diffus est expérimentalement vérifiée dans les salles de concert.

### 2.3.1.2 Densité de retours. Temps de mélange

La Section précédente montre que la partie diffuse d'une réponse impulsionale de salle est statistiquement distribuée et qu'elle peut être modélisée par un bruit blanc Gaussien. Sachant que la transformée de Fourier d'un tel signal est un bruit blanc Gaussien, il est naturel de vouloir synthétiser la partie diffuse de la réponse par un tel bruit. Moorer [19] a vérifié à l'aide de tests perceptifs la pertinence de synthétiser une réponse impulsionale par un bruit blanc Gaussien pondéré par une exponentielle décroissante dont la constante de temps dépend de la fréquence.

---

<sup>2</sup>L'Eq. 2.13 montre que la densité est une fonction quadratique de la fréquence.

Dans sa thèse d'état [9], Polack présente la réponse impulsionale comme le produit de deux fonctions dont l'une varie lentement, et l'autre beaucoup plus rapidement. Les variations lentes du spectre de la réponse  $h(t)$  sont liées aux variations rapides du signal temporel  $s(t)$  auquel est attaché une échelle de temps rapide notée  $\tau$ , où :

$$h(t) = s(t)e^{-\delta(f)t}. \quad (2.17)$$

Les variations rapides du spectre de  $h(t)$  sont liées à la variation lente de l'exponentielle (Eq. 2.17) à laquelle est attachée une échelle de temps notée  $t$ . Alors que la variable  $t$  est de l'ordre de la milliseconde,  $\tau$  est de l'ordre de la seconde. On retrouve ici les observations faites par Schroeder dans le domaine spectral. Ainsi, Polack [9] modélise la réponse impulsionale comme suit :

- la réponse impulsionale suit une loi de Gauss centrée à chaque instant,
- le spectre moyen de Wigner-Ville de  $h(t)$  est donné par :

$$\langle W_h(t, f) \rangle = \int_{-\infty}^{+\infty} \langle h(t - \tau/2).h(t + \tau/2) \rangle .e^{-2j\pi f\tau} d\tau, \quad (2.18)$$

où  $\langle W_h(t, f) \rangle = F^2(t, f)$  avec  $F(t, f) = B(f)e^{-\gamma(f)t}$  qui est une fonction qui varie lentement avec  $t$  et  $f$ .

La précédente Section montre qu'il existe une fréquence de transition entre deux domaines du spectre, un en basses fréquences où les modes sont "discrets" et l'autre en plus hautes fréquences où la densité de modes devient constante. Dans le domaine temporel, il existe également un temps de transition entre deux états du système, il s'agit du *temps de mélange*.

Considérons une salle dont les parois sont réfléchissantes, dans laquelle on place une source qui émet une impulsion dans toutes les directions de l'espace à l'instant  $t = 0$ . Comme nous l'avons vu précédemment, cette source génère une infinité de sources-images qui émettent toutes simultanément le même signal. Ce phénomène est représenté dans la Fig. 2.7.

Dans l'intervalle de temps  $[t_1 : t_1 + \Delta t]$ , les impulsions reçues au point  $M$  sont celles qui sont issues des sources-images comprises entre les cercles de rayon  $c_0 t_1$  et  $c_0(t_1 + \Delta t)$ . Autrement dit, ces sources sont situées dans le volume :

$$\Delta V_S = \frac{4}{3} \times \pi c_0^3 \times [(t_1 + \Delta t)^3 - t_1^3]. \quad (2.19)$$

En négligeant les termes en  $\Delta t^3$ , l'équation précédente devient :

$$\Delta V_S = 4\pi c_0^3 (t_1^2 \Delta t + t_1 \Delta t^2). \quad (2.20)$$

Si la salle considérée a un volume  $V$ , alors le nombre de retours qui arrivent en  $M$  dans l'intervalle de temps  $\Delta t$  est donné par :

$$\Delta N = \frac{\Delta V_S}{V}. \quad (2.21)$$

Nous en déduisons que la densité temporelle  $D_e(t)$  de retours, entre  $t_1$  et  $t_1 + \Delta t$ , peut s'exprimer en première approximation comme :

$$D_e(t) = \frac{\Delta N}{\Delta t} = \frac{4\pi c_0^3 t_1^2}{V} + \frac{4\pi c_0^3 t_1 \Delta t}{V}. \quad (2.22)$$

Polack propose de définir le temps de mélange comme le temps auquel au moins 10 retours sont compris dans une fenêtre temporelle de 24ms [12]. Cette durée correspond à la durée

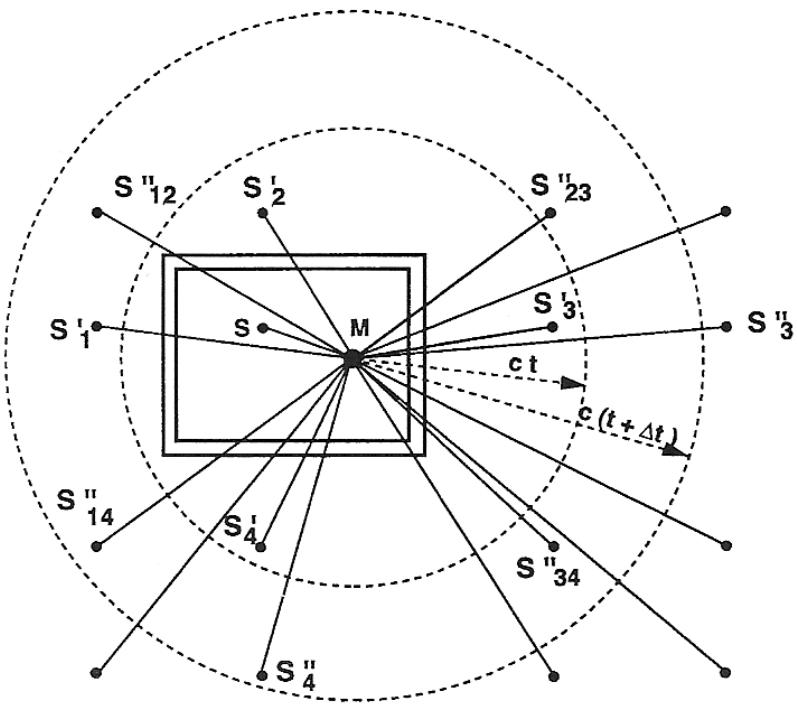


FIG. 2.7 – Représentation bidimensionnelle d'une densité temporelle de sources-images.  $S'_i$ ,  $S''_i$ , ... sont les sources-images que donnent les parois de la source  $S$ . Il est possible de tracer tous les rayons virtuels  $S'_i M$ ,  $S''_i M$ , ... et de définir le nombre  $\Delta N$  de sources comprises entre les cercles de rayons  $c_0 t$  et  $c_0(t + \Delta t)$  [11].

d'intégration du système auditif humain. Connaissant le nombre moyen de retours par unité de temps (en première approximation) :

$$N(t) = \frac{4\pi c_0^3 t^3}{3V}, \quad (2.23)$$

nous obtenons une formulation heuristique du temps de mélange donnée par :

$$T_M = \sqrt{\frac{NV}{4\pi c_0^3 \Delta t}} \approx \sqrt{V}, \quad (2.24)$$

où  $N = 10$  retours,  $\Delta t = 24ms$ ,  $V$  est le volume de la salle en  $m^3$  et  $T_M$  est exprimé en  $ms$ .

Le temps de mélange dépend de l'étalement de l'impulsion sonore émise par la source et de l'échelle d'observation considérée (ici, 24ms). Notons qu'en pratique, cette formule théorique du temps de mélange est surestimée. En effet, d'une part, l'absorption des parois augmente l'étalement temporel des retours à chaque réflexion, et d'autre, une partie de l'énergie réfléchie par les parois l'est de manière diffuse, ce qui conduit à un mélange plus rapide.

Il est utile de s'arrêter quelques instants sur ce qu'est le mélange. Précédemment, nous avons présenté le temps de mélange tel qu'il est proposé par Polack. Dans la littérature [3], nous trouvons également que le mélange est une propriété de certains systèmes dynamiques

qui sont également ergodiques<sup>3</sup>. Le caractère mélangeant d'une salle est associé à la divergence (exponentielle ou non) des rayons sonores émis par une source et initialement adjacents au point d'émission. Ce sont les aspérités des parois, par exemple, qui vont créer ces divergences de trajectoires. Lorsque système est mélangé, la densité de rayons dans l'espace est isotrope et homogène. Les trajectoires deviennent indépendantes de leur origine, c'est-à-dire du point source d'émission. Les rayons ont alors la même probabilité (nous parlons d'équiprobabilité) d'atteindre une région de l'espace des phases. Le théorème d'équipartition de l'énergie est valable pour les systèmes ergodiques et à l'équilibre thermodynamique, pour lesquels tous les états de même énergie ont la même probabilité d'être occupés. Dans un système ergodique, il est possible d'échanger de l'énergie -sous ses différentes formes- dans le système, afin qu'elle puisse se répartir uniformément au cours du temps. Dans le cas des gaz parfaits, ces sont les chocs entre les molécules qui assurent ces échanges d'énergie et qui répartissent l'énergie équitablement en moyenne entre toutes les molécules. Dans le cas de l'acoustique des salles, nous retrouvons l'hypothèse selon laquelle, l'énergie est uniformément répartie dans l'espace lorsque le champ diffus est établi.

Nous sommes ici confrontés à trois définitions du temps de mélange : une, qui est ancrée dans le domaine de l'acoustique des salles et qui définit ce temps comme le temps de transition entre les premiers retours (ou premières réflexions), une qui se base sur des critères perceptifs –celle de Polack– et enfin la dernière qui fait référence à la théorie des systèmes dynamiques. La question que nous nous posons ici, et à laquelle nous répondons par la suite, est : peut-on trouver un ou des estimateurs objectifs du temps de mélange ?

Notons dès à présent que nous ne répondrons pas à cette question selon la définition des systèmes dynamiques. De plus, nous ne pouvons pour l'instant pas dire si la première définition du temps de mélange est liée à la seconde.

Les Chapitres 4-6 répondent à ces questions, et nous discutons nos résultats dans un Chapitre 7 en revenant aux différentes définitions du temps de mélange.

### 2.3.2 Réflexion diffuse

Comme cela a été vu en introduction à ce Chapitre, les rayons sonores qui se propagent dans une salle ne subissent pas uniquement des réflexions spéculaires mais aussi diffuses, causées par les éléments de décoration (balcons, fosse d'orchestre, sièges, etc.). Alors qu'un rayon réfléchi spéculairement repart de la paroi avec le même angle que l'angle d'incidence selon la loi de Snell, l'onde réfléchie de manière diffuse repart suivant plusieurs directions. Dans ce cas, la réflexion peut être modélisée comme la somme pondérée d'une réflexion spéculaire et d'une réflexion totalement diffuse. La pondération est fonction des aspérités de la surface, de l'angle d'incidence et de la fréquence de l'onde incidente. S'il y a diffusion totale, la distribution angulaire des directions de départ des rayons réfléchis est équiprobable et indépendante de l'angle d'incidence [3]. En revanche, l'énergie réfléchie est fonction de l'angle d'incidence et de réflexion.

Au lieu de s'intéresser au cas de chaque rayon, on voudra caractériser la direction de réflexion des rayons par un angle solide. La loi de Lambert permet une modélisation de la réflexion de l'énergie lorsque la réflexion est totalement diffuse (Fig. 2.8).

Soit  $I(r)$  l'intensité reçue en un point distant de  $r$  d'un élément de surface  $dS$  et caractérisé par un angle  $\eta$  à la normale  $n$  :

$$I(r) = I_0 \frac{\cos(\eta_0)\cos(\eta)}{\pi r^2} dS = I_0 \cos(\eta_0) d\Omega, \quad (2.25)$$

---

<sup>3</sup>Notons que l'inverse n'est pas vrai : un système ergodique n'est pas obligatoirement mélangeant, alors qu'un système mélangeant est ergodique.

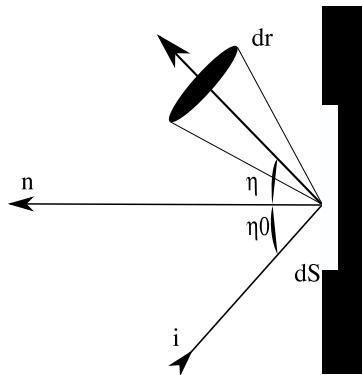


FIG. 2.8 – Réflexion d'un rayon suivant la loi de Lambert.

avec  $dS$  l'élément de surface touché par le rayon sonore,  $\eta_0$  l'angle d'incidence,  $I_0$  l'intensité de rayon incident, et  $d\Omega$  l'angle solide défini par :

$$d\Omega = \frac{\cos(\eta)dS}{r^2}. \quad (2.26)$$

Joyce [3] montre que qu'elles que soient les propriétés des matériaux, si les rayons incidents sur une paroi arrivent de manière diffuse, alors la réflexion est idéalement diffuse. Cremer [12] (p.250) poursuit et conclut : le champ sera toujours à peu près diffus, tant que les réflexions ne seront pas tout à fait uniquement spéculaires, mais qu'elles présenteront un peu de diffusion. Ceci est d'autant plus vrai que le phénomène de diffusion suit un effet "boule de neige" à chaque réflexion. Joyce suppose que les petites irrégularités sur les surfaces d'une salle conduiront fatallement à un état diffus total. La question soulevée par Cremer [12] est de savoir combien de temps et quelle énergie sont nécessaires pour atteindre un champ diffus total. Les Chapitres 4, 5 et 6 tentent de répondre à cette question en proposant des méthodes objectives d'estimation des temps d'arrivée des retours et du temps de mélange, temps auquel les réflexions sont idéalement diffuses.

## 2.4 Méthodes de synthèse de réponses impulsionales de salles

### 2.4.1 Introduction

La synthèse de réponses impulsionales de salles est utilisée lors de projets architecturaux, et plus particulièrement lorsque l'on souhaite entendre (ou auraliser) l'acoustique du lieu avant sa construction. L'idée d'estimer la réponse impulsionnelle d'une salle n'est pas nouvelle. En effet, avant l'avènement de l'informatique, les modèles étaient réalisés en miniaturisant (à l'échelle 1/10<sup>e</sup> ou 1/50<sup>e</sup>) la salle dont il était question (Section 2.6). De nombreuses études ont ainsi été réalisées, et le sont toujours (la Philharmonie de Paris est un exemple récent). Les protocoles

expérimentaux ne sont pas triviaux : choix et calibration de la source et du récepteur, acquisition des données, traitement post-acquisition des données. Les sources employées sont souvent des sinus glissants, ou bien encore des étincelles. Le logiciel MIDAS (et par conséquent OpenMIDAS) [20, 21] (Chapitre 3), par exemple, est un logiciel qui permet de faire l'acquisition de réponses impulsionales à pleine échelle ou dans des modèles réduits et de traiter les données acquises en basculant à pleine échelle grâce au facteur d'échelle employé lors de la réalisation de la maquette, et en compensant l'atténuation de l'air.

Dans les années 1970 [4], les puissances de calcul des ordinateurs deviennent suffisantes et permettent la modélisation informatique de réponses de salles. Même si aujourd'hui les modèles informatiques sont plus rapides que les mesures sur maquettes, ils n'en restent pas moins approximatifs.

L'objet de cette partie n'est pas de confronter deux moyens de modélisation : la maquette et l'informatique. Le lecteur intéressé par la mesure sur maquette est invité à consulter le livre de Cremer [12] (pp.164-185) afin d'avoir une présentation non-exhaustive des moyens et techniques employés, ainsi que des travaux majeurs dans ce domaine, et des principales publications sur ce sujet.

Cette partie présente succinctement les différents modèles informatiques utilisés pour l'estimation de réponses impulsionales de salles (méthode source-image, méthode du lancer de rayons et méthode des cônes), ainsi que leurs limites. Puis, une attention toute particulière est portée à l'approche statistique qui mène à définir un modèle stochastique de réponses impulsionales de salles (Section 2.5).

Ce Chapitre présente en détails les concepts qui mènent à la notion de réponse d'un système tel qu'une salle, à une impulsion. En particulier, il y est expliqué comment les rayons sonores émis par la source dans toutes les directions de l'espace parcourent des chemins différents, appelés *trajectoires*, avant d'atteindre le récepteur qui capte la réponse de la salle au signal d'excitation émis par la source. L'air contenu dans le volume de la salle est homogène et isotrope, par conséquent l'énergie acoustique se propage selon des rayons normaux aux fronts d'onde émis par la source. Ces ondes sonores se propagent en ligne droite jusqu'à rencontrer un obstacle qui leur fait perdre de l'énergie et les filtre selon les propriétés intrinsèques des matériaux qui composent l'obstacle. L'onde est réfléchie et change de direction de propagation, pour suivre une autre trajectoire. Les rayons qui ont suivi différents chemins avant d'atteindre le récepteur sont appelés *retours*. Parmi ces trajectoires, la plus courte est celle empruntée par le *son direct*, retour qui possède le plus souvent le plus d'énergie, et qui constitue le premier retour de la réponse impulsionale. Suivent ensuite -dans le signal temporel de la réponse impulsionale- les retours des rayons qui ont parcouru de plus longues distances que le son direct. Ils ont donc des formes d'onde plus étalées, et sont souvent d'énergie plus faible que le son direct. Nous avons montré précédemment que l'évolution temporelle de l'énergie et du nombre de retours croît selon :

$$N(t) = \frac{4\pi c_0^3}{3V} t^3, \quad (2.27)$$

où  $c_0$  est la célérité du son en  $m.s^{-1}$ , et  $V$  est le volume de la salle en  $m^3$ .

La prédiction de la réponse d'une salle pour un couple de position source/récepteur consiste à déterminer les différents trajets qu'empruntent les rayons sonores émis par la source. A chaque trajectoire est associée une impulsion dont l'énergie est une fonction de la distance parcourue et du nombre de réflexions subies. La somme des contributions de retours forme la réponse impulsionale de la salle. L'Eq. 2.27 met en évidence l'augmentation rapide du nombre de retours avec le temps. Passé un certain temps, le *temps de mélange*, la densité de retours est

beaucoup plus grande qu'aux premiers instants de la réponse, de telle sorte que le calcul de la réponse entière n'est actuellement pas envisageable puisque les puissances des ordinateurs ne sont pas suffisantes. Passé ce temps, la réponse de la salle peut être vue comme un processus aléatoire et peut être estimée de manière globale grâce aux résultats de la théorie statistique de l'acoustique des salles.

Les méthodes de simulation de salles usuelles (présentées en Annexe A) relèvent de l'acoustique géométrique, qui est particulièrement bien adaptée à la détermination des trajectoires des particules sonores dans une salle dont les dimensions sont grandes devant les longueurs d'ondes considérées, ce qui est le cas des salles de spectacles.

En guise de remarque générale, signalons qu'aucune des méthodes usuelles ne synthétisent la diffusion et la diffraction. Cependant, des modèles hybrides proposent différentes approches de ces phénomènes [22, 23, 24, 25, 26].

Les méthodes de synthèses de réponses impulsionales de salles sont présentées dans l'Annexe A.

#### 2.4.2 Remarque générale -Vers une méthode probabiliste

Les méthodes telles que les sources-images, le lancer de rayons ou encore les méthodes hybrides présentent de nombreux inconvénients. En premier lieu vient la nécessité de modéliser en deux ou trois dimensions la géométrie de la salle considérée, ainsi que les matériaux qui la composent. Cette première étape est bien souvent fastidieuse et longue. En second lieu intervient le temps de calcul qui constitue une véritable limite aujourd'hui et qui constraint les utilisateurs à tronquer le signal synthétisé à un certain temps, ou à un certain ordre de réflexion, et à modéliser le reste du signal en se basant sur des considérations statistiques de l'acoustique des salles.

L'idée selon laquelle le signal peut n'être synthétisé à l'aide d'une de ces trois méthodes que jusqu'au temps de mélange -moment où le signal adopte un comportement aléatoire- est illusoire et peu pertinente. En effet, cette idée repose sur l'acceptation sans vérification expérimentale de la formulation du temps de mélange proposée par Polack [9]. En réalité, Polack suppose que l'on peut définir le mélange d'un point de vue perceptif, c'est-à-dire quand 10 retours sont captés dans un laps de temps de  $24ms$ , durée correspondant au temps caractéristique du système auditif humain [12]. Dans ce Chapitre, rappelons que nous avons montré que la formule du temps de mélange s'écrit :

$$T_M \approx \sqrt{V} \quad (2.28)$$

où  $T_M$  est le temps de mélange, exprimé en  $ms$ .

A la connaissance de l'auteur, aucune estimation expérimentale du temps de mélange robuste n'a été présentée jusqu'à présent. Le lecteur intéressé par des méthodes d'estimation du temps de mélange et une exploration expérimentale des propriétés statistiques des réponses impulsionales de salles de spectacles s'orientera vers les Chapitres 4, 5 et 6, qui montrent que la formule théorique du temps de mélange n'est pas exactement vérifiée. Bien que ce temps ait pour limite supérieure  $\sqrt{V}$  ( $ms$ ), il est aussi une fonction de la distance source/récepteur. En d'autres termes, les méthodes d'estimation que nous mettons en œuvre plus loin ne vérifient pas un temps de mélange à peu près égal dans toute une salle.

Dans la continuité de cette Section, ce qui suit reprend brièvement la théorie sous-jacente au mélange et présente le modèle stochastique développé par Polack [27] qui se base sur des probabilités des temps d'arrivée des retours (ayant chacun subi un nombre probable de réflexions) en un point de l'espace en se basant non pas sur la géométrie complexe ou simple de la salle, ni sur les matériaux qui la composent, mais sur le temps de réverbération moyen à  $30dB$ , l'absorption moyenne, la distance source/récepteur et enfin le volume de la salle.

## 2.5 Modélisation stochastique de l'acoustique des salles de spectacles

### 2.5.1 Introduction

Le modèle stochastique présenté dans cette Section a été proposé par J-D. Polack en 1993 [27]. A partir de la théorie statistique de l'acoustique des salles et des systèmes dynamiques que sont les billards, le modèle stochastique dont il est question ici propose une approche probabiliste et statistique du comportement des trajectoires des rayons sonores dans une salle [6, 27, 28]. Reprenant les travaux de Joyce [3], Polack ne s'intéresse plus aux particules sonores (les phonons) mais bien aux trajectoires des rayons, et définit la diffusion comme une propriété de ces rayons et non du champ sonore.

Un rayon sonore est assimilé au parcours d'une particule dans la salle au cours du temps. Cette particule occupe au cours du temps différentes positions et prend différentes directions. On définit alors l'espace des phases comme l'espace qui décrit la position et la direction de la particule à chaque instant. Joyce émet l'hypothèse que les grandes salles sont ergodiques, c'est-à-dire que les trajectoires passent par presque tous les points de l'espace des phases avec la même probabilité, et y séjournent en moyenne le même temps, et mélangeantes. La propriété de mélange assure une diffusion des particules. Rappelons que le mélange assure une existence équitable à une particule en tout point de l'espace des phases.

Dans les salles, ces propriétés sont liées à la géométrie de la cavité et à l'état des surfaces des matériaux. Ainsi, une salle parallélépipédique rectangle n'est pas ergodique (et donc pas mélangeante), car un rayon ne peut emprunter que huit directions possibles.

### 2.5.2 Le modèle stochastique

Les Sections précédentes montrent que l'hypothèse du champ diffus n'est jamais rigoureusement vérifiée dans une salle dont les parois sont absorbantes. Polack montre [6, 27] que les propriétés d'ergodicité et de mélange permettent de retrouver l'expression de la décroissance exponentielle de l'énergie [12]. L'énergie moyenne perdue au bout d'un temps  $t$  par un rayon sonore au cours de multiples réflexions est donnée par :

$$E(t) = E_0 \sum_n R^n P_n(t) = E_0 \sum_n (1 - \alpha)^n P_n(t), \quad (2.29)$$

où  $n$  représente le nombre de réflexions subi par le rayon sonore,  $R$  est l'énergie restituée à chaque réflexion,  $\alpha$  est le coefficient d'absorption de Sabine, et  $P_n(t)$  représente la probabilité qu'un rayon ait subi  $n$  réflexions à l'instant  $t$ .

Dans le cas d'un mélange instantané, les rayons possèdent les propriétés statistiques correspondant à un processus de Poisson [29] :

- la probabilité  $P_n$  d'avoir  $n$  réflexions pendant un intervalle de temps de durée  $t$  ne dépend que de  $t$ , et non de la position de  $t$  sur l'axe temporel ;
- le nombre de réflexions  $n$  se produisant pendant deux intervalles de temps disjoints sont des variables aléatoires indépendantes ;
- la probabilité que deux réflexions ou plus se produisent dans un petit intervalle de temps  $\Delta t$  est négligeable par rapport à la probabilité qu'il n'y ait qu'une seule réflexion. Ou d'une manière plus formelle :

$$Q_n(t) = \begin{cases} o(\Delta t), & (k \geq 2) \\ \lambda t + o(\Delta t), & (k = 1) \\ 1 - \lambda t + o(\Delta t), & (k = 0). \end{cases} \quad (2.30)$$

On montre alors que la probabilité  $P_n(t)$  s'écrit [29] :

$$P_n(t) = \frac{(\lambda t)^t}{n!} e^{-\lambda t}. \quad (2.31)$$

Le paramètre  $\lambda$  (ou intensité du processus de Poisson) a pour expression [6, 27] :

$$\lambda = \frac{Sc_0}{4V} = \frac{c_0}{\bar{l}}, \quad (2.32)$$

où  $c_0$  est la célérité du son,  $S$  la surface des parois,  $V$  le volume de la salle et  $\bar{l}$  le libre parcours moyen.

On obtient alors immédiatement l'expression de l'énergie dans la salle  $E(t)$  :

$$E(t) = E_0 e^{-\alpha c_0 t / \bar{l}} = E_0 e^{-\alpha S c_0 t / 4V}. \quad (2.33)$$

On retrouve ici l'expression obtenue dans l'hypothèse du champ diffus. Notons cependant que si cette relation est valable pour un rayon, elle ne l'est pour l'énergie de la salle que si l'hypothèse des salles ergodiques est valide.

### 2.5.3 Améliorations proposées

Le modèle stochastique proposé par Polack synthétise des réponses impulsionales hautes fréquences, sans filtrage de la salle et sans diffusion. Cette Section présente trois ajouts au modèle :

- synthèse de la diffusion ;
- synthèse du filtrage de la salle ;
- synthèse des angles d'incidence des rayons sonores.

Ces ajouts, qui sont qualifiés d'*améliorations* dans le titre de cette Section, peuvent être utiles d'une part, si l'on souhaite approcher une validité écologique du modèle, et d'autre part, si l'on souhaite tester des méthodes d'analyse et d'estimation de paramètres (tels les retours -Chapitre 5) sur un modèle *proche* de la réalité. Rappelons ici que des paramètres de signaux qui sont jugés utiles du point de vue physique ne le seront pas obligatoirement du point de vue perceptif. En d'autres termes, toute l'information n'est pas contenue dans le signal. On peut tout aussi bien dire que toute l'information contenue dans un signal n'est pas forcément pertinente perceptivement. Une bonne illustration de ces propos sont les tests d'écoute conduits par Moorer dont le résultat est qu'un bruit blanc filtré et exponentiellement décroissant est perçue comme une réponse de salle réaliste [19].

Les travaux présentés dans la suite de ce manuscrit permettront de juger du réel apport de ces ajouts et conclure quant à l'amélioration qu'ils constituent.

#### 2.5.3.1 Synthèse de la diffusion

Nous ne nous intéressons ici qu'aux réflexions spéculaires et diffuses qui induisent un filtrage du rayon incident sur la paroi. La diffusion apparaissant après la première réflexion, nous ne pouvons l'observer qu'après le premier retour dans le signal de la réponse impulsionale. Chaque particule sonore est porteuse d'une énergie élémentaire. Nous pouvons définir dans l'espace des phases [30, 31] :

- $w(\vec{r}, t)$  la densité d'énergie définie par :

$$w(\vec{r}, t) = \int \int \int f(\vec{r}, \vec{v}, t) d^3 v; \quad (2.34)$$

–  $\vec{J}(\vec{r}, t)$  le flux d'énergie, défini par :

$$\vec{J}(\vec{r}, t) = \int \int \int \vec{v}.f(\vec{r}, \vec{v}, t) d^3 v, \quad (2.35)$$

où  $f(\vec{r}, \vec{v}, t)$  est la fonction de distribution des particules, autrement dit, le nombre de particules de vitesse<sup>4</sup>  $\vec{v}$  en  $\vec{r}$ , à l'instant  $t$  :

$$f(\vec{r}, \vec{v}, t) \approx \frac{1}{4\pi} w(\vec{r}, t) + \frac{3}{4\pi ||v||^2} \cdot \vec{v} \cdot \vec{J}(\vec{r}, t). \quad (2.36)$$

L'Eq.2.36 comprend également un terme qui est négligé ici car  $\vec{v}$  est indépendant du temps et de l'espace.

De plus, le flux d'énergie est relié à la densité par le coefficient de diffusion  $D$  par la relation :

$$\vec{J}(\vec{r}, t) \approx -\frac{D}{2} \vec{\nabla} w(\vec{r}, t), \quad (2.37)$$

où,  $\vec{\nabla} w(\vec{r}, t) = -\frac{\vec{r}}{Dt} \cdot \vec{w}(\vec{r}, t)$ , et  $D = \bar{l}c_0/3$  ( $\bar{l}$  étant le libre parcours moyen).

On peut donc réécrire l'Eq. 2.36 :

$$f(\vec{r}, \vec{v}, t) \approx \frac{1}{4\pi} \left[ w(\vec{r}, t) - \frac{3}{c_0^2} \frac{D}{2} \vec{v} \cdot \vec{J}(\vec{r}, t) \right]. \quad (2.38)$$

De plus, la densité d'énergie est décrite (pour les trois dimensions de l'espace) par la formule [30] :

$$w(\vec{r}, t) = \frac{1}{(2\pi Dt)^{3/2}} e^{-\frac{r^2}{2Dt} - \sigma t}, \quad (2.39)$$

où  $\sigma$  est un terme relatif à l'absorption moyenne de la salle :  $\sigma = c_0 \times \alpha_{Sabine}/\bar{l}$ .

Après calculs, il vient :

$$f(\vec{r}, \vec{v}, t) \approx \frac{w(\vec{r}, t)}{4\pi} \left[ 1 + \frac{3}{2c_0 t} \times r \times \cos(\theta) \right], \quad (2.40)$$

où  $\theta$  est l'angle formé entre les vecteurs  $\vec{r}$  et  $\vec{v}$ .

Dans la pratique, le modèle détermine de manière équiprobable l'angle  $\theta$  de telle sorte que l'on obtient la probabilité d'existence de particules de diffusion au point de réception. Notons que ces particules ne peuvent exister au point d'observation que si  $r \geq t$ .

### 2.5.3.2 Synthèse du filtrage de la salle

Le modèle stochastique est un modèle hautes fréquences. De plus, le trajet des rayons dans la salle, ainsi que les matériaux qui composent les parois et les angles d'incidence des rayons sur ces parois ne sont pas calculés. Dans sa version d'origine, le modèle de Polack ne calcule donc pas le filtrage des rayons sonores par les parois de la salle.

Les temps d'arrivée des rayons en un point de l'espace sont calculés selon un processus de Poisson dont le paramètre est le nombre moyen de retours par intervalle de temps (Eq. 2.27). Chaque rayon subi un nombre de réflexions calculé également selon un processus de Poisson (Eq. 2.31). En définissant un certain nombre de filtres, à chaque réflexion est associée aléatoirement un filtre, qui peut être plusieurs fois le même. Le nombre de filtres, ainsi que leur spectre sont fixés arbitrairement ici.

Trois familles de filtres sont élaborées :

---

<sup>4</sup>La vitesse de propagation des particules sonores est la célérité du son dans l'air à 20°C, soit  $c_0 = 340 m.s^{-1}$ .

### 1. "Materiaux" (passe-bas/passe-bande/passe-haut)

Les filtres de cette famille sont basés sur une grossière approximation des matériaux que l'on peut rencontrer dans une salle : matériaux absorbant les hautes fréquences, passe-bande et absorbant les basses fréquences du son incident. Arbitrairement, nous choisissons trois filtres à réponses impulsionales finies. La fréquence de coupure du filtre passe-bas est fixée à  $700Hz$ ; la bande passante du filtre passe-bande est fixée à  $1000 - 4000Hz$ ; la fréquence de coupure du filtre passe-haut est fixée à  $6000Hz$ .

### 2. Bandes d'octave :

Les filtres de sept bandes d'octaves sont choisis :  $125Hz$ ;  $250Hz$ ;  $500Hz$ ;  $1000Hz$ ;  $2000Hz$ ;  $4000Hz$ ;  $8000Hz$ .

### 3. Variations Spectrales Aléatoires (VSA) :

Cette approche se base sur le spectre de l'impulsion émise par la source. Il ne s'agit pas d'agir sur l'énergie du spectre, mais sur les fréquences auxquelles le spectre prend ses valeurs. En d'autres termes, le filtre VSA agence dans un ordre aléatoire les valeurs du spectre. Dans un premier temps, il s'agit de trouver les fréquences auxquelles le spectre n'est pas nul. Puis, dans un second temps, le spectre de l'impulsion est modifié en choisissant aléatoirement, et de manière équiprobable, les fréquences auxquelles le spectre prend ses nouvelles valeurs. Les variations se font selon un processus normal, centré sur la valeur initiale et avec un écart-type que l'on fixe arbitrairement. Le choix de ce dernier paramètre est issu de l'observation et de l'écoute. En effet, pour des variations très élevées, les variations du spectre sont trop grandes et donc le filtrage de la salle paraît soudain et peu naturel.

Lorsqu'un retour subi deux réflexions, il subit par conséquent deux filtrages. Pour chacune de ces réflexions, la famille de filtres ainsi que le filtre sont choisis de manière équiprobable. Le second filtrage se fait évidemment à partir du retour ayant subi un premier filtrage. Une voie de recherche future possible est l'estimation du filtrage global de la salle subi par chaque rayon avant d'atteindre le récepteur, et ce à l'aide de méthodes d'estimation des retours (Chapitre 5). Une large étude menée sur de nombreuses mesures et dans de nombreuses grandes salles permettrait peut-être d'identifier des tendances de filtrage, et donc de ne plus opérer comme nous l'avons fait ici de manière arbitraire.

Les Fig. 2.9-2.13 présentent les spectres de chaque filtre du modèle. Les Fig. 2.14-2.17 montrent, quant à elles, quelques exemples de réponses impulsionales générées par le modèle ainsi que leur spectrogramme.

#### 2.5.3.3 Synthèse en 2 et 3 dimensions

La restitution sonore (ou auralisation [32]) de l'acoustique d'une salle en deux (stéréophonie) ou trois dimensions (Ambisonics, binaurale, etc.) n'est possible que si le modèle de réponses impulsionales utilisé génère les angles d'incidence des rayons sonores en un point de l'espace. Selon que la restitution est stéréophonique ou faite en trois dimensions, seul l'azimut ou l'azimut et le site doivent être calculés respectivement.

Dans les modèles des source-images et du lancer de rayons (Section 2.4), ces angles d'incidence sont calculés de manière déterministe. Ceci est possible car ces modèles tiennent compte de l'architecture de la salle (avec plus ou moins de détails). En revanche, dans l'approche probabiliste qui est celle du modèle présenté, la géométrie de la salle n'est pas définie. Autrement dit, le modèle stochastique génère des salles aléatoires. La connaissance des angles d'incidence des rayons au point de réception est par conséquent impossible. Dans ce modèle, on choisit une distribution aléatoire de ces angles, basée sur des observations faites à partir de mesures réalisées

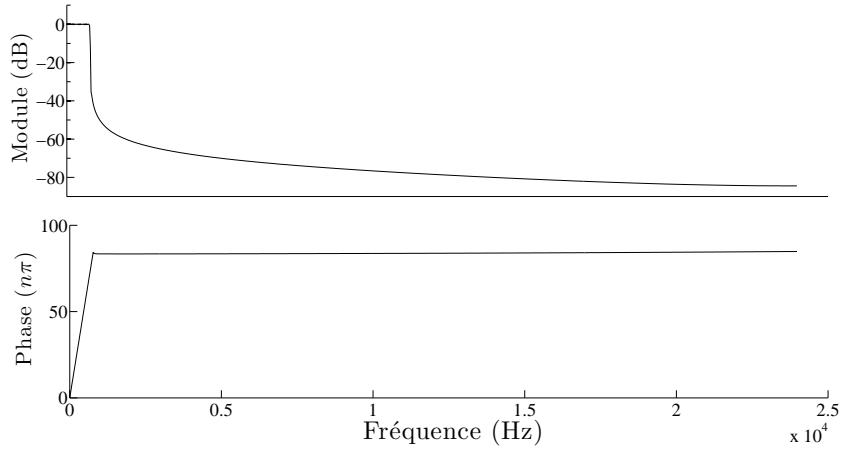


FIG. 2.9 – Spectre (module et phase) du filtre type matériaux passe-bas. Fréquence de coupure : 700Hz. Spectre calculé sur  $N = 2048$  points.

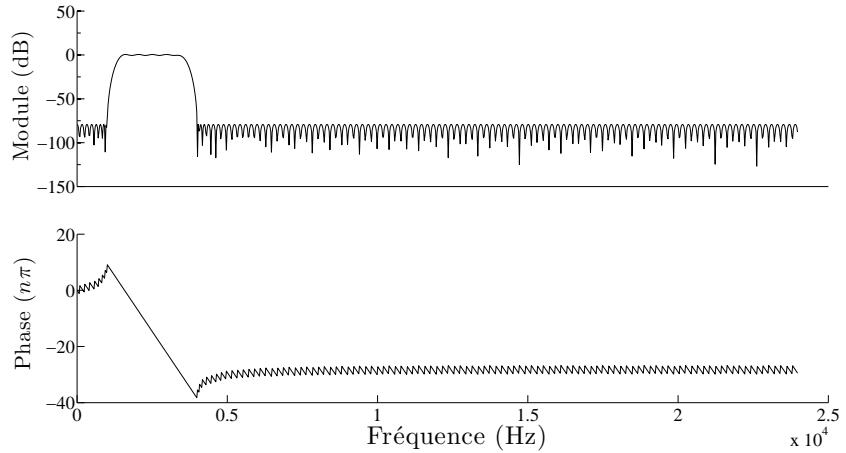


FIG. 2.10 – Spectre (module et phase) du filtre type matériaux passe-bande. Bande passante : 1000 – 4000Hz. Spectre calculé sur  $N = 2048$  points.

dans des salles de spectacles.

Le Chapitre 5 présente une méthode d'estimation des retours captés en un point de l'espace d'une salle [33]. Cette méthode est appliquée à des mesures de réponses impulsionales de la Salle Pleyel réalisées à l'aide d'un microphone Soundfield Ambisonics composé de plusieurs capsules qui captent le champ de pression en trois dimensions (plus le champ omnidirectionnel) en le décomposant selon les quatre premières harmoniques sphériques. A partir des signaux enregistrés par le microphone Ambisonics, il est possible d'estimer l'azimut et le site d'incidence des rayons dont les temps d'occurrence ont été estimés (Chapitre 5, Section 5.3.5). Ce mémoire ne traitant pas de la distribution angulaire des rayons en un point de l'espace au cours du temps, aucune étude approfondie n'a été menée sur ce sujet. Cependant, quelques observations montrent que la distribution des angles d'incidence suit une loi normale centrée autour de 0, lorsque le récepteur fait face à la source, tant pour l'azimut que pour le site. Dans cette dernière configuration, le son direct atteint le récepteur avec un azimut et un site égaux à 0°. Notons

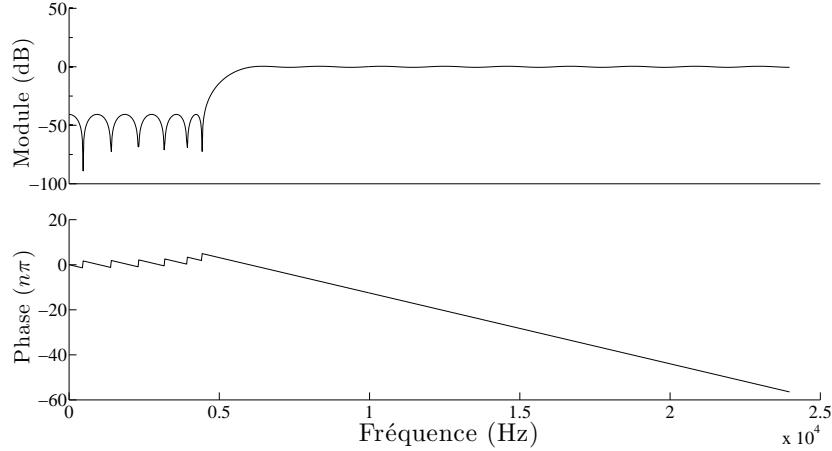


FIG. 2.11 – Spectre (module et phase) du filtre type matériaux passe-haut. Fréquence de coupure : 6000Hz. Spectre calculé sur  $N = 2048$  points.

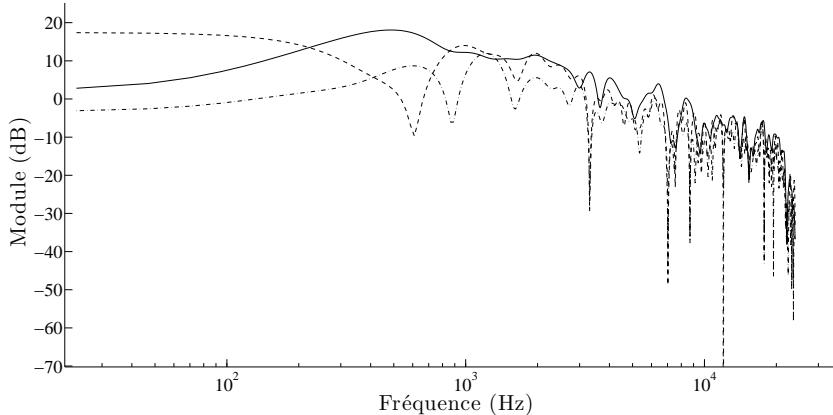


FIG. 2.12 – Spectres (modules) du filtre *Variations Spectrales Aléatoires (VSA)*. Trait plein : son direct (impulsion de revolver). Trait interrompu : première réflexion. Trait interrompu pointillé : deuxième réflexion. Spectres calculés sur  $N = 2048$  points. Ici, l'écart-type des variations est fixé à 0.1.

ici qu'un paramètre important de cette distribution est l'étalement de la Gaussienne, qui n'a pas été caractérisé dans ce travail mais qui devrait certainement retenir l'attention de futures recherches, et être comparé aux travaux de Bassuet [34] et Schodder [35], entre autres.

La distribution angulaire des rayons du modèle est calquée sur ces observations. Dans les travaux présentés, l'écart-type de la distribution normale centrée réduite est fixée arbitrairement (ce qui le rend très discutable).

**Remarque** Le modèle synthétise le signal de la réponse en énergie. Afin d'obtenir un signal en pression, le signe des retours est déterminé aléatoirement de manière équiprobable (sauf le son direct dont le signe est gardé positif) en calculant plus ou moins la racine carré du signal en énergie. Ce choix est basé sur la remarque de Kuttruff [36] qui rappelle que Becker et Mommertz ont montré que la phase des retours n'avaient subjectivement que peu d'importance.

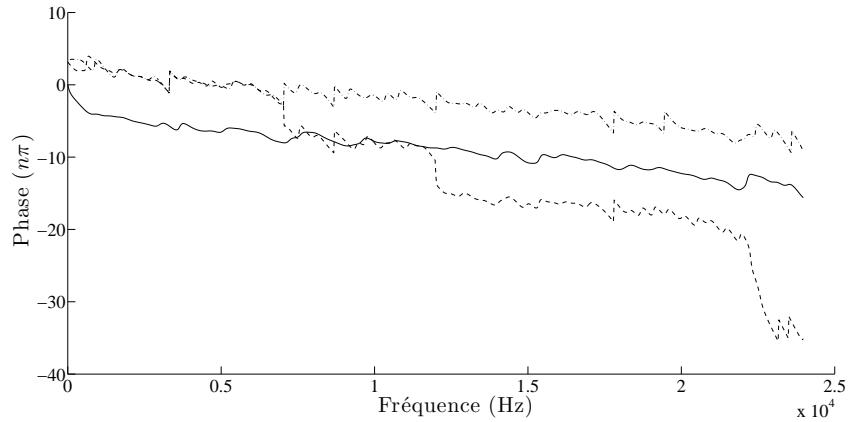


FIG. 2.13 – Spectres (phases) du filtre *Variations Spectrales Aléatoires (VSA)*. Trait plein : son direct (impulsion de revolver). Trait interrompu : première réflexion. Trait interrompu pointillé : deuxième réflexion. Spectres calculés sur  $N = 2048$  points. Ici, l'écart-type des variations est fixé à 0.1.

#### 2.5.4 Conclusion

Le modèle stochastique que nous venons de présenter a déjà été validé par Polack [7] du point de vue des indices acoustiques des salles (Section 2.6). Un des enjeux du travail que nous présentons dans ce document est la validation du modèle du point de vue physique. Autrement dit, nous allons par la suite étudier la validité de la statistique des retours générés par le modèle et la pertinence des ajouts que nous proposons. En effet, les filtrages que nous proposons ici sont arbitraires et basés sur l'observation, l'intuition et l'écoute. Nous verrons également l'intérêt de la synthèse de la diffusion. La validation physique du modèle se fera en comparant les réponses synthétisées à celles que nous aurons mesurées dans des salles de spectacles. Nous tentons de répondre à ces questions dans les Chapitres 5, 6. Une seconde approche est nécessaire : celle-ci est perceptive. Le modèle stochastique est-il pertinent perceptivement ? Quels sont les impacts de la diffusion ? du filtrage ? existe-t-il un filtrage plus vraisemblable que d'autres ? Ces questions devraient trouver des réponses dans des travaux futurs (des résultats préliminaires sont cependant présentés en Annexe E).

## 2.6 Qualité acoustique d'une salle

### 2.6.1 Rappel des phénomènes physiques mis en jeu

Lors d'une représentation au théâtre ou à l'opéra par exemple, il n'est pas rare d'observer des spectateurs cherchant les emplacements qui leur assurerait la meilleure écoute, c'est-à-dire la meilleure restitution sonore de la pièce ou de l'œuvre jouée. Le spectateur se base sur son expérience pour définir le choix de son emplacement. Il sait que le message qui lui parvient aux oreilles est une version modifiée de celui émis depuis la scène. Les modifications du signal sonore sont dues à la salle, aux propriétés des matériaux qui la composent et au nombre de réflexions subies par les rayons émis depuis la scène, ainsi qu'à la diffraction des front d'ondes sur des ornements du mobilier de la salle ou encore sur les rebords des balcons. Ces phénomènes modifient le son ; l'auditeur juge si ces modifications lui plaisent ou non. Les acousticiens des salles, depuis longtemps, cherchent à quantifier la qualité perçue de l'acoustique d'une salle à l'aide d'indices physiques, basés sur les phénomènes physiques mis en jeu lors de la propagation

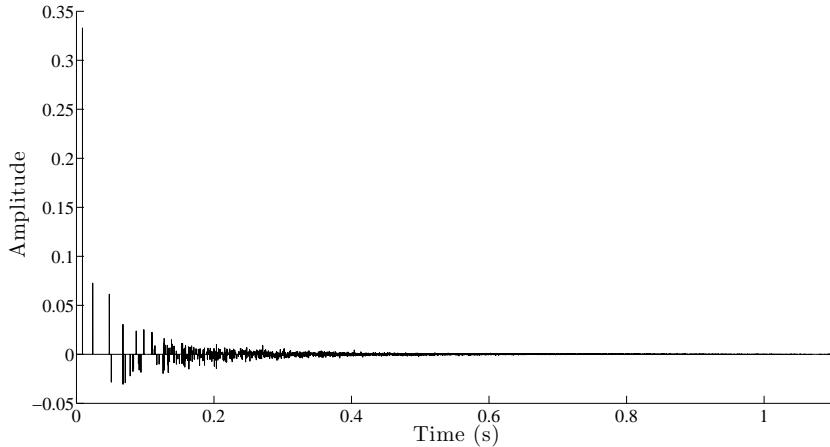


FIG. 2.14 – Signal temporel d'une réponse impulsionnelle synthétisée par le modèle (sans diffusion, sans filtrage).

du son dans la salle.

Avant de présenter quelques indices acoustiques (Section 2.6.2), rappelons brièvement les différents phénomènes physiques mis en jeu et la composition d'une réponse impulsionnelle de salle. Une représentation utile à la compréhension des phénomènes mis en jeu est celle des sources/images (Section 2.4 ; Fig. 2.1 et 2.4). Lorsqu'un son est émis, l'auditeur reçoit d'abord le son direct (Fig. 2.5), c'est-à-dire l'onde qui se propage en ligne droite du point d'émission au point de réception. Ensuite, arrivent les fronts ondes issues des différentes réflexions sur les parois de la salle. Comme cela a été vu précédemment (Section 2.2), différents phénomènes entrent en jeu : les réflexions spéculaires, les réflexions diffuses, la diffraction, etc. (Fig. 2.1). Le nombre de rayons sonores arrivant au point de réception augmente rapidement avec le temps, de telle sorte qu'au-delà d'un certain temps, le temps de mélange, il devient impossible pour l'auditeur de distinguer chaque retour. A partir de ce temps, on considère que le champ diffus s'est établi (Fig. 2.5).

Une approche physique de ces phénomènes consiste à traiter le signal capté au point de réception, c'est-à-dire à modéliser les modifications subies par le son émis par la source. La mesure de la réponse impulsionnelle de la salle peut être réalisée de plusieurs manières. Des impulsions (revolvers, ballons de baudruches, pétards) ou bien des séquences sonores (sinus glissants, séquence de Hadamard, bruit rose, etc.) [37] constituent des sources sonores couramment employées dans le domaine de l'acoustique des salles. La réponse de la salle au signal d'excitation peut être mesurée à l'aide de microphone omnidirectionnel, de tête de mannequin dans laquelle sont placés deux microphones au niveau des oreilles (mesure binaurale) [38], d'antenne de microphones, ou encore à l'aide de microphone captant le champ sonore en trois dimensions, telles les mesures Ambisonics [39]. La mesure de la réponse  $h(t)$  de la salle obtenue, l'échogramme se calcule à l'aide de :

$$E(t) = 20 \times \log_{10}(|h(t)|) - 20 \times \log_{10}(h_{ref}). \quad (2.41)$$

On a alors la réponse en énergie de la salle.

Les architectes souhaitant réaliser de *bonnes* salles, les acousticiens des salles s'intéressent à relier les phénomènes physiques mis en jeu lors de l'émission d'un son dans une salle à la perception de ce son par les auditeurs. Pour ce faire, des questionnaires sont distribués à un

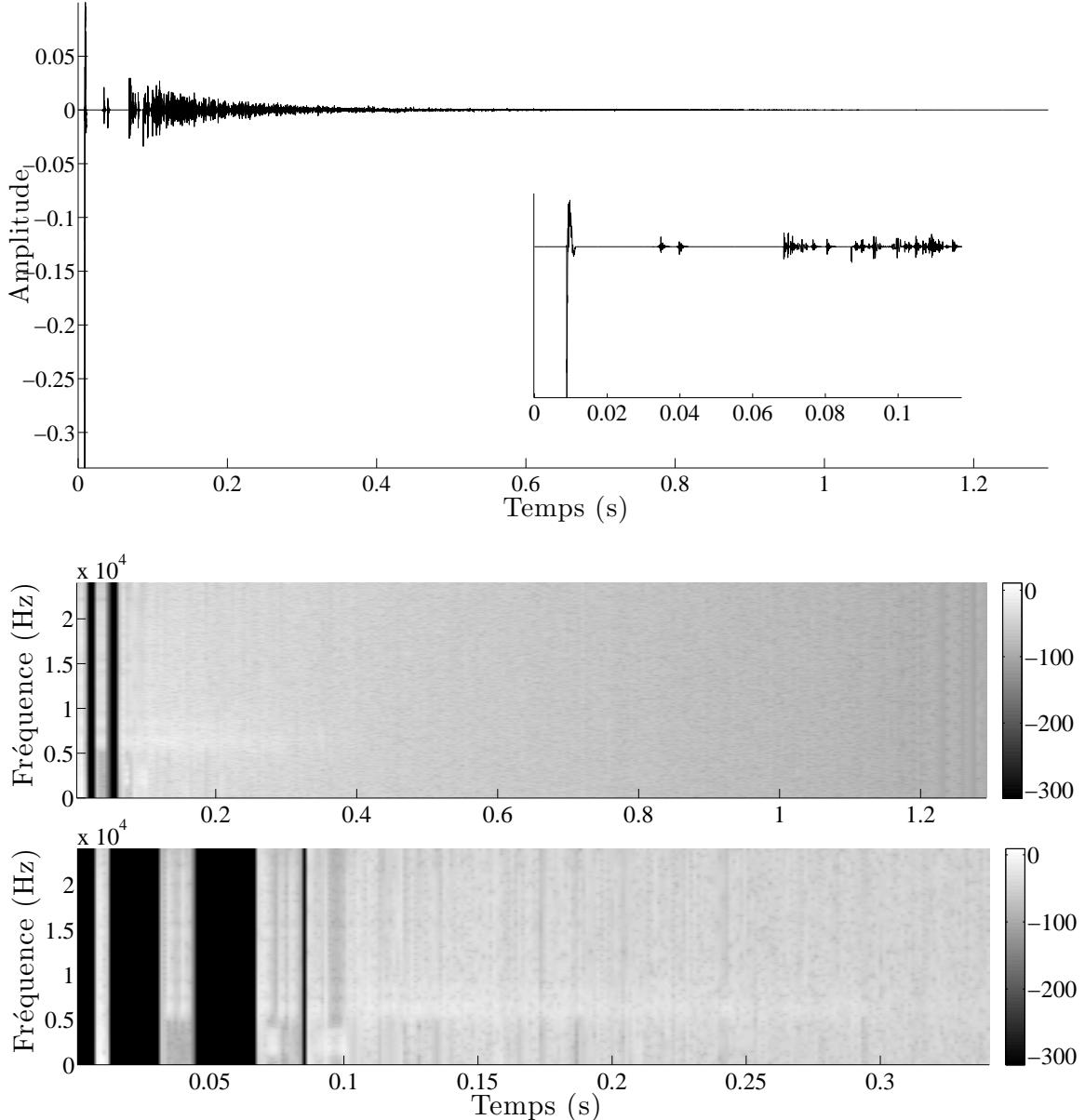


FIG. 2.15 – Haut : signal temporel d'une réponse impulsionnelle synthétisée par le modèle (avec diffusion et filtrage type matériaux). Milieu et bas : spectrogramme de la réponse impulsionnelle synthétisée ( $N_{FFT} = 2048$  pts). Haut : signal temporel entier ( $M=512$  pts temporel). Bas : zoom sur les premiers instants de la figure du haut ( $M=128$  pts temporel). L'échelle des couleurs est en décibels.

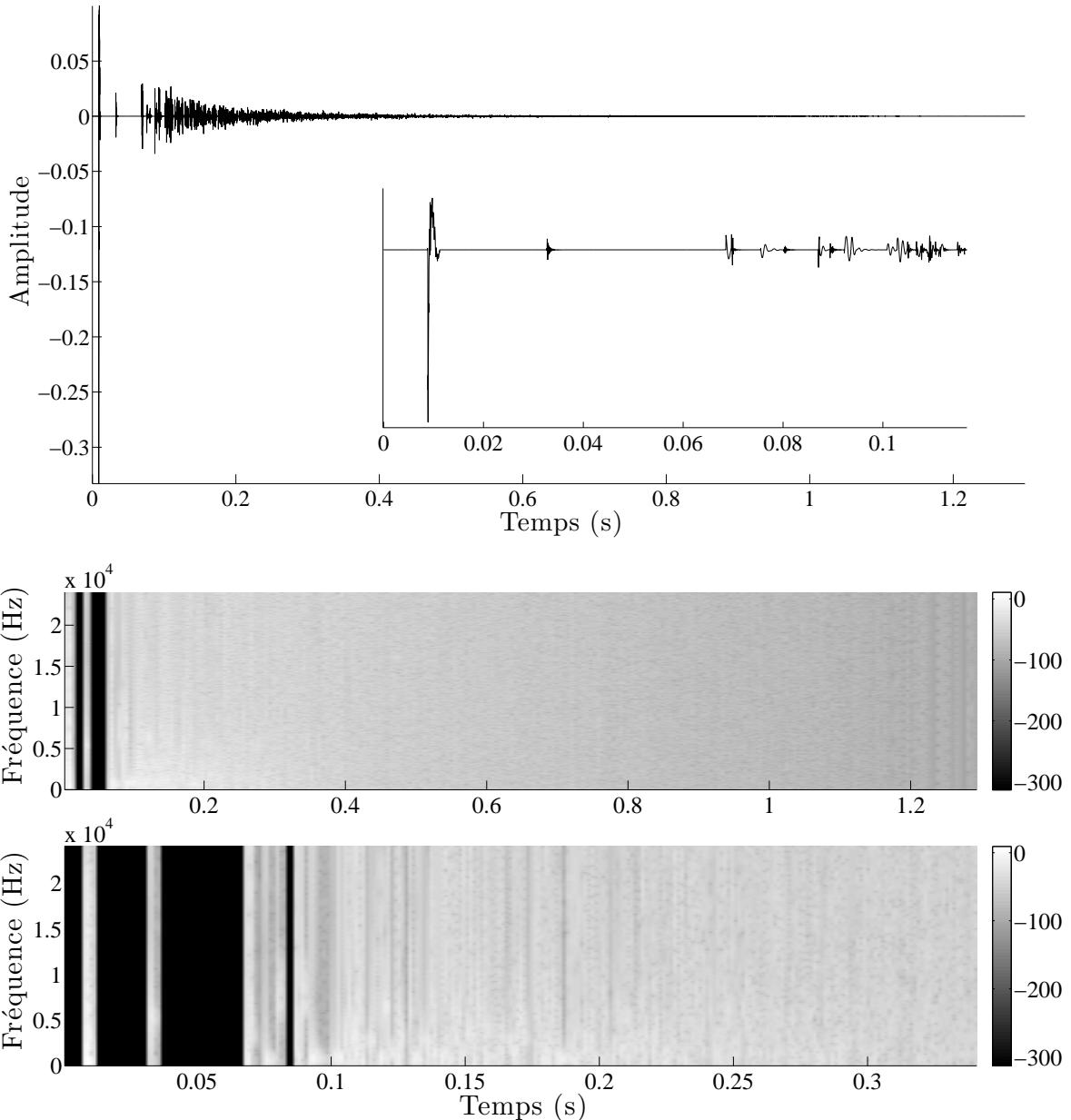


FIG. 2.16 – Haut : signal temporel d'une réponse impulsionnelle synthétisée par le modèle (avec diffusion et filtrage en bandes d'octave). Milieu et bas : spectrogramme de la réponse impulsionnelle synthétisée ( $N_{FFT} = 2048$  pts). Milieu : signal temporel entier ( $M=512$  pts temporel). Bas : zoom sur les premiers instants de la figure du haut ( $M=128$  pts temporel). L'échelle des couleurs est en décibels.

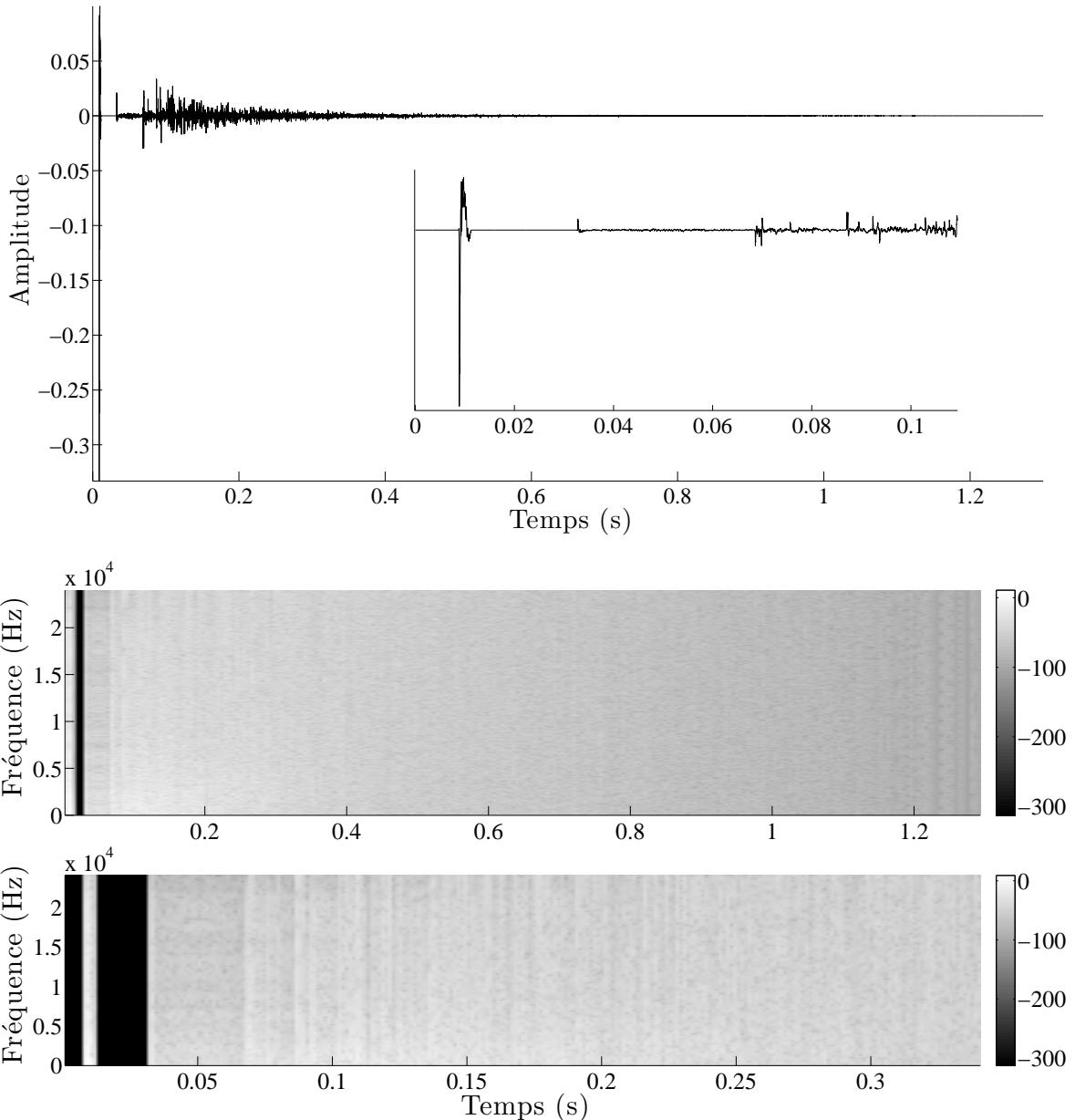


FIG. 2.17 – Haut : signal temporel d'une réponse impulsionnelle synthétisée par le modèle (avec diffusion et filtrage type VSA). Milieu et bas : spectrogramme de la réponse impulsionnelle synthétisée ( $N_{FFT} = 2048$  pts). Milieu : signal temporel entier ( $M=512$  pts temporel). Bas : zoom sur les premiers instants de la figure du haut ( $M=128$  pts temporel). L'échelle des couleurs est en décibels.

grand nombre d'auditeurs, et ce dans différentes salles de concert où les pièces musicales sont différentes. Les données obtenues sont confrontées aux mesures réalisées dans ces mêmes salles. De là, sont apparus une palette d'indices acoustiques, dont le fameux temps de réverbération. Pelorson *et al.* [40] conclut, après analyse multi-dimensionnelle de ces indices, que tous ces indices sont corrélés et qu'il est possible de n'en garder que cinq. La section suivante (2.6.2) présente ces quelques indices, sans pour autant procéder à une analyse fine de chacun d'entre eux. Il est à noter également que les parties de ce manuscrit utilisant des indices acoustiques ne se réfèrent qu'à certains d'entre eux pour des raisons expliquées dans les parties concernées (Chapitres 5 et 6).

## 2.6.2 Quelques indices acoustiques

Cette section présente certains indices acoustiques, donne leur définition et leur interprétation. Notons que quiconque voudra les calculer et les utiliser devra se référer à la norme ISO 3382 [41]. De plus, l'ouvrage de Beranek [42] fournit une description détaillée et approfondie de chacun de ces indices ainsi que de nombreux autres.

### 2.6.2.1 Durée de réverbération

Le temps de réverbération est une mesure du temps qui sépare l'instant de l'arrêt de la source de celui où celui-ci n'est plus audible. Par définition, le temps de réverbération en un point de la salle est le temps écoulé entre le moment où le niveau sonore de la source commence à décroître et le moment où le niveau sonore est inférieur ou égal au niveau initial moins 60dB [43]. Dans la pratique, la dynamique de 60dB n'est jamais atteinte. De plus, la décroissance de l'énergie n'est jamais tout à fait linéaire et comporte de nombreuses irrégularités qui diffèrent d'un point à un autre de la salle [44].

Le temps de réverbération est obtenu à partir de la pente des courbes de décroissances par bande d'octave obtenues par intégration rétrograde du carré de la réponse impulsionnelle [43]. La pente de la courbe de décroissance est déterminée par la droite de régression linéaire correspondant à la portion appropriée de la courbe de décroissance. Ainsi, pour l'estimation du temps de réverbération à 30dB, on calcule la droite de régression linéaire entre les temps à -5dB et à -35dB.

### 2.6.2.2 Durée de réverbération initiale : Early Decay Time (EDT)

L'EDT [45, 46] correspond au temps que met l'énergie sonore pour décroître de 60dB, en considérant que la décroissance s'effectue à la même vitesse pendant les 10 premiers décibels (de -5dB et -15dB). Cet indice se rapporte à la balance temporelle énergétique du son parvenant à l'auditeur. L'EDT est subjectivement plus important que le temps de réverbération  $T_R$ , car il semble correspondre à l'impression de réverbération, tandis que  $T_R$  fait référence aux propriétés acoustiques de l'auditorium.

### 2.6.2.3 Précision

La précision est un indice se rapportant au rapport de l'énergie précoce sur l'énergie tardive. Il peut être calculé pour une limite temporelle  $xx = 50ms$  ou  $xx = 80ms$ , selon que l'on considère une salle de conférence ou de concert, et est défini par :

$$C_{xx} = 10 \times \log_{10} \left( \frac{\int_0^{xx} h(t)^2 \cdot dt}{\int_{xx}^{+\infty} h(t)^2 \cdot dt} \right), \quad (2.42)$$

où  $xx$  est une limite temporelle exprimée en  $ms$ , et  $h(t)$  est le signal en pression acoustique instantanée de la réponse impulsionale mesurée au point de réception.

Certains acousticiens emploient également la définition  $D_{xx}$ , rapport entre l'énergie précoce et l'énergie totale, en particulier pour les salles de conférence :

$$D_{xx} = \frac{\int_0^{xx} h(t)^2 dt}{\int_0^{+\infty} h(t)^2 dt}, \quad (2.43)$$

où  $xx$  est une limite temporelle exprimée en  $ms$ , et  $h(t)$  est le signal en pression acoustique instantanée de la réponse impulsionale mesurée au point de réception.

La définition et la précision sont deux indices acoustiques liés par la relation suivante :

$$C_{xx} = 10 \times \log_{10} \left( \frac{D_{xx}}{1 - D_{xx}} \right), \quad (2.44)$$

où  $C_{xx}$  et  $D_{xx}$  sont la précision et la définition à  $xx$   $ms$  respectivement. Il n'est donc pas nécessaire de mesurer les deux grandeurs.

La précision, la définition, ainsi que le temps central font référence à l'équilibre perçu entre précision et réverbérance, ainsi qu'à l'intelligibilité de la parole.

#### 2.6.2.4 Temps central

Le temps central  $T_c$  est une mesure différente de la précision ou de la définition (Section 2.6.2.3) de la répartition de l'énergie au cours du temps. En effet, cet indice évite d'avoir recours à une frontière arbitraire entre l'énergie précoce et l'énergie tardive de la réponse. Cet indice, qui est une mesure du centre de gravité de la réponse impulsionale quadratique, est défini comme suit :

$$T_c = \frac{\int_0^{+\infty} t \times h(t)^2 dt}{\int_0^{+\infty} h(t)^2 dt}, \quad (2.45)$$

où  $h(t)$  est le signal en pression acoustique instantanée de la réponse impulsionale mesurée au point de réception.

### 2.6.3 Mesure de réponses impulsionales de salles

L'acoustique d'une salle est documentée en mesurant entre deux points de l'espace<sup>5</sup> (un point d'émission et un point de réception) la réponse de la salle à un signal d'excitation. La réponse impulsionale (ou dualement la fonction de transfert dans l'espace de Fourier) renseigne non seulement sur le système physique qu'est la salle, mais aussi sur les qualités acoustiques du lieu. Nous pouvons ainsi calculer des indices acoustiques (Section 2.6.2), ou encore envisager la pertinence de l'emploi d'une salle pour d'autres prestations que celles pour lesquelles celle-ci a été conçue. Les mesures de réponses impulsionales sont également utiles pour la prédiction de l'acoustique de la salle. Outre la synthèse numérique (Chapitre 2.4), les mesures dans les maquettes de salles permettent de prédire plus ou moins précisément l'acoustique du lieu.

---

<sup>5</sup>Ces deux points peuvent être interchangés selon le principe de réciprocité.

### 2.6.3.1 Sources sonores

A pleine échelle, principalement deux types de sources sonores peuvent être employées pour exciter la salle en un point. Nous distinguons les sources impulsives et les systèmes de haut-parleurs. Le premier type de sources comprend généralement des ballons de baudruche qui sont éclatés, des revolvers, des pétards ou des plaquettes de bois entrechoquées. Le second type de source peut simplement être un seul haut-parleur ou bien un système multi-haut-parleurs. A moins de privilégier une directivité particulière, les sources employées lors de mesures doivent autant que possible présenter un rayonnement omnidirectionnel et être plat en fréquence (généralement entre 100 et 5000Hz)<sup>6</sup> [15]. Certaines études peuvent se concentrer sur des directivités particulières de la source afin d'étudier la propagation de l'énergie dans une ou plusieurs directions. Loin d'être inutile, ce choix peut s'avérer judicieux lorsque nous souhaitons émettre l'énergie sonore comme le ferait approximativement un instrument de musique ou un locuteur sur scène. En effet, si les normes [41] conseillent d'employer des sources omnidirectionnelles, nous pouvons nous interroger sur la pertinence d'une telle indication : un orchestre, quelques instruments ou encore des acteurs n'émettent pas le son dans toutes les directions de l'espace, mais bien selon des directivités privilégiées.

Les sources impulsives (ballons, pétard, etc.) émettent des impulsions, dont la reproductibilité temporelle/spectrale et spatiale reste à étudier en détails, même si des premiers travaux ont abordé ce sujet [47, 48, 49, 50]. Les haut-parleurs servent soit à émettre un sinus glissant (les fréquences sont parcourues à une vitesse logarithmique) soit une séquence de nombre aléatoirement choisis (Maximum Length Sequence, séquence de Hadamard, etc.). La réponse impulsionnelle de la salle est alors obtenue en déconvoluant la réponse de la salle par le signal d'excitation.

### 2.6.3.2 Récepteurs

Les récepteurs employés sont des microphones omnidirectionnels, ou privilégiant des directivités particulières (B-format, cardioïdes, figure-de-huit, etc.). Les têtes de mannequin équipées de petits microphones placés dans les oreilles servent aux mesures binaurales. Dans la pratique, il est courant d'échantillonner l'espace, autrement dit, de mesurer non pas une unique réponse impulsionnelle, mais en plusieurs points préalablement choisis. Ainsi, si la salle est symétrique, il est inutile de mesurer des réponses des deux côtés de l'axe de symétrie. En général, il est communément admis de se référer aux indications de Gade [51] dont nous nous sommes servi dans [52].

### 2.6.3.3 Mesures dans les maquettes

L'étude de l'acoustique d'une salle avant ou après sa réalisation, à partir de sa maquette, date du début des années 1930 [12].

Les modèles de tracés lumineux sont relativement simples à mettre en œuvre. Ils permettent l'étude des chemins des rayons sonores (la source sonore est remplacée par une source lumineuse). Dans les modèles lumineux, les surfaces qui nous intéressent sont recouvertes d'un miroir ou d'une matière réfléchissante, afin de se focaliser sur les réflexions spéculaires et non diffuses. Les lasers sont les sources lumineuses les plus couramment employées, car elles émettent des rayons lumineux fins et concentrés en un point. Notons cependant que ce type de modèle ne permet l'étude des réflexions que jusqu'au premier voire au second ordre. Au-delà, les chemins deviennent trop complexes pour être visualisés [12]. Le principale inconvénient du modèle lumineux est qu'il

---

<sup>6</sup>L'émission de signaux de fréquence pure est rarement utilisée, car elle favorise un seul mode de la salle.

ne permet pas d'analyser les temps d'arrivées des rayons en un point de l'espace. Nous mesurons l'importance de cet inconvénient quand on sait l'importance perceptive que revêtent les différences de temps d'arrivée des rayons lumineux. L'intensité lumineuse en un point peut être reliée à un temps de parcours (l'intensité sera élevée si ce temps de parcours est faible, par exemple) ; mais cela peut tout aussi bien être le fruit de la superposition de plusieurs rayons. Par conséquent, afin d'obtenir des informations pertinentes sur les temps d'arrivée des rayons, le modèle lumineux doit être agrémenté de la mesure des temps de parcours entre chaque réflexion. Finalement, les longueurs d'onde de la lumière sont tellement courtes qu'elles ne modélisent la propagation de l'énergie sonore uniquement dans le domaine des très hautes fréquences. Les phénomènes de diffraction et de réflexions diffuses qui ont lieu aux basses fréquences ne sont pas prises en compte.

Les maquettes remplies d'eau présentent l'avantage de prendre en compte la diffraction et les réflexions diffuses<sup>7</sup>. De plus, la vitesse de propagation d'une onde est de l'ordre de  $20 - 30\text{cm}^{-1}$  (soit 1/1000 de la célérité du son). Par conséquent, la propagation de l'onde dans l'eau peut être facilement suivie à l'oeil nu et filmée à l'aide d'une caméra rapide, par exemple. L'excitation du milieu peut se faire soit en déplaçant rapidement un objet de petite taille, soit en laissant choir une goutte d'eau sur la surface<sup>8</sup>. L'inconvénient de ce modèle est qu'une simple excitation donne naissance à plusieurs front d'ondes réfléchis de longueurs d'onde différentes. Les différences de longueurs d'onde, l'absence de dispersion, ainsi que la tension de surface de l'eau et les effets de la gravité rendent ce genre de modèle peu attractif.

L'emploi de la technique de "Schlieren" [53] permet d'éviter les inconvénients des précédentes méthodes : la dispersion peut être visualisée et la vitesse de propagation est la même à toutes les fréquences. De plus, la visualisation de la propagation de l'énergie est nettement plus claire que dans une maquette remplie d'eau [12]. La source utilisée est généralement une étincelle qui génère en champ proche un front d'onde cylindrique qui devient sphérique en champ lointain. Les fronts d'onde sont mis en évidence par des lignes plus ou moins sombres en raison de la variation de l'indice de réfraction créée par la perturbation du champ acoustique. Les fronts d'onde sont par conséquent aisément visualisés et peuvent être filmés à l'aide d'une caméra rapide. Notons que la technique de 'Schlieren' a été employée pour la première fois en acoustique des salles par Wallace Clement Sabine [2, 12]. L'inconvénient de cette approche est qu'elle ne permet la visualisation de la propagation du son uniquement dans un plan.

Outre les précédentes techniques, la modélisation par translation de la fréquence du champ sonore reste sans doute la plus employée. En effet, tous les effets rencontrés lors de la propagation de l'énergie sonore sont pris en compte et la mesure est réalisée en tenant compte des trois dimensions de l'espace, sans pour autant permettre de visualisation de la propagation de l'énergie au cours du temps. Le champ de pression est enregistrée en un point de l'espace, comme dans les mesures à pleine échelle. L'étincelle constitue sans doute le moyen technologique le plus robuste pour exciter la salle [12]. Le signal enregistré au point de réception doit être ramené dans le domaine des fréquences audibles grâce au facteur d'échelle [12]. Spandock, et plus tard Jordan [12], sont les premiers à prouver que les mesures dans les maquettes sont possibles avec des sources et des récepteurs électroacoustiques [12, 54]. Notons que la difficulté rencontrée avec les haut-parleurs miniatures est le contrôle de leur rayonnement.

---

<sup>7</sup>Les premières expériences datent de 1843, d'après Cremer [12].

<sup>8</sup>Les inconvénients de telles méthodes ne sont pas discutées par Cremer [12]. En effet, nous pouvons nous interroger sur la reproductibilité de ces sources et des effets indésirables (gouttes secondaires, maladresse de l'expérimentateur, etc.).

Le logiciel MIDAS [20] créé dans les années 1980 permet la mesure et l'analyse de réponses impulsionnelles dans des salles à pleine échelle ou dans des modèles réduits. Présenté dans le Chapitre 3, MIDAS compense l'atténuation de l'air aux hautes fréquences et le bruit de fond et permet ainsi, outre une analyse physique de la mesure, une évaluation perceptive de l'acoustique d'une salle avant que celle-ci ne soit construite.

## Chapitre 3

# Mesure et analyse de réponses impulsionales de salles

### 3.1 Introduction

L'acoustique des salles a pour but de décrire la qualité acoustique d'un lieu. Dans la pratique, un grand nombre de mesures de réponses impulsionales entre des couples de points d'émission (la source) et de réception (le récepteur). Les Chapitres précédents présentent l'architecture de la réponse impulsionnelle, ainsi que les critères objectifs (indices acoustiques) employés pour qualifier la qualité de la salle.

Dans les années 1980, Polack *et al.* développent le logiciel MIDAS qui permet la mesure et l'analyse de réponses impulsionales dans des salles à pleine échelle ou dans des maquettes. Resté longtemps une référence, MIDAS n'est plus guère employé aujourd'hui. En effet, il a été créé pour fonctionner sur des ordinateurs Macintosh Apple des années 1980. Depuis, l'architecture des ordinateurs a évolué, rendant presque difficile l'utilisation de MIDAS.

Une des réalisations de cette thèse est la rénovation du logiciel. Initialement écrit en Fortran 77, il a été porté en langage MatLab et son organisation a été repensée. En effet, MIDAS a gagné sa réputation grâce à ses routines et ses techniques de mesures étalonnées, ce que peu de logiciels -même encore aujourd'hui- offrent. La rénovation de MIDAS a été menée dans le but de créer un outil léger, gratuit et dont le code est accessible à tous. La Section 3.2 présente les modifications apportées, et la nouvelle organisation du logiciel qui se nomme désormais OpenMIDAS.

Le reste de ce Chapitre se concentre sur deux aspects principaux de la mesure et de l'analyse des réponses impulsionales de salles. L'intérêt porté aux deux sujets que nous traitons par la suite vient du travail sur MIDAS, de la compréhension de ses fonctions. Autrement dit, le travail sur MIDAS nous a permis d'aborder certains points de la mesure et de l'analyse des données qui ne sont pas triviales :

1. la détection du temps d'arrivée du son direct, autrement appelé *onset* (Section 3.3)

Nous montrons comment de faibles erreurs de détection de l'onset peuvent engendrer de grandes erreurs sur les indices acoustiques calculés à partir de ce temps. Nous proposons et évaluons plusieurs méthodes d'estimation dont l'erreur de détection est faible ;

2. la détection des limites temporelles (onset et offset) du son direct (Section 3.4)

Dans de nombreux cas, les réponses impulsionales ont été et sont encore mesurées avec des sources sonores dont les paramètres physiques (reproductibilité, diagramme de rayonnement et niveau) ne sont pas connus : pistolet, ballons de baudruche, pétards, etc. La connaissance *a posteriori* de la source peut s'avérer utile, en particulier si l'on souhaite blanchir la réponse, ou estimer

les temps d'arrivée des retours [33] (Chapitre 5). Nous confrontons ici deux méthodes d'estimation des limites temporelles du son direct. L'étude est menée sur des réponses expérimentales mesurées avec les trois sources sonores citées précédemment. Nos méthodes permettent également de montrer les différences de reproductibilité entre les trois sources considérées.

## 3.2 From MIDAS to OpenMIDAS

### 3.2.1 Introduction

The MIDAS package [20] was created in the late 1980's, and has been written in Fortran77. It assures on the one hand, the acquisition of impulse responses (IRs) in all-scales and in model-scales, and on the other hand, the processing of the IR, in order to obtain room acoustical indices such as Reverberation Times, Early Decay Time, Clarity, Central Time, Speech Transmission Index, Bass Ratio, etc. More details about acoustic criteria can be found in [41, 42]. The OpenMIDAS (written in MatLab) is a refurbishment of the MIDAS system and is constituted of two main modules<sup>1</sup> : Acquisition (A) -which is currently under development- and Processing (P). The OpenMIDAS software provides an easy acces and understanding of the code, and aims at creating a user community who could share and replace original functions by their own, as long as they respect data labels. In this work, we have focused on module P<sup>2</sup>.

### 3.2.2 Description of module P

Module P has conserved its original organization. It first starts from the lecture of a parameter ('*MIDAS-Parameters.txt*') file derived from module A (temperature, relative humidity, number of channels, sampling rate, number of samples, time-step, etc.). Module P provides either a time/frequency visualization of the room impulse response decay, or (and) calculates room acoustical indices. Once the acquisition file is read, module P computes constants (number of octave bands, number of points for the FFT, etc.), calculates the filter to apply to the response and then reads (from a text file created by the user) the room acoustical criteria chosen by the user. If the sound source is known, one is able to calibrate the signal. Further, based on MIDAS, OpenMIDAS keeps the possibility to carry out measurements in scale-models. In that case, the air absorption is compensated. It is also possible to compensate for the background noise, based on a threshold set by the user. Finally, the Lateral Energy Fraction, Transmission indices, the signal decay and usual room acoustical indices are computed. All data generated during the process are saved in a MatLab file ("*.mat*") and in a text file.

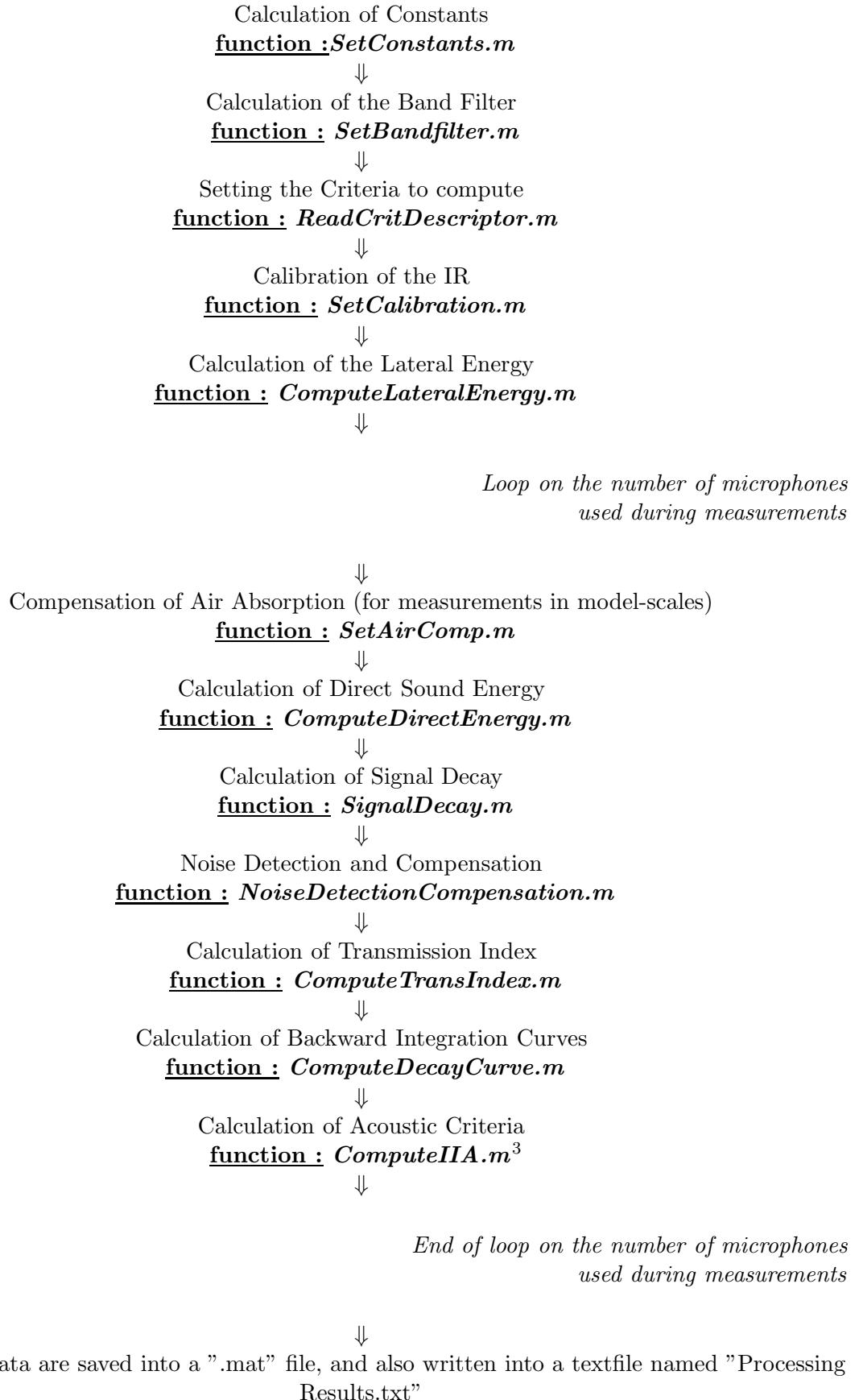
We have chosen to organize module P as a combination of functions. Each routine receives data from a previous one in a sub-structure, and outputs processed data into another sub-structure. Once the computation is finished, the user has the opportunity to access -through the MatLab file- each step of the computation, that is, all data (sub-structure) calculated by each function. In the following, we describe the global organization and give a brief description of each function. A detailed description of module P is given in Appendix B.

Read parameters from Acquisiton file  
**function : IndexComputation.m**  
 ↓

---

<sup>1</sup>All functions and parameters names mentionned in this Section are those found in the original MIDAS code.

<sup>2</sup>J-D. Polack has now adapted the original MIDAS user interface to MatLab, so that module P can currently be ran with the original friendly user interface.




---

<sup>3</sup>IIA : Integrated Impulse Answer

### 3.3 Détection de l'onset de réponses impulsionales

#### 3.3.1 Préambule

De nombreux indices acoustiques (Chapitre 2.6) sont calculés à partir de réponses impulsionales de salles sans que le temps de propagation entre la source et le récepteur soit pris en compte. Durant ce temps, le capteur placé au point de réception enregistre du bruit de fond, du bruit de mesure, et parfois quelques bruits parasites (claquement de portes, etc.). Les normes internationales [41] de mesures en acoustique des salles ne précisent pas la ou les méthodes requises pour estimer le temps d'arrivée du son direct au point de réception, c'est-à-dire, le premier point *valide* de la réponse. Nous appelons cet instant *l'onset* de la réponse impulsionale. Une estimation précise et d'erreur faible et connue de l'onset est nécessaire car, comme nous le montrons, une détection erronée peut avoir d'importantes répercussions sur les indices que l'on calcule. Aussi, plusieurs méthodes d'estimation de l'onset sont présentées et évaluées dans la suite comparativement à une méthode de référence que l'on estime être la plus précise et robuste. L'évaluation des erreurs de détection des méthodes d'estimation de l'onset est un point difficile que nous discutons par la suite. Cette étude conclut quant à la nécessité d'introduire dans les normes une méthodologie d'estimation de l'onset dont l'incertitude est connue.

Cette Section a été publiée en tant qu'*Express Letter* dans la revue *Journal of the Acoustical Society of America* sous le titre : " *Finding the onset of a room impulse response : straightforward ?*" (G.Defrance, L.Daudet and J-D. Polack) [55].

---

#### 3.3.2 Abstract

This letter deals with precision issues in the determination of the timing of the Room Impulse Responses (RIRs) onset. First, it is shown that while errors of onset timing estimation do not have that much effect on temporal indices, an erroneous onset estimation leads to significant differences in energetic and statistic acoustical indices. Twelve automatic onset detection methods are compared, in terms of precision, robustness and complexity. Experimental validation made on a set of 100 RIRs provides good evidence in favor of spectral and/or energetical methods, according to the type of sound source.

#### 3.3.3 Introduction

A standard way to document the acoustics of a room is to measure a set of Room Impulse Responses (RIRs). A RIR should ideally be recorded within absolute silence, a condition which of course is never met in practice. It then belongs to the acoustician to identify the edges of the RIR. Acousticians have for a long time designed various methods, as attested in the ISO 3382 [41] standard, for determining the last point of the reverberation tail, which is mixed with background noise. However, nothing is said about the onset of the RIR, implicitly assuming that this is a straightforward task : each acoustician, or software, can have its own method. In

order to find the onset time, the simplest ways that come first to mind would be 1) to determine it manually, i.e. visually on the waveform; or 2) to consider that the RIR always starts by direct sound, which often has the largest amplitude of the signal. Measurements in a concert hall typically leads to a hundred of RIRs, so that the first method becomes easily cumbersome. The second method, as we shall see, can in some cases provide poor estimates of the onset time. For instance, when strong scattering attenuates the direct sound, it is the first reflection that presents the largest amplitude, and the RIR maximum and its beginning can differ from less than  $0.1ms$  up to more than  $50ms$  (Figure 3.3 -(b)). Therefore, direct comparison between different authors and/or methodologies may become unreliable.

The first goal of this letter is to show that, for some of the most commonly used room acoustical indices, the above methods lead to large errors. The ISO 3382 standard lists some of these indices. Energetic and statistic, such as Clarity ( $C_{80}$  in dB, at  $80ms$ ) and Central Time ( $T_C$  in ms), are derived from a ratio of integration of the energy of the RIR; while temporal indices, such as Early Decay Time ( $EDT_{10}$ ) and the Reverberation Times ( $RT_{20}$  and  $RT_{30}$ ) are obtained from a linear regression made on the integration of the energy of the RIR. We do not attempt to discuss the relevance of such indices, be they orthogonal or not. This has already been done for instance by Pelorson *et al.*[56]. We do attempt, however, to propose and evaluate a number of different automatic onset detection methods for their precision, robustness, and complexity. This is the second goal of this letter.

The letter is constructed as follows. In section 3.3.4, we investigate how the precision for the localization of RIR onsets influences the result for 5 commonly used acoustic indices. Section 3.3.5 describes 12 methods for the automatic determination of RIR onsets. The performance of these methods are empirically compared in sections 3.3.6. The last section (section 3.3.7) discusses guidelines for choosing the right method according to the application at hand, and concludes on the need to present reliable results with documented uncertainty.

### 3.3.4 Precision of onset timing vs precision of the indices

Assuming that  $t_0$  is the reference time index corresponding to the maximum of the RIR, we compute on one full-band typical RIR (from Salle Pleyel, Paris [52]) the  $EDT_{10}$ ,  $RTs$ ,  $C_{80}$  and  $T_C$  indices using different onset times  $t' = t_0 - \Delta t_0$ , with  $\Delta t_0$  varying from  $0.05ms$  to  $10ms$ . Table 3.1 shows how acoustic indices can be affected differently by differences on onset timing. Even with  $\Delta t_0 = 10ms$ , errors on  $EDT_{10}$  and  $RTs$  still remain weak (around 0.6%). The authors would like to point out that calculating the  $EDT$  according to Jordan's definition [46] leads automatically to large errors, since it looks for the time that the total energy has decayed by  $10dB$ .

On the contrary,  $\Delta t_0 > 2ms$  leads to a difference in clarity of  $1.45dB$  (i.e., 55%). Variations of  $T_C$  remain inferior to 20% for the RIR tested here, with similar values for other RIRs. This way, even if the method used to determine the onset index is the maximum of the RIR, the user should prefer to calculate the central time than the clarity, because they both refer to the density repartition of the energy in the signal. This simple example shows the importance of exact onset detection of impulse responses in room acoustics. As errors on temporal indices are low, and would be considered as insignificant by most experts, as documented in [21], they will not be discussed in this letter. Nevertheless, attention is paid to energetic and statistic indices, since they have a strong dependency on the precision of the onset timing.

### 3.3.5 Proposed onset detection methods

This letter presents twelve different methods for onset detection. As the study of the acoustics of a hall can lead to typically hundreds of measurements (hence 100 RIRs), this article is focused

TAB. 3.1 – Variations of acoustical indices as a function of errors in the onset time  $\Delta t_0$  :  $\Delta EDT_{10}(\%)$ ,  $\Delta RT_{20}(\%)$ ,  $\Delta RT_{30}(\%)$ ,  $\Delta C_{80}(dB)$ ,  $\Delta T_C(\%)$ .

$\Delta t_0(ms)$	$\Delta EDT_{10}(\%)$	$\Delta RT_{20}(\%)$	$\Delta RT_{30}(\%)$	$\Delta C_{80}(dB)$	$\Delta T_C(\%)$
0.05	0.002	0.002	0.002	0.25	3.6
0.1	0.005	0.006	0.006	0.31	4.5
0.2	0.01	0.01	0.01	0.37	5.3
1.0	0.05	0.06	0.06	1.29	17.6
2.0	0.11	0.11	0.12	1.42	18.7
4.0	0.23	0.23	0.24	1.35	16.5
10.0	0.57	0.59	0.6	0.92	9.9

on methods that can be computed automatically. The methods presented below relies on the idea that the onset is linked to abrupt changes in one or more properties of the audio signal [57], and can therefore be detected by detecting the changes.

### 3.3.5.1 Temporally based methods

When observing the temporal structure of an RIR, it is noticeable that the occurrence of an onset is accompanied by a sudden increase of amplitude. The first methods of onset detection were based on this property by using a detection function which follows the envelope of the signal [57].

Four simple functions are presented here : (1) Maximum ( $M$ ) : As seen in introduction, the first idea is to consider the onset  $t_0$  to be the maximum of the absolute value of the RIR. (2) Maximum minus 5 milliseconds ( $M_5$ ) : A few softwares, such as the MIDAS package [20], consider that the onset can be defined 5ms before the maximum of the RIR. (3) Mean over time ( $D_E$ ) : An envelope follower can easily be constructed by low-pass filtering the local energy [57]. The maximum of  $E(n)$  is detected, and the signal analyzed from its beginning to its maximum. The ratio of two successive windows is calculated ; the index is found when the ratio is maximum (Eq.3.1). This can be written as :

$$E(n) = \sum_{m=-N/2}^{N/2} w(m)x^2(n.h + m) \quad (3.1)$$

$$t_0 = h \times \text{argmax}_n(E(n+1)/E(n)) \quad (3.2)$$

where  $w$  is a smooth windowing function, and  $h$  the time step between two windows.

(4) Threshold ( $E$ ) : This method works on the energy of the RIR, which is windowed by rectangular windows  $w(m)$ .  $t_0$  is here defined as the first time index where this local energy is  $K$  (typically  $K = 3$ ) times higher than its median running on past windows.

### 3.3.5.2 Spectrally based method

Since the spectra of the direct sound and the firsts reflections of an impulse response are very different from the background noise, a function based on spectra comparison is expected to give to good results, as seen in [58]. One can expect an increase of low frequencies components when looking at the direct sound.

(5) Mean over spectra ( $D_S$ ) : The idea is almost identical to  $D_E$  method, but ratio are calculated over spectra (Eq.3.3).

$$\tilde{E}(n) = \sum_k |X(n.h, k)| \quad (3.3)$$

$$t_0 = h \times \text{argmax}_n(\tilde{E}(n+1)/\tilde{E}(n)) \quad (3.4)$$

where  $X(n, k)$  is the Short Time Fourier Transform of the signal  $x(n)$ , and  $h$  the time step between two windows.

### 3.3.5.3 Time frequency method

(6) Wavelet Transform ( $W$ ) : This method, often used for denoising [59], is a natural tool for analyzing transient signals, since its time frequency resolution provides an increasingly finer time resolution at smaller scales. Let  $\Psi$  be a zero-mean real gaussian wavelet. The wavelet transform of  $x(t)$  is defined as :

$$Wx(u, s) = \int_{-\infty}^{+\infty} x(t) \frac{1}{\sqrt{s}} \Psi^* \left( \frac{t-u}{s} \right) dt \quad (3.5)$$

where  $u$  is the translation parameter and  $s$  the scale factor.

In other words,  $Wx(u, s)$  measures variations of  $x(t)$  near  $u$  within an equivalent window of size  $s$  (Eq. 3.5). When  $s \rightarrow 0$ , the decrease of the wavelet coefficients characterizes the regularity of  $x$  around  $u$ . The onset of the transient signal is estimated by applying a threshold to the wavelet coefficients [59], such as :

$$Tm = \sigma_m \sqrt{2 \log_e(N)} \quad (3.6)$$

where  $\sigma_m$  is the standard deviation of the noise, and  $N$  the length of the signal. After inverse wavelet transform, the onset is estimated as the first non-zero sample.

### 3.3.5.4 Refinement step

The six methods presented above can be refined by a method based on linear prediction ( $L_P$ ). Linear prediction analysis finds the coefficients of a FIR linear filter that predicts the current value of the real-valued time series based on past samples, minimizing the prediction error in the least squares sense.  $L_P$  residual is computed on adjacent windows  $[t-m : t]$  and  $[t : t+m]$ , with  $m = 0.8ms$ .  $t_0$  is the time index that maximizes the likelihood of having a stationary gaussian residual in both backward and forward windows, with a  $\chi^2$  goodness-of-fit test [60]. In other words,  $t_0$  corresponds to a changepoint in the behaviour of the system. This method is potentially very precise, but cannot be used as such, since it detects any spurious event regardless of its size. Instead, it is used only as refinement step locally around the times provided by the previous methods. We call the six improved methods :  $ML_P$ ,  $M5L_P$ ,  $D_E L_P$ ,  $D_S L_P$ ,  $EL_P$ , and  $WL_P$ .

### 3.3.5.5 Reference method

An important issue regarding the evaluation of these methods is that the exact onset timing is not known (no absolute truth). These reference points could be determined manually by experts, at the cost of a tedious hand-labeling and a potential lack of consistency between experts. Here, we decide to choose method  $D_S L_P$  as reference. The rationale for choosing this method is the following. First, on all the RIRs that we have at hand,  $D_E$  and  $D_S$  are the only methods which always return onset indices validated at hand by the authors (this is not always the case for  $M$ ,

$M_5$ ,  $E$ ,  $W$ ). As differences of onset estimation between methods  $D_E$  and  $D_S$  are inferior to  $0.1ms$ , we need to look at the enhanced methods. Indeed, method  $D_{ELP}$  estimates the onset at the beginning of a small ripple that we call *precursor* (Figure 3.3 -(c) and (d)), while method  $D_{SLP}$  detects the onset at the end of the precursor. It is assumed that the precursor does not belong to the RIR itself, but is either an artifact of the pistol shot (maybe due to the cylinder rotation), or more probably an artifact of an anti-aliasing filter in the A/D conversion. Second, we assume that the most precise method should return the best onset index.  $L_P$  being a refinement step, this leaves  $D_{SLP}$  as reference.

### 3.3.6 Comparison of the 12 automatic onset detection methods

#### 3.3.6.1 Evaluation method

The twelve detection methods are tested over 100 audio wavfiles. These RIRs have been measured in Salle Pleyel in Paris [52], according to the international standard [41], with pistol shots as sound sources. For each RIR, the onset time is estimated by each of the 12 methods. From this, acoustical indices ( $C_{80}$ ,  $T_C$ ) are calculated and compared to the reference value, returned by method  $D_{SLP}$ . Onset times that are not roughly consistent with the reference onset (i.e. not within  $100ms$ ) are not taken into account for the statistics.

#### 3.3.6.2 Robustness to noise

A way to assess the robustness of the presented estimators is to vary the background noise level of the RIRs. Figure 3.1 shows the mean variations of the onset time estimated by the first 6 methods on 100 experimental RIRs, with an  $SNR$  varying from 25 to  $90dB$ , by step of  $5dB$ . For each method, the reference onset time is the one obtained without adding noise to the signal. As expected from 3.3.5.3, the most robust method is  $W$ , followed by  $M$  and  $M_5$ ; the worst method is  $E$ , followed by  $D_S$ .

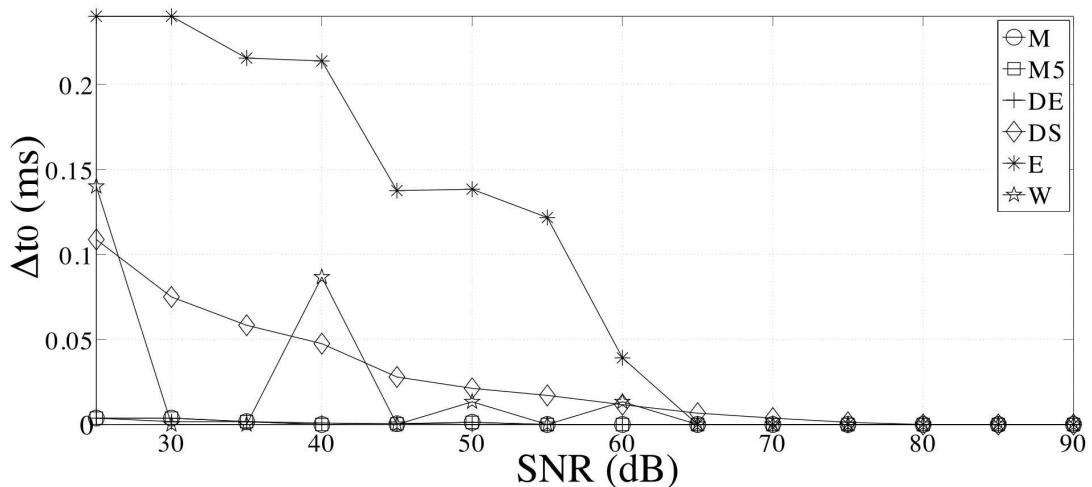


FIG. 3.1 – Variations of onset times estimation as a function of  $SNR$ .

#### 3.3.6.3 Results and Discussion

Except for large variations, which are discussed later, results (Figure 3.2) show that adding the method  $L_P$  to any other method improves considerably the accuracy of onset detection, and

hence, decreases errors, except for method  $M_5$ .

Methods  $M$  and  $M_5$  are extremely robust to artifacts, but present important variations, since they do not account for a potential scattering effect. Figure 3.3 presents 3 different RIRs. The first one (a) starts very near its maximum. One can expect that detection results should not vary from a method to another. The second RIR (b) presents a long scattering effect ( $\approx 30ms$ ), caused by a balcony. Methods  $M$  and  $M_5$  provide bad estimations in that case. In cases (c) and (d), the onset is not the maximum, but a *precursor*, as introduced in 3.3.5.5. These small variations cannot be detected by  $M$  and  $M_5$ , but also by method  $E$ , since its threshold is not always adapted to the RIR's precursor amplitude. Moreover, because of its threshold, that the user has to set differently for each RIR,  $E$  is not suited to an automated analysis.

Methods  $D_E$ ,  $D_S$  and  $W$  seem to be particularly indicated for a quick and precise onset index determination, method  $W$  being more robust to noise than the other methods, as seen in section 3.3.6.2. The resulting estimated  $C_{80}$  is always within  $\pm 0.01dB$  (i.e.  $\pm 3\%$ ) of our reference method given by  $D_S L_P$ . Indeed, even if method  $L_P$  improves on other methods, the computation time is significantly increased for a negligible gain in precision (typically below 0.1%).

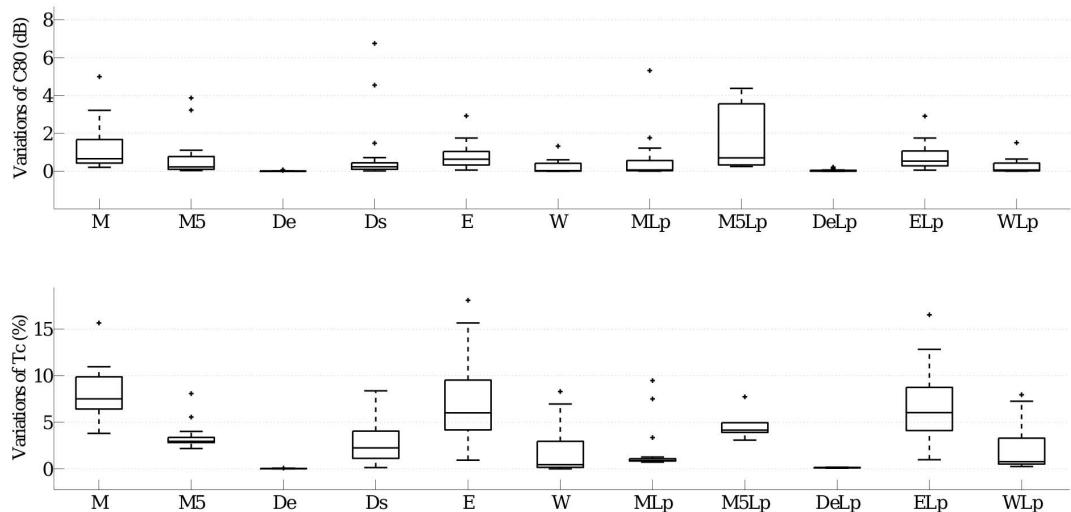


FIG. 3.2 – Variations of  $C_{80}$  (top) and  $T_C$  (bottom) for the 11 methods,  $D_S L_P$  being taken as reference.

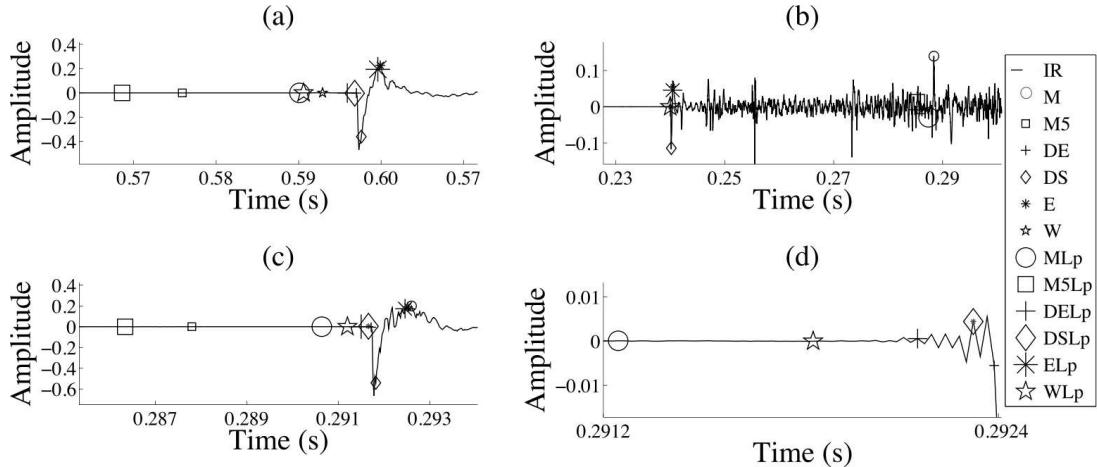


FIG. 3.3 – Example of three different RIRs (note the different time scales). a) Simple case onset ; b) Scattering effect ; c) Precursor ; d) Details of a precursor presented in (c).

### 3.3.6.4 Measurements with balloon bursts

The same experiment has been carried out on 100 RIRs measured with balloon bursts in salle Pleyel (for the same source and receiver positions). Method  $D_{SLP}$  being also taken as reference, results slightly differ from those obtained with the pistol shots. Although method  $L_P$  improves all methods, except  $M_5$ , method  $D_S$  offers the best estimate ( $\bar{\Delta}_{C_{80}} = 0.02dB$ ,  $\sigma_{C_{80}} = 0.02dB$ ;  $\bar{\Delta}_{T_C} = 0.3\%$ ,  $\sigma_{T_C} = 0.01\%$ ), instead of  $D_E$  for pistol shots. This can be explained as follows. The pistol shot has a much sharper increase of energy than balloon burst [50], typically 1.5 times faster. Thus, differences between pistol shots and balloon bursts are both spectral and temporal, and also related to their directivity, as explored in[50]-[61]. Further studies are required to refine these claims. It should be noticed that non-linear effects that may affect pistol shots are not detected by these methods since they intrinsically are parts of the response.

### 3.3.7 Conclusion

The main goal of this letter is to raise awareness on a loophole in the ISO 3382 standard for the computation of room acoustics indices. If one is interested in statistic and energetic indices such as  $T_C$  and  $C_{80}$ , a robust and precise method for determining the onset time is necessary. For instance, a variation of only 2ms can generate high variations of clarity (around 1.5dB, i.e. 55%). Furthermore, it is shown that for some of RIRs, there are large differences in the results given by obvious used onset detection methods, inducing significant differences in the acoustic indices.

Our experimental tests also show that methods based on energetical differences (method  $D_E$ ), for pistols shots, on spectral differences ( $D_S$ ) for balloon bursts, and on time-frequency analysis, such as wavelet transform ( $W$ ), seem to provide reliable estimates, with a precision that is appropriate for most uses. Nevertheless, the computation time needed by method  $W$ , and the slight difference with results obtained with  $D_S$  do not justify its use. These results highlight the inner spectral and temporal differences of frequently used sound sources. A statistically-based refinement method is also a viable approach, but the slight gain in precision does not seem to justify the additional computational complexity. Further studies should test these methods on other sets of RIRs, including Ambisonics measurements of Salle Pleyel, and adjust the different important parameters such as window lengths, for an extended set of indices.

One may question the expected outcome of such detection algorithms. In the case of complex RIRs such as those with a precursor, what is the most relevant onset time from a perceptual point of view ? Finding the beginning of the precursor may not be the best choice, since the precursor could be inaudible due to temporal masking effects. Such studies that are definitely beyond the scope of this letter, would require extensive listening tests. However, they remind us that these automatically-generated signal processing indices are only meaningful if they provide information that has a *perceptual* relevance.

## 3.4 Caractérisation de sources sonores couramment employées

### 3.4.1 Préambule

Nous présentons ici différentes méthodes d'estimation du son direct de réponses impulsionales mesurées dans des salles. Connaître le son direct (c'est-à-dire ses limites temporelles) c'est connaître la source sonore qui émet l'impulsion dans la salle. Ceci peut s'avérer très utile lorsqu'on ne connaît pas précisément cette source. C'est le cas des impulsions de revolver, de pétards, des éclatements de ballons de baudruche, etc. Ces sources sont pratiques à employer, et ne nécessitent que peu de moyens technologiques comparativement à l'émission d'un sinus glissant par exemple.

En présentant et évaluant plusieurs méthodes d'estimation du son direct, cette étude met en évidence la reproductibilité de la forme d'onde de ces sources. En revanche, la reproductibilité de la directivité n'est pas étudiée ici, car seuls des enregistrement monophoniques sont pris en compte. Cependant, les méthodes mises en oeuvre peuvent facilement être étendues aux cas des mesures spatiales.

Les travaux présentés ici ont été présentés lors de la conférence de l'Institute Of Acoustics, à Oslo (Norvège) en novembre 2008 dans un acte de congrès intitulé : *Characterizing sound sources for room acoustical measurements* (G.Defrance, L.Daudet and J-D. Polack) [47].

---

### 3.4.2 Abstract

This paper explores two methods for characterizing frequently used sound sources in room-acoustical measurements : gun shots, balloons, and crackers. As these sources are usually recorded in situ, reflections are superimposed to the direct sound and make it difficult to accurately delimit the start and end points of the direct sound. Of the two methods presently investigated, Short Time Fourier Transform offers a poor precision, whereas Matching Pursuit, based on the simulation of the impulse response by delayed and scaled versions of the direct sound calculated iteratively, gives accurate estimates. Goodness of fit is measured by the similarity between indices measured with the original impulse response and with the synthesized response, and the temporal boundaries of the direct sound are estimated by minimizing the number of iterations ran by Matching Pursuit. Time-frequency analysis, for example with the Wigner-Ville distribution, gives then information on the spectral and temporal properties of the direct sound.

### 3.4.3 Introduction

In room acoustics, room impulse responses (RIR) are measured to characterize quantitatively and qualitatively the acoustics of a hall. A convenient manner to measure RIRs is to use impulsive sound sources, such as : balloon bursts, gun shots, and crackers. However, unlike sweep or white noise, this kind of sources is known to be hardly reproducible. That is, the spectral and temporal

components may differ from a measure to another. According to the nature of the source, sound source impulses do not have the same duration [61] [50] [48]. Moreover, for some experimental RIRs carried out with such sources, it becomes difficult to identify clearly these boundaries when the sound source is not recorded in the near field, or in a damped chamber. The identification of these boundaries may vary from an expert to another, since the background noise may often disturb the readability of the signal, as seen in Figure 5.5.

This study investigates two methods for estimating the impulse duration of some impulsive sound sources (cited above). The knowledge of the direct sound provides useful information on the sound source itself, and allows to whiten the RIR. This study aims at going one step further in the characterization of some frequently used sound sources in room acoustics measurements.

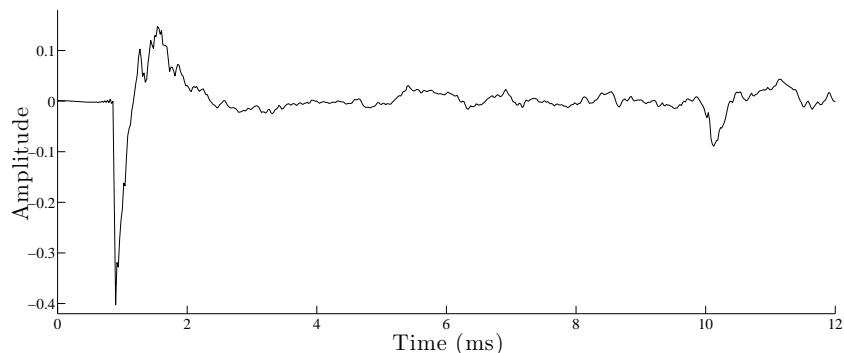


FIG. 3.4 – Experimental RIR for which determining the direct sound is not obvious.

Section 3.4.4 proposes to visualize temporal and spectral differences of such sources using the smoothed Wigner-Ville distribution. Section 3.4.5 presents two methods for estimating the temporal boundaries of the direct sound of experimental RIRs. The first method (Section 3.4.5.1) consists in low-pass filtering the local spectral energy of the signal, using the Short Time Fourier Transform (*STFT*), with high precision in time. Since this method shows well-known limitations, a second method, Matching Pursuit (MP) (Section 5.2.4.1), focuses on another approach, based on the inter-correlation between the RIR and the direct sound itself [62]. The estimation of the direct sound duration is achieved by learning the dictionary of atoms needed by MP. Section 5.2.6 presents experimental results derived from a hundred of measurements of RIRs in salle Pleyel carried out with balloon bursts on the one hand, and with gun shots and crackers on the other hand [52].

Finally, the last section (Section 3.4.8) compares the results to impulses duration measured in a damped chamber, as well as to the relevance of each method, and discusses guidelines for choosing the best sound source.

#### 3.4.4 Visualizing the direct sound using the Wigner-Ville distribution

This section aims at providing an original visualization and detection of impulses in noise in the time and frequency domains. A time-frequency energy distribution which is particularly indicated for this purpose is the Wigner-Ville distribution (*WVD*), which is defined in [63] as :

$$W_x(t, \nu) = \int_{-\infty}^{+\infty} x(t + \tau/2) \cdot x^*(t - \tau/2) \cdot e^{-2j\pi\nu\tau} d\tau \quad (3.7)$$

This distribution satisfies a large number of desirable mathematical properties. In particular, the *WVD* is always real-valued, on the one hand, and it preserves time and frequency shifts, on the other hand.

Unlike spectrogram, the *WVD* has non-zero interference terms regardless of the time-frequency distance between two signal terms. These interference terms are troublesome since they may overlap with auto-terms (signal terms) and thus make it difficult to visually interpret the *WVD* image. This drawback is solved by using a smoothing function, as presented in [58]. Figure 3.5 presents three examples of the direct sound of experimental RIRs measured with a balloon burst, a gun shot and a cracker impulse, respectively. As a first observation, this visualization highlights the longest duration of the balloon burst compared to the gun shot and to the cracker. Moreover, the time signal of the balloon burst presents large and long oscillations, making difficult to identify the temporal boundaries of the impulse precisely. Frequency components of the balloon burst are evolving with time, which is somehow related to the tearing of the balloon during its burst. The cracker impulse is short and relatively well localized in the frequency domain. However, its spectral contents is impoverished compared to the gun shot. Indeed, even if this latest is not as brief as the cracker impulse, it covers a larger frequency range. The three sources studied here are all mainly low frequency .

### 3.4.5 Methods for estimating the direct sound duration

#### 3.4.5.1 Estimating the direct sound duration with STFT

When observing the temporal structure of a RIR, it is noticeable that the occurrence of the direct sound is accompanied by a sudden increase of amplitude, followed by a decay. The method proposed here aims at analyzing mean variations of spectra from the beginning (the onset) to the end (the offset) of the direct sound, windowing the signal. Temporal boundaries are estimated using an envelope follower, which can easily be constructed by low-pass filtering the local energy [55]. The onset time is estimated as follows. The maximum of  $E(n)$  is detected, and the signal analyzed from its beginning to its maximum ( $t_{max}$ ). The ratio of two successive windows is calculated ; the index ( $T_{On}$ ) is found when the ratio is maximum (Eq.3.8). While the offset is estimated as follows. The signal is analyzed from its maximum up to 5ms from the onset time, assuming a duration inferior or equal to 5ms. The ratio of two successive windows is calculated ; the index ( $T_{Off}$ ) is found when the ratio is minimum (Eq.3.8). This can be written as :

$$\tilde{E}(n) = \sum_k |X(n.h, k)| \quad (3.8)$$

$$t_{On} = h \times argmax_n(\tilde{E}(n+1)/\tilde{E}(n)) \quad (3.9)$$

$$t_{Off} = h \times argmin_n(\tilde{E}(n+1)/\tilde{E}(n)), n > t_{max} \quad (3.10)$$

where  $X(n, k)$  is the Short Time Fourier Transform (*STFT*) of the signal  $x(n)$ , and  $h$  the time step between two windows.

In practice, we choose a window analysis width (that is the temporal resolution) of  $h = 0.1ms$ .

#### 3.4.5.2 Matching Pursuit applied to RIR

Since details and issues of using MP on RIRs has been developed in [64], this section briefly presents theoretical and practical aspects of MP, and focuses on the method for estimating the direct sound. A RIR can be seen as a linear set of occurrences of the direct sound translated in time, filtered by reflections on the surfaces of the hall. For this latter reason, it is believed that MP can help for understanding more deeply the architecture of a RIR, since this algorithm, introduced by [62], provides information, which can be seen as maxima of correlation (Eq.6.20) [65] between two signals : the RIR and a dictionary composed of the direct sound (the mother

atom) delayed in time. The temporal boundaries of the direct sound are estimated by defining precisely the mother atom, that is, by learning the dictionary (Section 5.2.4.3).

**Theoretical reviews** In theory, any signal  $x$  can be perfectly decomposed in a linear set of atoms for an infinity of iterations.

---

*Algorithm of Matching Pursuit*

---

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

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

$$n = 0$$

$$R^0 x \leftarrow x$$

Repeat (until a stopping criterium is reached) :

$$\gamma_{opt}^{(n)} \leftarrow argmax_{\gamma \in \Gamma} |\langle R^n x, \phi_\gamma \rangle| \quad (3.11)$$

$$\alpha^{(n)} \leftarrow |\langle R^n x, \phi_{\gamma_{opt}^{(n)}} \rangle| \quad (3.12)$$

$$R^{n+1} x \leftarrow R^n x - \alpha^{(n)} \phi_{\gamma_{opt}^{(n)}} \quad (3.13)$$

$$n \leftarrow n + 1 \quad (3.14)$$

$$x^{(n)} = \sum_{k=0}^{n-1} \alpha^{(k)} \phi_{\gamma_{opt}^{(k)}} \quad (3.15)$$


---

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

In practice, this number of iterations must be finite and a stopping criterium has to be set. The authors propose to use the signal/noise ratio (*SRR*) in dB of  $x$  over the residual ( $R$ ). The quality of the decomposition of  $x$  in atoms depends on the value of *SRR*. On the one hand, for a *SRR* too low, 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 *SRR* too high 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 [41], calculated on  $x$  to those calculated on  $x^{(n)}$  for different values of a *SRR* allows to set the stopping criterium on a rational ground. This is achieved by running MP on a RIR for which the determination of the direct sound is obvious. The acoustical indices used here are the reverberation times at 20dB ( $RT_{20}$ ), 30dB ( $RT_{30}$ ), the Early Decay Time at 10dB ( $EDT_{10}$ ), and the Central Time ( $T_C$ ). According to [64], a convenient *SRR* would be 20dB, since variations of acoustical indices lie under 5%. Figure 5.4 shows a RIR, its direct sound, the linear set of coefficients and the residual.

### 3.4.5.3 Detection of the direct sound : learning the dictionary of atoms ( $\phi_\gamma$ )

It is extremely important to know the exact temporal boundaries of the direct sound, since the number of iterations of MP is the lower for a more precise atom definition, and bigger for a totally approximative atom. A fine estimation is possible using a method of grid search, as follows.

As in Section 3.4.5.1, we assume the impulse duration to be inferior to 5ms, defining the mother atom  $\gamma$ . The time index ( $t_0$ ) of the maximum of the mother atom is detected. A dictionary of atoms  $\phi_\gamma$  is constituted of atoms with temporal boundaries that are varying (with a step

of  $0.1ms$ ) from  $0ms$  to  $t_0$  and from  $t_0$  to  $5.00ms$  for the first and last indices respectively. For each couple of [first index : last index], MP is ran. The temporal 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.

### 3.4.6 Reference impulses

For the three types of sound sources, 20 measurements of their impulse is carried out in a damped room with a  $1/2"$  omnidirectionnal microphone B&K (type 4191), a amplifier B&K Nexus 2692C (calibrated at  $316\mu V/Pa$ ) and a numerical recorder NAGRA V. Recordings were made at a rate of  $48kHz$  and for  $24bits$ . The receiver is placed normally to the sound source, that is : pointing towards the piercing point of the balloon, perpendicular to the barrel of the gun, and facing the cracker. Temporal boundaries are estimated at hand for each impulse. Figure 3.7 shows the different durations measured. Variations of results are a consequence of the poor reproducibility of these sources.

### 3.4.7 Results

The two detection methods are tested over 100 audio wavfiles. These RIRs have been measured in Salle Pleyel in Paris [52], according to ISO3382 [41], with balloon bursts, gun shots and cracker impulses. The sound sources were of the same type than those used in the damped room. For each RIR, the onset and offset times are estimated by each method (figures 3.8-3.9), and compared to the references (figure 3.7). Statistics of the results are presented as interquartiles. Boxes have lines at the lower quartile, median, and upper quartile values. Whiskers show the extent of the rest of the data, while outliers are data with values beyond the ends of the whiskers. On the one hand, the two methods return results that are in agreement with the reference durations : balloon bursts last longer than gun shots and cracker impulses. Durations of balloon bursts estimated present larger variations than for the two other kind of sources. On the other hand, time durations estimated by MP are closer (in terms of average and spreading) to the references than those estimated using the *STFT*. This point is further discussed in Section 3.4.8. The good agreement between results from MP and references is a hint for choosing this correlation based method to estimate the direct sound duration.

### 3.4.8 Discussion

Large variations of results are not only related to the robustness of the presented estimators, but also to the lack of reproducibility of the considered sound sources. According to results, the shorter and more reproducible source is the cracker (since variations are the smallest), followed by the gun shot and then by the balloon burst. Large spreading of results returned by the *STFT* method are a consequence of the drawback of this method : that is, its resolution. Indeed, the time resolution is proportional to the effective duration of the analysis windows. Similarly, the effective frequency resolution is proportional to the effective band-width of the analysis window. Consequently, for the *STFT* a *trade-off* between time and frequency resolution has to be made : on the one hand, a good time resolution requires a short window analysis ; on the other hand, a good frequency resolution requires a long window. This limitation leads to choose another method.

It is believed that the estimation of the impulse duration using MP provides more accurate and realistic results than the *STFT* method. First, results are closer to the references, and second, MP does not undergo any frequential or temporal constraint. However, MP seems to underevaluate impulse durations, except for the crackers. This can be explained by the fact that

reference impulses duration have been determined by hand, and the determination of the offset mixed into the background noise leads to variations, as mentionned in Introduction. Furthermore, differences between experimental conditions of measurements in the damped room and in Salle Pleyel also explain these variations of estimation.

The three kind of sources studied in this article are certainly not recommended for accurate RIRs measurements, because of the lack of reproducibility in the time and frequency domains, yet they are often used by practitioners. Temporal and frequential properties of balloon bursts depend on the volume of the balloon, on the homogeneity of the latex, and overall on the location of the piercing point relatively to the microphone. Guns shots are also poorly reproducible for two main reasons. First, during the explosion of the bullet, the sound wave not only escapes the barrel, but also the cylinder, since it is in fact a revolver. Second, the bullets used with gun are not exactly filled with the same amount of powder, according to manufacturers. Among the three sources studied, cracker impulse is the shortest, and the more reproducible. However, the drawback of this source is its impoverished spectra, mainly a low frequency spectrum. As for the gun, variations of durations of the cracker impulse can be explained by its manufactory issues. All these parameters have a strong influence on the directivity, on the impulse duration and on frequency components of the sources.

### 3.4.9 Conclusions

Three different types of frequently used sound sources are studied : balloon bursts, gun shots and cracker impulses. Their relative duration are estimated using a method based on the *STFT* and Matching Pursuit. Results are compared to some refrence impulses measured in a damped room. Matching Pursuit returns the best results. This method is believed to be more accurate to determine the temporal boundaries of the direct sound of an experimental RIR than a method based on *STFT*. High disparities in results, specially for the balloon bursts and the guns shots, underlies their poor reproducibility in time. Cracker impulses are the sharpest, followed by gun shots and then by balloon bursts. Results confirm that these sound sources must be used with knowledge of their physical limitations. Finally, further studies should investigate on the one hand, the effect of window analysis width used in the two presented methods, and on the other hand, the directionnality of such sound sources.

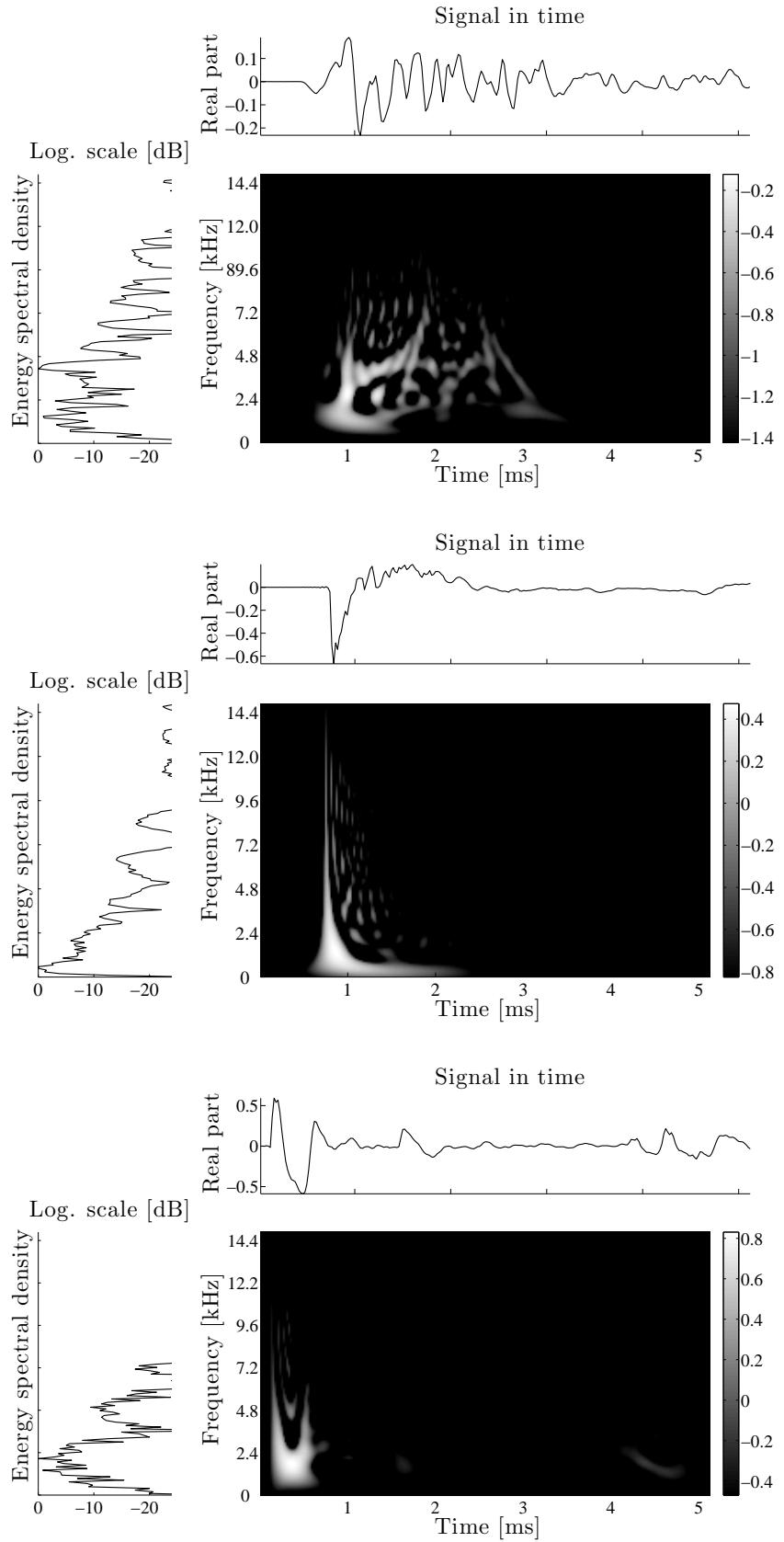


FIG. 3.5 – Smoothed Wigner-Ville distributions of the direct sound of three experimental RIRs measured with : - a balloon burst (top) ; - a gun shot (middle) ; - a cracker (bottom) (at 48kHz).

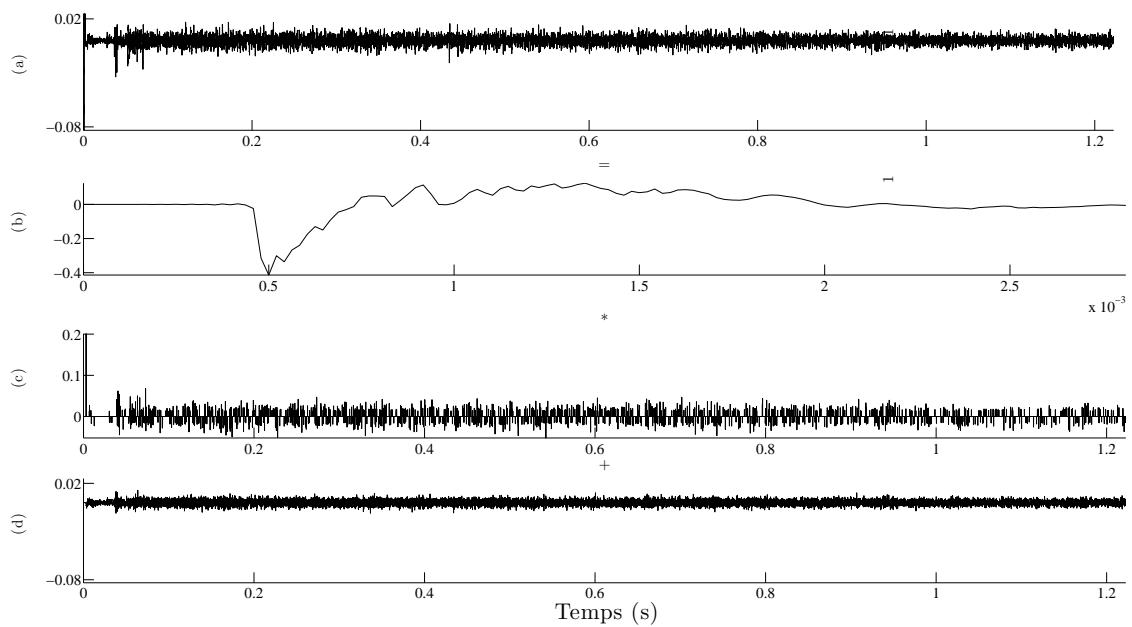


FIG. 3.6 – Matching Pursuit ran on an experimental RIR ( $SRR = 20dB$ ) [ (a) : experimental RIR, (b) : direct sound, (c) : linear set of coefficients  $(\gamma_{opt}^{(n)})$ , (d) : residual]. Note this is a slightly modified version of the original figure.

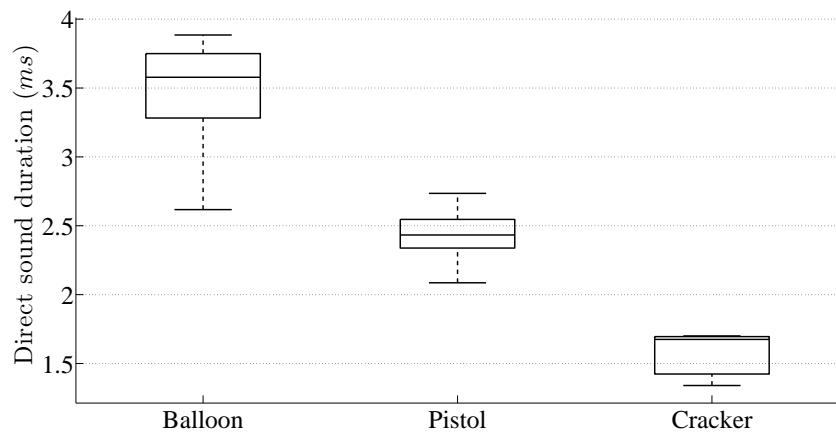


FIG. 3.7 – Impulse duration of balloon bursts, gun shots, and crackers measured in a damped room.

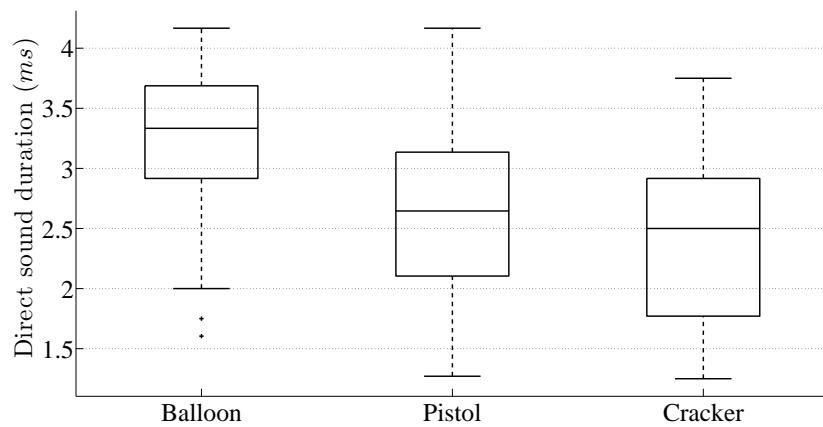


FIG. 3.8 – Impulse duration calculated on RIRs using the *STFT* method.

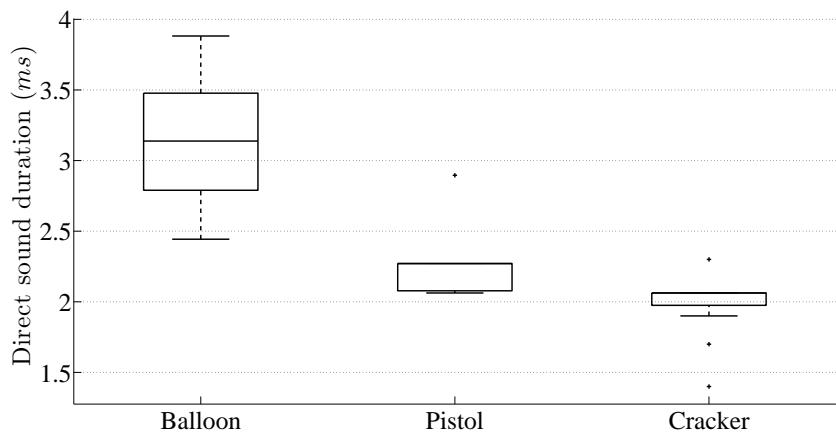


FIG. 3.9 – Impulse duration calculated on RIRs using Matching Pursuit ( $SRR = 20dB$ ).



## Chapitre 4

# Détection du temps de mélange à l'aide de tests de Gaussianité

Dans le Chapitre 2, nous avons présenté le temps de mélange tel qu'il est défini en acoustique des salles et selon le critère de Polack. Nous avons également noté qu'une autre définition de ce temps particulier existe et est relative à la théorie des systèmes dynamiques. Selon cette dernière, le temps de mélange est le temps nécessaire à deux rayons initialement adjacents au point d'émission (i.e., le point source) pour se disperser uniformément –pour que la densité de rayons soit homogène et isotrope– dans l'espace.

Dans les trois prochains Chapitres nous nous intéressons à définir plusieurs estimateurs objectifs du temps de mélange, tel qu'il est défini en acoustique des salles : temps de transition entre les régimes des premières réflexions et la réverbération tardive.

Dans un premier temps, nous tentons une estimation du temps de mélange en détectant le temps auquel le signal de la réponse impulsionale devient Gaussien. Dans un second temps (Chapitre 5), nous détournons l'emploi d'un algorithme utilisé en décomposition parcimonieuse de signaux audio afin de détecter les retours présents dans le signal. Nous montrons comment l'étude de la statistique des retours estimés permet d'estimer le temps de mélange. Dans un troisième temps (Chapitre 6), nous présentons un outil de traitement du signal (l'eXtensible Fourier Transform – XFT) que nous avons développé. Cet outil temps/fréquence permet de visualiser l'évolution temporelle des composantes spectrales d'un signal. Notre idée est de détecter le temps à partir duquel les retours sont statistiquement distribués, temps à partir duquel le signal adopte une distribution Gaussienne. Enfin, dans un dernier Chapitre (7) nous revenons sur les différents résultats que nous obtenons à l'aide de ces estimateurs et les confrontons aux différentes définitions du temps de mélange que l'on trouve dans la littérature.

### 4.1 Introduction

Dans ce Chapitre, nous tentons une approche simple de la détection du temps de mélange, basée sur la théorie statistique de l'acoustique des salles. Rappelons que le temps de mélange, en acoustique des salles, définit le temps de transition entre deux régimes de la réponse impulsionale : celui des premiers retours, et celui de la réverbération tardive. Selon cette dernière, la réverbération tardive serait distribuée Gaussiennement, puisque dans le champ diffus le spectre de la réponse impulsionale suit une même distribution. L'idée de mettre en œuvre des tests de Gaussianité afin de détecter le début de la réverbération tardive a déjà été employée par Stewart *et al.* et Defrance *et al.* [66, 67].

Dans la suite, nous détectons les instants où le signal suit localement une distribution Gaussienne. Dans un premier temps, nous estimons les valeurs prises par les moments d'ordre 2 et 4 (écart-type et kurtosis) d'une loi normale, puis nous testons les instants auxquels ces paramètres prennent des valeurs nulles - définissant ainsi les temps auxquels le signal suit une loi normale.

Dans un deuxième temps, nous mettons en oeuvre des tests de Gaussianité, tels que le test du  $\chi^2$  ou de Lilliefors, afin de détecter les temps auxquels la distance mesurée entre la distribution du signal et celle d'un loi normale est inférieure à un seuil de tolérance que nous fixons. Nous employons également le test de Kolmogorov-Smirnov qui nous permet de tester l'écart mesuré entre la distribution de deux échantillons du signal d'une réponse impulsionale. L'idée sous-jacente à l'emploi d'un tel test est de détecter le temps à partir duquel le signal suit une distribution à peu près identique, selon un seuil de tolérance que nous nous fixons.

Nous montrons ici que l'emploi des tests de Gaussianité n'est pas suffisante (mathématiquement) et physiquement. Mais ce n'est pas là notre principal intérêt. En effet, nous mettons en évidence le phénomène de ce que nous appellerons des *bouffées de cohérence*, c'est-à-dire, des instants où le signal de la réponse s'écarte fortement d'une distribution Gaussienne, alors que le champ diffus est vraisemblablement établi. A ces moments, le système redevient cohérent avec son état initial.

## 4.2 Les cumulants d'une loi normale

Il paraît inutile de rappeler toutes les propriétés d'une distribution normale. Un certain nombre d'ouvrages sont bien plus précis et complets que ce que nous pourrions dire ici en quelques lignes. Rappelons néanmoins qu'une distribution Gaussienne (centrée) présente une moyenne nulle, et approximativement deux-tiers des observations sont compris entre  $+\sigma$  et  $-\sigma$  ( $\sigma$  étant l'écart-type). L'écart à une Gaussienne est mesuré par une symétrie  $s$  par rapport à la moyenne (*skewness* en anglais), et un aplatissement mesuré par le kurtosis  $k$ .

La loi normale (ou de Laplace-Gauss) est une loi qui s'applique à une variable statistique qui est la résultante d'un grand nombre de causes indépendantes, dont les effets s'additionnent et dont aucune n'est prépondérante.

Nous disons qu'une variable aléatoire  $X$  suit une loi normale (ou de Laplace-Gauss) de moyenne  $\mu$  et de variance  $\sigma^2$  si sa densité de probabilité est donnée par :

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left[-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right], \quad (4.1)$$

où  $\sigma$  est également appelé l'écart-type de la distribution. La fonction de répartition, qui représente la probabilité que la variable aléatoire ait une valeur inférieure à  $x$ , a pour expression :

$$F(x) = P\{X < x\} = \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^x \exp\left[-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right] dx. \quad (4.2)$$

La loi normale est d'application très générale. Elle est, en effet, engendrée sous la condition très peu restrictive de variance finie, par *l'addition* de variables aléatoires indépendantes et de même loi. Il s'agit là d'un théorème important du calcul de probabilités : le *théorème central limite*, dont nous donnons ici un énoncé heuristique :

*Soit  $X_1, \dots, X_n, \dots$  une suite de variables aléatoires de même loi, faiblement corrélées, de variance finie, alors :*

$$P\left(\frac{(X_1 + \dots + X_n) - n\mu}{\sigma\sqrt{n}} < x\right) \xrightarrow{n \rightarrow \infty} F(x), x \in \mathbb{R}. \quad (4.3)$$

### 4.2.1 L'écart-type $\sigma$

Notons que dans une distribution normale centrée, approximativement deux-tiers des observations sont compris entre  $[-\sigma; +\sigma]$ , alors qu'approximativement un tiers de ces observations est compris entre  $]-\infty; -\sigma[$  et  $]\sigma; +\infty[$ . D'après la théorie des probabilités appliquée à l'acoustique des salles, le mélange est atteint lorsque la variable  $R$ , définie par l'Eq. 4.4 est égale à 2 :

$$R(n.h) = \sum_{n=0}^{N-1} \frac{N_{x(n.h)}[-\sigma; \sigma]}{N_{x(n.h)}[-\infty; -\sigma] + N_{x(n.h)}[\sigma; +\infty]}, \quad (4.4)$$

où  $N_{x(n.h)}[a; b]$  est le nombre d'observations du signal  $x$  (fenêtré par un fenêtre de longueur  $h$ ) dont l'amplitude est comprise entre  $a$  et  $b$ .

### 4.2.2 Le kurtosis $k$

Le cumulant du quatrième ordre d'un processus est appelé *kurtosis*, et est défini (après normalisation) par :

$$k = \frac{E(x - \mu)^4}{\sigma^4} - 3, \quad (4.5)$$

où  $E$ ,  $\mu$  and  $\sigma$  sont l'espérance, la moyenne et l'écart-type de  $x$  respectivement. Pour une distribution normale,  $k = 0$ . En théorie, cette valeur est donc prise lorsque le système a atteint un état diffus.

La Fig. 4.1 montre l'estimation de l'écart-type et du kurtosis sur une même réponse impulsionnelle. Outre la mise en évidence des retours de plus forte énergie, il apparaît nettement que les estimateurs passent par zéro avant le début de la réverbération tardive (dont nous estimons approximativement le début à l'oeil vers  $80ms$ ), et montrent des pics, des sauts, une fois le champ diffus établi, correspondant à des bouffées de cohérence. Ces dernières seraient liées à des retours fortement corrélés à l'état initial du système, et mériteraient une attention particulière dans le futur (Section 4.4). Une hypothèse vraisemblable serait que ces retours particuliers proviennent des plus longues dimensions de la salle. Ainsi, leur temps de parcours avant d'atteindre le récepteur peut être long, plus long que celui nécessaire pour que le champ diffus s'établisse. De plus, il est parfaitement concevable que ces bouffées ne soient pas observables partout dans la salle. La probabilité d'observer ces bouffées pourrait constituer le sujet d'études futures.

Il est important de souligner que la seule estimation de ces cumulants ne suffit pas à détecter de manière certaine le début de la réverbération tardive. En effet, la réalisation des hypothèses précédentes n'est mathématiquement pas suffisante pour établir que la distribution est normale. De plus, la Fig. 4.1 met en évidence des moments où le système suit une distribution normale dans les premiers instants puis passe à un autre comportement. Nous pouvons émettre l'hypothèse que ce phénomène est observable en présence de diffusion, et/ou de bruit de fond<sup>1</sup>. Ceci ne permet donc pas d'affirmer avec certitude que le mélange est atteint. Enfin, il faut souligner que les valeurs prises par  $R$  et  $k$  dépendent principalement de la taille de fenêtre d'analyse, et qu'à la connaissance de l'auteur, il n'existe aucun argument justifiant l'emploi d'une longueur de fenêtre plutôt qu'une autre.

Nous pourrions très bien continuer notre investigation de Gaussianité en calculant le critère de symétrie ou des moments d'ordre plus élevés, mais là encore, ce ne seront jamais des arguments mathématiquement suffisants. Des tests de Gaussianité plus robustes vont donc être proposés.

---

<sup>1</sup>De plus, ce n'est pas parce que nous mesurons localement que le système est diffus que le système entier est entré dans le régime de la réverbération tardive.

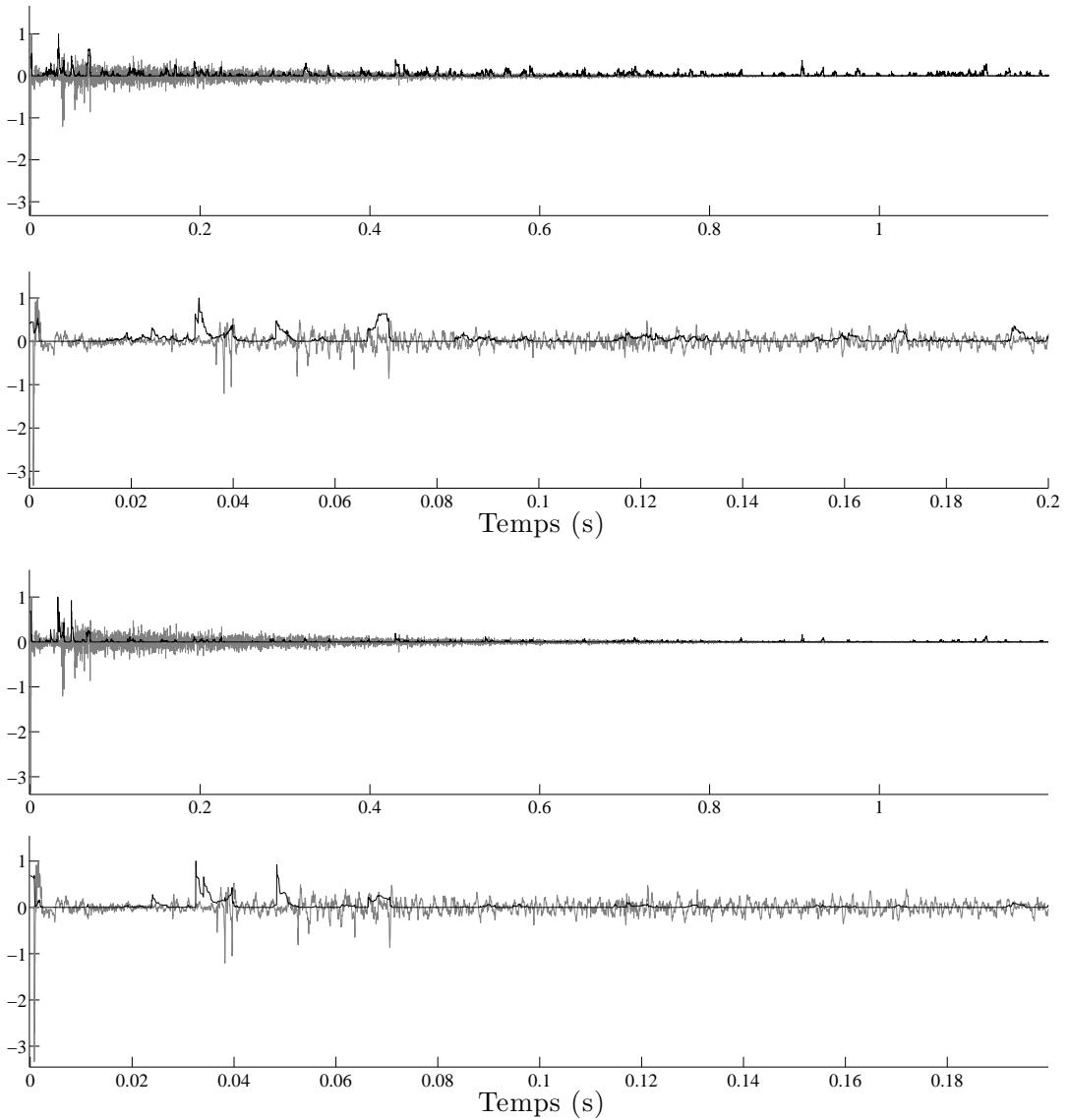


FIG. 4.1 – Haut (2 premiers graphiques) : Ecart-type ( $R$ ) (noir) d'une réponse impulsionale expérimentale (gris) calculé sur  $N = 200$  points (fréquence d'échantillonnage de  $F_s = 48kHz$ ) (signal entier et détails). Pour une meilleure lisibilité, on a tracé  $R - 2$ . Bas (2 derniers graphiques) : Kurtosis (noir) d'une réponse impulsionale expérimentale (gris) calculé sur  $N = 200$  points ( $F_s = 48kHz$ ) (signal entier et détails). Note : les signaux sont normalisés et les estimateurs passant par 0 marquent la réalisation de l'hypothèse de Gaussianité. On visualise nettement les retours de plus fortes énergie, ainsi que la présence de bouffées de cohérence une fois le champ diffus établi. Enfin, le signal utilisé ici est le même sur le graphique du haut et du bas.

### 4.3 Tests statistiques

L'idée d'employer des tests statistiques, que nous présentons par la suite, repose sur la théorie statistique de l'acoustique des salles. Nous cherchons un moyen d'identifier le temps auquel le système voit son comportement changer. En d'autre termes, nous cherchons le temps de passage

d'une distribution à une autre. Pour cela, plusieurs tests statistiques peuvent être mis en œuvre<sup>2</sup> :

- test du  $\chi^2$  ;
- test de Lilliefors ;
- test de Kolmogorov-Smirnov sur deux fenêtres adjacentes.

Les tests du  $\chi^2$  et Lilliefors testent l'adéquation entre la distribution des individus considérés et celle d'une distribution normale à partir de tables de valeurs, alors que le test de Kolmogorov-Smirnov teste l'adéquation entre deux distributions empiriques obtenues sur des fenêtres adjacentes.

### 4.3.1 Tests d'ajustement

Supposons qu'une variable aléatoire statistique  $X$  suive très exactement une loi de probabilité  $P$ . Si nous tirons un échantillon dans la population correspondant à cette loi, la distribution observée s'écartera toujours plus ou moins de la distribution théorique : les observations sont en effet entachées de fluctuations aléatoires.

Généralement, nous ne connaissons ni la forme de la loi de probabilité suivie par le phénomène observé, ni la valeur des paramètres de cette loi. C'est une réflexion sur la nature du phénomène et l'analyse de la distribution observée qui permet de choisir la loi de probabilité, celle qui paraît convenir. Dans notre cas, au regard de la théorie de l'acoustique statistique des salles, nous choisissons une loi normale.

Les écarts entre la loi théorique et la distribution observée peuvent être attribués :

- soit aux fluctuations d'échantillonnage,
- soit au fait que le phénomène observé ne suit pas en réalité la loi supposée.

Grossièrement, nous admettrons que les écarts sont imputables aux fluctuations aléatoires si ceux-ci sont suffisamment faibles. En revanche, s'ils sont trop élevés, nous en concluons qu'ils ne peuvent être expliqués par ces seules fluctuations et que, par conséquent, le phénomène ne suit pas une loi normale (dans notre cas).

Juger de la validité d'un ajustement consiste à tester l'hypothèse que le phénomène observé suit la loi théorique supposée. Pour ce faire, il faut d'abord définir une mesure de la distance existante entre la distribution empirique et la loi de probabilité théorique, et ensuite déterminer la loi de cette distance sous hypothèse d'adéquation. Connaissant cette loi, si nous constatons que dans l'hypothèse retenue, il y a une forte probabilité d'obtenir, par le seul fait des fluctuations aléatoires, une distance supérieure à celle observée, nous accepterons l'hypothèse, et nous admettrons que le phénomène suit bien la loi théorique supposée. Si au contraire, cette probabilité est faible (moins de 5% par exemple) il y a de grandes chances que les écarts constatés ne soient pas uniquement imputables aux fluctuations aléatoires, mais plutôt à l'inadéquation de la loi théorique retenue pour représenter le phénomène – dans ce cas, nous rejetons l'hypothèse.

#### 4.3.1.1 Test du $\chi^2$

Le test d'homogénéité du  $\chi^2$  consiste à se demander si une liste de valeurs observées peut dériver d'une loi de probabilité donnée.

Supposons que la variable aléatoire  $X$  suive une loi de probabilité  $P$  qui fait correspondre à  $k$  modalités ou classes :  $C_1, C_2, \dots, C_k$  les probabilités  $p_1, p_2, \dots, p_k$ , et par suite, les effectifs théoriques  $Np_1, Np_2, \dots, Np_k$ . Nous sommes alors en mesure de calculer la valeur  $d$  :

$$d = \sum_{i=1}^k \frac{(N_i - Np_i)^2}{Np_i}, \quad (4.6)$$

---

<sup>2</sup>Cette liste n'est pas exhaustive. En effet, d'autres tests pourraient être considérés, mais les conclusions auxquelles nous aboutirions seraient les mêmes.

prise par la variable aléatoire  $D$ , qui mesure la distance existant entre la distribution observée et la distribution théorique. Dans l'hypothèse où la distribution théorique est effectivement la loi  $P$ , cette distance  $D$  suit asymptotiquement une loi de  $\chi^2$  à  $\nu$  degrés de liberté :

$$\chi^2 = \sum_{i=1}^m \frac{(O_i - E_i)^2}{E_i}, \quad (4.7)$$

où  $O_i$  est la  $i$ ème valeur observée,  $E_i$  est la valeur attendue sous l'hypothèse d'homogénéité avec une distribution normale.

Le nombre  $\nu$  de degrés de liberté de cette loi dépend du nombre  $k$  de classes et du nombre  $r$  de paramètres estimés à partir des observations :  $\nu = k - r - 1$ . Pour l'ajustement d'une loi normale dont on estime les paramètres  $m$  par la moyenne observée  $\bar{x}$  et l'écart-type  $\sigma$  par la variance empirique, nous avons entre les  $N_i$  les relations :

$$m = \frac{1}{N-1} \sum_{i=1}^k N_i x_i \quad (4.8)$$

$$\sigma^2 = \frac{1}{N-1} \sum_{i=1}^k N_i (x_i - m)^2. \quad (4.9)$$

Dans le cas d'une loi normale,  $r = 2$ , soit  $\nu = k - 3$ .

Le test du  $\chi^2$ , introduit par K. Person, repose sur le raisonnement suivant<sup>3</sup>. Nous faisons l'hypothèse que le phénomène observé suit –dans notre cas– une loi normale. Dans ces conditions, la distance  $D$  est, du fait des fluctuations aléatoires, une variable  $\chi^2$  à  $\nu = k - 3$  degrés de liberté.

S'il y a une forte probabilité (déterminée par la consultation de la table de  $\chi^2$ ) que  $D$  prenne une valeur supérieure à la valeur  $d$  observée, les fluctuations aléatoires suffisent à expliquer la distance enregistrée. L'hypothèse (que l'on nomme hypothèse nulle) est jugée acceptable.

Si, en revanche, il n'y a qu'une faible probabilité (par exemple inférieure à 5%) d'obtenir une valeur  $D$  supérieure à la valeur  $d$  observée, il est beaucoup plus probable que cette valeur soit due à l'inadéquation de la loi théorique (loi normale). Nous rejetons alors l'hypothèse que le phénomène observé suit la loi normale – l'hypothèse nulle est rejetée.

Dans la pratique, pour que la distance  $D$  suive une loi de  $\chi^2$ , les conditions de convergence impliquent que les effectifs théoriques  $Np_i$  des différentes classes ne soient pas trop petits. On considère qu'ils doivent être au moins égaux à 4 ou 5. Par suite, il est parfois nécessaire de regrouper certaines classes, notamment aux extrémités de la distribution. Enfin, signalons que les seuils de probabilité à partir desquels on décide de rejeter l'hypothèse sont de 2 à 5%. Nous choisissons ici un seuil de 5%.

La Fig. 4.2 présente la probabilité de réalisation de l'hypothèse nulle, contre celle non-nulle, sur une réponse impulsionnelle expérimentale pour deux tailles de fenêtres différentes : 200 et 120 échantillons (cette dernière durée correspondant à la durée du son direct). Nous mettons clairement en évidence la dépendance entre les résultats et la taille de la fenêtre d'analyse. De plus, il nous faut remarquer que les résultats obtenus dépendent également du nombre de classes choisi, c'est-à-dire de la taille de ces classes.

Ce test ne permet pas d'estimer de manière robuste le temps de mélange. En effet, il est difficile de trouver un critère permettant de décider de ce temps : l'hypothèse nulle est vérifiée dans les tous premiers instants de la réponse et ensuite par intermittence. De plus, les instants où le signal suit localement une distribution proche de celle d'une loi normale dépendent également de la marge d'erreur que nous fixons (ici 5%). L'utilisation d'une autre valeur changera naturellement les résultats que nous présentons ici.

---

<sup>3</sup>Ce type de raisonnement statistique est désigné sous le nom de *test d'hypothèse*.

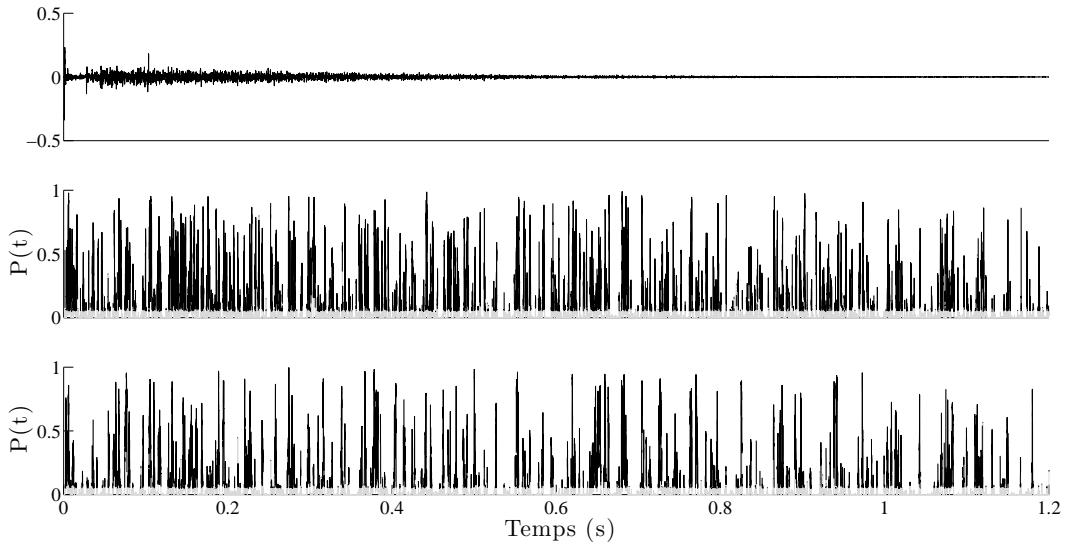


FIG. 4.2 – Haut : réponse impulsionnelle expérimentale (fréquence d'échantillonnage de  $F_s = 48kHz$ ) (signal entier et détails). Milieu : probabilité de validité de l'hypothèse nulle du test de  $\chi^2$  sur des fenêtres de taille 200 échantillons. Bas : probabilité de validité de l'hypothèse nulle du test de  $\chi^2$  sur des fenêtres de taille 120 échantillons. Note : en noir, les instants auxquels  $P(t) \geq 0.05$ . Remarquons que la validation de l'hypothèse nulle, contre celle non-nulle, dépend de la taille de la fenêtre d'analyse. De plus, quelque soit le niveau de détails auquel nous pourrions observer les graphiques, nous visualiserions toujours une zone de rejet de l'hypothèse nulle.

#### 4.3.1.2 Test de Lilliefors

Le test de Lilliefors vérifie l'adéquation entre une distribution empirique et celle d'une distribution normale à l'aide de tables de valeurs issues de simulations de Monte-Carlo, dans lesquelles le nombre d'échantillons est inférieur à 1000.

Le test de Lilliefors calcule :

$$L = \max_x |F_1(x) - F_2(x)|, \quad (4.10)$$

où  $F_1$  et  $F_2$  sont les fonctions de répartition de la variable empirique et de la distribution normale de moyenne et écart-type égaux à ceux de la variable observée, respectivement.

La Fig. 4.3 présente la probabilité de réalisation de l'hypothèse nulle sur une réponse impulsionnelle expérimentale pour deux tailles de fenêtres différentes : 200 et 120 échantillons (cette dernière durée correspondant à la durée du son direct). Nous mettons clairement en évidence la dépendance entre les résultats et la taille de la fenêtre d'analyse.

De plus, ce test ne permet pas plus que celui du  $\chi^2$  d'estimer le temps de mélange. En effet, il est difficile de trouver un critère permettant de décider de ce temps : l'hypothèse nulle est vérifiée dans les tous premiers instants de la réponse et ensuite par intermittence.

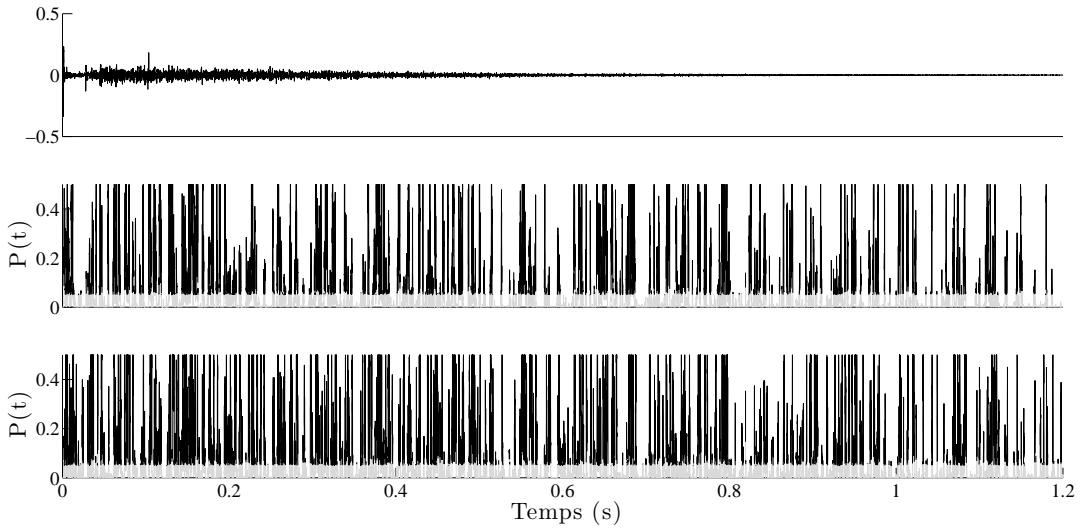


FIG. 4.3 – Haut : réponse impulsionnelle expérimentale (fréquence d'échantillonnage de  $Fs = 48kHz$ ) (signal entier et détails). Milieu : probabilité de validité de l'hypothèse nulle du test de Lilliefors sur des fenêtres de taille 200 échantillons. Bas : probabilité de validité de l'hypothèse nulle du test de Lilliefors sur des fenêtres de taille 120 échantillons. Note : en noir, les instants auxquels  $P(t) \geq 0.05$ . Remarquons que la validation de l'hypothèse nulle, contre celle non-nulle, dépend de la taille de la fenêtre d'analyse. La réponse impulsionnelle étudiée ici est la même que celle de la Fig. 4.2.

#### 4.3.2 Test de Kolmogorov-Smirnov

Le test de Kolmogorov-Smirnov<sup>4</sup> est utilisé pour les lois à fonctions de répartition continue, et permet de :

- de déterminer si un échantillon suit bien une loi donnée à répartition continue ;
- de tester si deux échantillons suivent la même loi.

Ce test repose sur les propriétés de convergence des fonctions de répartition empirique. Si  $(x_1, \dots, x_n)$  est un échantillon de  $n$  variables aléatoires indépendantes à valeurs réelles, on définit la fonction de répartition empirique<sup>5</sup> de cet échantillon par :

$$F_n(x) = \frac{1}{n} \sum_{i=1}^n \delta_{x_i \leq x}, \quad (4.11)$$

avec

$$\delta_{x_i \leq x} = \begin{cases} 1, & x_i \leq x \\ 0, & \text{sinon} \end{cases} \quad (4.12)$$

Le test de Kolmogorov-Smirnov compare les distributions de deux échantillons d'individus. Si ces deux échantillons proviennent d'une même loi, la distance  $\sup_x (|F_{n,1}(x) - F_{n,2}(x)|)$  tendra vers 0. Plus précisément, la distance  $\sqrt{n} \sup_x (|F_{n,1}(x) - F_{n,2}(x)|)$ , parfois appelée distance de Kolmogorov, suit asymptotiquement une loi dont on sait calculer numériquement la fonction de répartition. Quand cette distance est grande, la probabilité que les deux échantillons proviennent

<sup>4</sup>Ce test est, par exemple, employé pour tester la qualité d'un générateur de nombres aléatoires.

<sup>5</sup>La fonction de répartition empirique est un processus qui prend ses valeurs dans l'espace des fonctions croissantes comprises entre 0 et 1.

d'une même loi reste faible, et l'hypothèse nulle est rejetée ( $H = 1$ ) ; sinon,  $H = 0$ . Pour chaque valeur  $x$  prise par les signaux ( $X_1$  et  $X_2$ ) à comparer, le test compare la proportion des valeurs de  $X_1$  inférieure à  $x$  au regard des valeurs de  $X_2$  inférieures à  $x$ . Le test estime le maximum de différence entre chaque distribution pour chaque valeur de  $x$ . Cela revient à écrire :

$$\sup_x (|F_{n,1}(x) - F_{n,2}(x)|), \quad (4.13)$$

où les  $n, i$  sont des indices.

Ainsi, le test de Kolmogorov-Smirnov sur deux fenêtres d'échantillons teste l'adéquation entre la distribution de deux fenêtres du signal de la réponse expérimentale. L'emploi de ce test doit permettre d'estimer un temps de rupture dans la statistique du signal lorsque deux fenêtres adjacentes ne suivent plus la même distribution.

La Fig. 4.4 montre la probabilité de validité de l'hypothèse nulle pour deux tailles de fenêtres d'analyse différentes : 200 et 120 échantillons (cette dernière largeur correspondant à la durée du son direct). Nous vérifions graphiquement que les résultats sont une fonction de la taille de la fenêtre d'analyse.

De plus, il est difficile de conclure de manière certaine à la détection du temps de mélange. En effet, la Fig. 4.4 met en évidence la présence de plusieurs périodes de temps où deux fenêtres consécutives ont la même distribution, et ce, même dans le premiers instants de la réponse de la salle. Puis, ce phénomène n'est observé que par intermittence. Enfin comme auparavant, les résultats que nous présentons dépendent du seuil de tolérance que nous fixons (ici 5%).

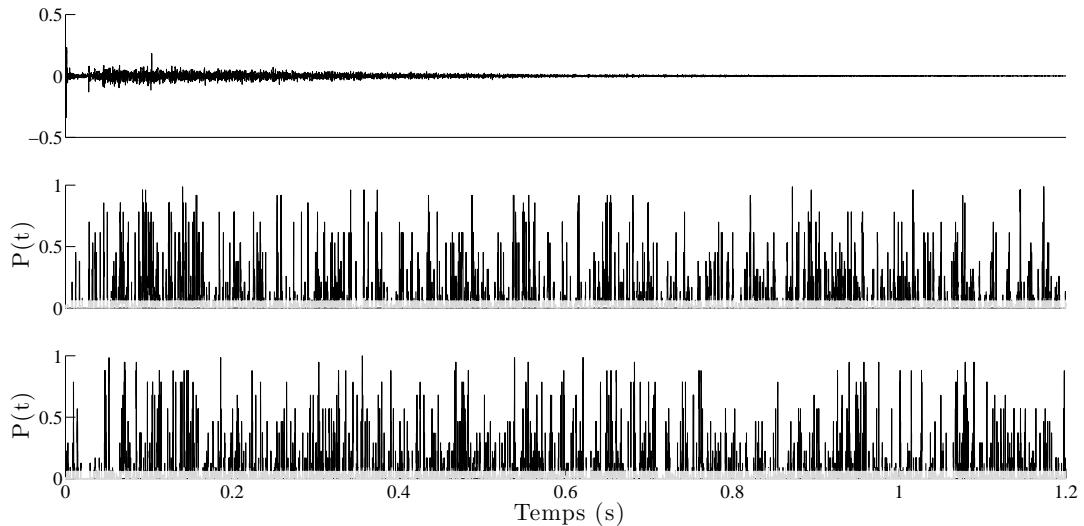


FIG. 4.4 – Haut : réponse impulsionnelle expérimentale (fréquence d'échantillonnage de  $F_s = 48kHz$ ) (signal entier et détails). Milieu : probabilité de validité de l'hypothèse nulle du test de Kolmogorov-Smirnov sur 2 fenêtres de taille 200 échantillons. Bas : probabilité de validité de l'hypothèse nulle du test de Kolmogorov-Smirnov sur 2 fenêtres de taille 120 échantillons. Note : en noir, les instants auxquels  $P(t) \geq 0.05$ . Remarquons que la validation de l'hypothèse nulle, contre celle non-nulle, dépend de la taille de la fenêtre d'analyse. La réponse impulsionnelle étudiée ici est la même que celle de la Fig. 4.2.

Notons que nous pourrions également utiliser un autre test, celui dit de Cramer-Von Mises. Ce test a les mêmes applications que le test de Kolmogorov. La différence entre ces deux tests

réside dans le fait que pour le test de Kolmogorov seul l'écart maximum entre la distribution empirique et la distribution d'ajustement entre en considération alors que l'indicateur d'écart du test de Cramer-Von Mises prend mieux en compte l'ensemble des données en ce sens que la somme des carrés des écarts intervient également. Le test de Kolmogorov est donc beaucoup plus sensible à l'existence de points abérents dans un échantillon que le test de Cramer-Von Mises<sup>6</sup>.

#### 4.4 Discussion et conclusion

Les quelques tests de Gaussianité que nous avons présentés ne constituent naturellement pas une liste exhaustive du genre. Nous avons cependant pu voir que si la distribution locale d'une réponse suit parfois celle d'une distribution Gaussienne, ce n'est vrai que par intermittence et cela dépend de la taille de la fenêtre d'analyse. Par conséquent, nous ne pouvons recommander l'emploi de ces tests afin d'estimer le temps de mélange. L'Annexe D présente les mêmes estimateurs appliqués à des réponses impulsionales modélisées par un bruit blanc Gaussien exponentiellement décroissant. Il apparaît que même en présence de distributions Gaussiennes ces estimateurs ne constituent pas des outils robustes (Fig. 4.5).

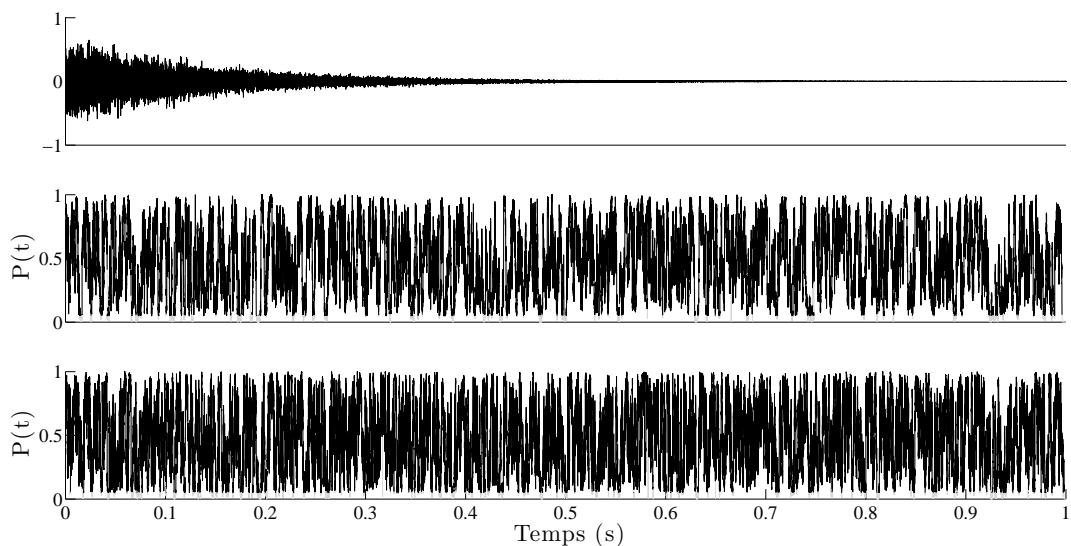


FIG. 4.5 – Haut : réponse impulsionale modélisée par un bruit blanc Gaussien (fréquence d'échantillonnage de  $F_s = 48kHz$ ). Milieu : probabilité de validité de l'hypothèse nulle du test de  $\chi^2$  sur des fenêtres de taille 200 échantillons. Bas : probabilité de validité de l'hypothèse nulle du test de  $\chi^2$  sur des fenêtres de taille 120 échantillons. Note : en noir, les instants auxquels  $P(t) \geq 0.05$ . Remarquons que la validation de l'hypothèse nulle, contre celle non-nulle, dépend de la taille de la fenêtre d'analyse.

Les quelques résultats présentés précédemment mettent en évidence un phénomène qui n'a pas retenu l'attention de Stewart *et al.* : des instants où le signal ne suit plus une distribution normale, alors que le champ diffus semble établi. Nous appelons ces laps de temps des *bouffées de cohérence*. Ces dernières correspondent à un retour du système dans un état proche de celui

<sup>6</sup>Les résultats auxquels nous aboutirions et les remarques que nous formulierions à la suite du test de Cramer-Von Mises seraient les mêmes que ceux nous avons déjà énoncés précédemment.

initial, et font particulièrement penser à un regain d'information du système alors qu'il est déjà entré dans un régime diffus. En d'autres termes, il existe des retours captés en un point de la salle dont l'entropie est plus faible que celle des instants précédents et futurs.

D'après la théorie de l'information de Shannon [68], nous savons que si tous les états d'un système sont équiprobables, alors l'entropie dudit système est maximum. De plus, le principe d'entropie maximum prédit que tant que l'état d'un système n'est pas désordonné au maximum, il continue d'évoluer jusqu'à ce qu'il ne contienne plus que les seules informations dont il dispose compte-tenu des contraintes auxquelles il est soumis.

On représente l'état macroscopique du système au moyen d'une loi de probabilité  $p$  sur l'espace des micro-états. Cela signifie qu'à chaque micro-état  $q \in \Omega$  ( $\Omega$  étant l'espace des possibles), on associe un nombre  $p(q) \geq 0$  tel que  $\sum_{q \in \Omega} p(q) = 1$ <sup>7</sup>. L'information contenue dans cet état est déterminée par son entropie  $s$  statistique (théorème de Shannon) :

$$s(p) = \sum_{p \in \Omega} -p(q) \times \ln(p(q)). \quad (4.14)$$

La Fig. 4.6 présente un exemple d'estimation de l'entropie sur une réponse expérimentale dont la décroissance de l'énergie est compensée. Les bouffées de cohérence apparaissent nettement. Là encore, la visualisation des bouffées de cohérence dépend de la largeur de la fenêtre d'observation du signal.

L'augmentation de l'entropie, observée dans la Fig. 4.6, met en évidence la perte d'information du système au cours du temps, c'est-à-dire son caractère dissipatif.

La question que nous pouvons nous poser -et qui pourrait être le sujet de futures recherches- est de savoir si le temps d'observation du système est suffisamment long pour vérifier que celui-ci a atteint son équilibre.

Enfin, il est utile de remarquer que l'estimation de l'entropie présentée dans la Fig. 4.6 dépend, comme les précédents tests, de la largeur de la fenêtre d'analyse (Annexe D). Une alternative à ce problème réside sans doute dans l'exploitation de l'estimation des retours par Matching Pursuit [33] (Chapitre 5), c'est-à-dire en estimant le contenu spectral de chacun d'entre eux relativement au son direct, ou état initial du système.

---

<sup>7</sup>Cette description est un peu plus subtile dans le cas des systèmes quantiques.

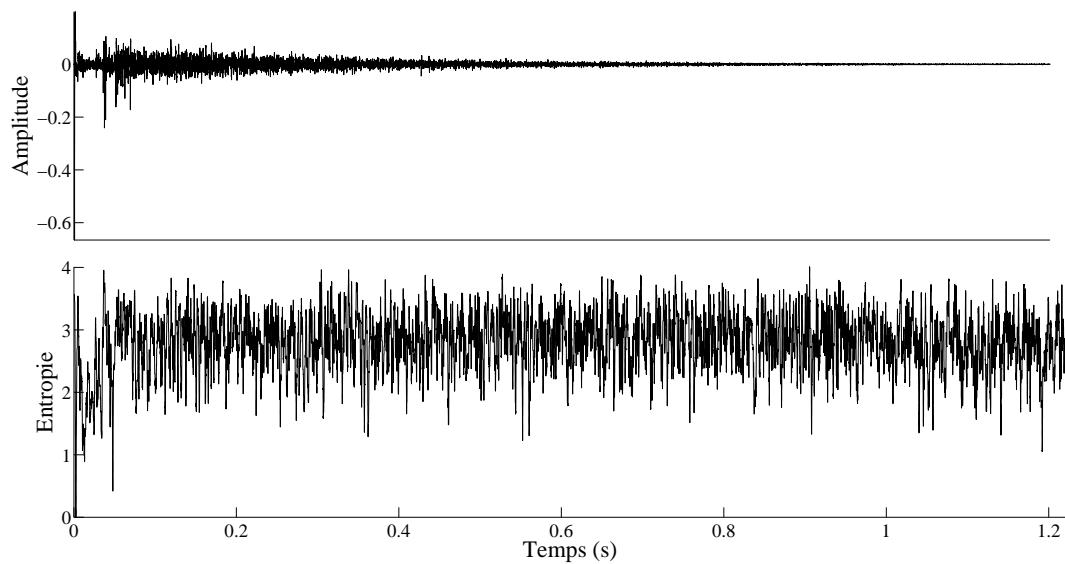


FIG. 4.6 – Haut : réponse impulsionnelle expérimentale (fréquence d'échantillonnage de  $F_s = 48kHz$ ) Bas : entropie calculée sur des fenêtres de 120 échantillons. Note : les instants où l'entropie du système est proche de celle initiale se distinguent nettement. La décroissance de l'énergie de la réponse est compensée pour le calcul de l'entropie, mais n'est pas montrée sur la figure du haut.

## Chapitre 5

# Estimation du temps de mélange à l'aide de Matching Pursuit

### 5.1 Préambule

Ce Chapitre propose d'estimer les retours de réponses impulsionales de salles à partir d'un algorithme (Matching Pursuit) utilisé pour la décomposition parcimonieuse de signaux. Plus précisément, il s'agit de trouver les instants où une forme d'onde (le son direct) se trouve reproduite dans le signal de la réponse de la salle. Naturellement, l'étude proposée ici est simple de par sa mise en oeuvre<sup>1</sup>. En effet, il serait plus judicieux d'utiliser plusieurs formes d'ondes qui seraient autant de versions filtrées du son direct. Ceci permettrait d'estimer le filtrage global en un point de la salle et d'avoir une connaissance approximative de la perte de mémoire du système, en d'autres termes, de mesurer son entropie.

Cela étant, ce Chapitre montre qu'à partir d'une réalisation simple de Matching Pursuit sur des réponses de salles, il est possible d'étudier la statistique des retours dans le temps et d'estimer le temps de mélange de la salle.

Ce Chapitre fait l'objet d'un article dans le journal *Acta Acustica United with Acustica* qui devrait paraître prochainement sous le titre : G. Defrance, L. Daudet and J-D. Polack, *Using Matching Pursuit for estimating mixing time within Room Impulse Responses*.

---

### 5.2 Using Matching Pursuit for estimating mixing time within Room Impulse Responses

#### 5.2.1 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

---

<sup>1</sup>Mais sans doute plus pertinente que de précédentes études estimant les retours en seuillant la réponse impulsionnelle [69, 70].

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

#### 5.2.2 Introduction

In 1975, Joyce [3] 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.5.1). 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 [8].

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* [6], 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 demonstrated in numerous books [42] [12]. But so far, no demonstration of the role of the mixing time has been given, since its estimation was lacking of consistency [66]. 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 5.2.3 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) [62] (Section 5.2.4). 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 5.2.4.1) 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 [62], it maintains an energy conservation which guarantees its convergence<sup>2</sup>. In practice the number of iterations

---

<sup>2</sup>En réalité, il est plus juste d'écrire : Although MP is a non-linear iterative decomposition process, it possesses a Parseval-like energy conservation property like a linear orthonormal transform [62]. Furthermore, MP guarantees that the approximation it builds converges onto the least-squares projection of the signal into the subspace spanned

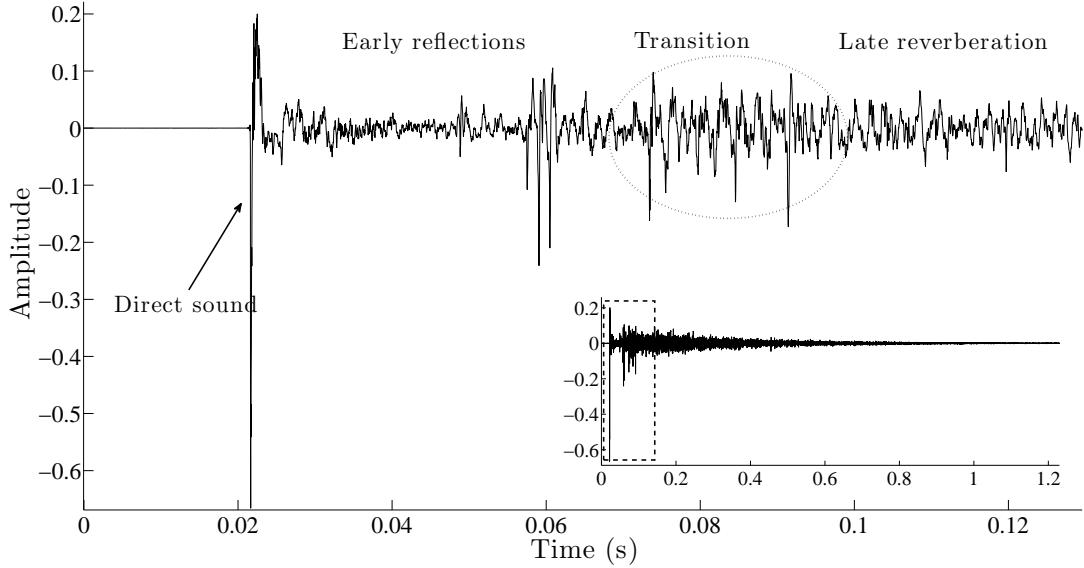


FIG. 5.1 – Early times of an experimental RIR (encapsulated graphic), measured in Salle Pleyel.

must be finite, leading to only an approximate decomposition of the signal (Section 5.2.4.2). In these experiments, the dictionary of waveforms (or atoms) 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 5.2.4.3 and explored in [47].

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 5.2.5 compares the temporal density of arrival to the theory presented in Section 5.2.3 and in [27], 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 5.2.6). 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 5.2.7. The issues of estimating mixing times using MP are discussed in Section 6.11.

### 5.2.3 Theoretical review of ergodic rooms

Since Weyl's path-breaking paper on eigenfrequency density in bounded rooms [13], 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 [14] 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 [14], 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,

---

by the dictionary. If the dictionary is at least complete, then the approximation is guaranteed to converge onto the actual signal as the number of iterations proceeds to infinity.

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 [?] 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, just as they are in the frequency domain. Indeed, Krylov's theory of mixing [8] 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 [71], teaches us that the density  $D_e(t)$  (Eq.(6.2)) 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).

$$D_e(t) = \frac{dN}{dt} = 4\pi c_0^3 \frac{t^2}{V} \quad (5.1)$$

where  $D_e(t)$  is in number of arrivals per second,  $N$  is the number of arrivals up to time  $t$ ,  $c_0$  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 distribution becomes constant (i.e.,  $D_e = 10$ ). The present paper concentrates therefore on the experimental determination of this transition time, which we call *mixing time* in accordance with [8], and investigate the robustness of different estimators of it.

In [9], 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  $N = 10$  reflections overlap within the characteristic time resolution of the auditory system, taken equal to  $\Delta t = 24ms$  [12]. Then Eq.(6.2) leads to :

$$t_m \approx 0.16 \cdot 10^{-3} \sqrt{\frac{V}{\Delta t}} \quad (5.2)$$

$$t_m \approx \sqrt{V} \quad (5.3)$$

where  $t_m$  is the mixing time, expressed in  $ms$  in Eq.(6.4),  $\Delta t$  is the characteristic time resolution of the auditory system, and  $V$  is the volume of the room in  $m^3$ .

In [9], this value was proposed as a reasonable approximation for the transition time between early reflections and late reverberation (Fig.5.1). The latter is then defined as an exponentially decaying noise, that is, as the superposition of pulses at a constant rate. It is shown in [27] and [9] that the exponentially decaying stochastic noise [72] is compatible with 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 [27] [73].

### 5.2.4 Matching Pursuit applied to RIR

A RIR can be seen as a linear combination of occurrence of the direct sound reproduced in time, and filtered by reflections on the surfaces of the hall. Figure 5.2 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.

#### 5.2.4.1 Theoretical reviews

Matching Pursuit can help understanding more deeply the architecture of a RIR, since this algorithm introduced by [62] provides information, which can be seen as maxima of correlation (Eq.(6.22)) [65] 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  $\gamma$  of a dictionary  $\phi$ , using inner products :

$$\forall \gamma \in \phi : CORR(x_m, \gamma) = |\langle x_m, \gamma \rangle| \quad (5.4)$$

The dictionary  $\phi$  is a set of atoms  $\gamma$ , 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 :

$$\tilde{\gamma}_m = argmax(CORR(x_m, \gamma))_{\gamma \in \phi} \quad (5.5)$$

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

$$x_{m+1} = x_m - \alpha_m \tilde{\gamma}_m \quad (5.6)$$

$$x_R^{(m)} = \sum_{k \leq m} \alpha_k \tilde{\gamma}_k \quad (5.7)$$

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,  $\gamma$  the atom (here, the direct sound),  $\phi$  the dictionary of atoms  $\gamma$ , 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<sup>3</sup>. 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 L2 of  $x$  over the norm L2 of the residual ( $R$ ) defined as :

$$SRR = 20 \log_{10} \left( \frac{\|x\|_2}{\|R\|_2} \right) \quad (5.8)$$

---

<sup>3</sup>Nous devons apporter ici une correction. Il est plus juste d'écrire : Given a dictionary that spans the space in which the signal exists, then MP guarantees that the error will be monotonic decreasing – i.e., it approaches 0 from above – which may take an infinite number of iterations.

#### 5.2.4.2 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 [41], 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 ( $RT$ ) measures the energy decay, and is probably the most widely used index. It is measured using the Schroeder integrated impulse response technique [43], 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 [74], is measured by the same method as  $RT$ , between  $0$  and  $10dB$ .
- Central Time, proposed by [12] and [75], expressed in seconds, is a measure of the centre of gravity of the impulse response energy :

$$T_C = \frac{\int_0^\infty t \cdot x^2(t) dt}{\int_0^\infty x^2(t) dt} \quad (5.9)$$

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 practical background. Figure 5.3 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 (Fig.5.2). This RIR presents the particularity that the direct sound (pistol shot) is immediately followed by the first reflection. This way, the identification of temporal boundaries of the direct sound, in order to constitute the dictionary  $\phi$  for that particular RIR, is facilitated.

According to [76], acceptable variations of acoustical indices are  $5\%$  and below. Thus, Figure 5.3 indicates that a convenient  $SRR$  would be  $20dB$ . Reasons of variations of acoustical indices can depend upon several factors [76] : 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 [55]. Figure 5.4 shows a RIR and the linear set of coefficients.

#### 5.2.4.3 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 sound sources in room acoustics measurements [47] [77] [48], 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. Moreover, the visual identification of these boundaries may vary from expert to expert [78] (Fig.5.5).

The direct sound is detected as follows. The impulse duration is assumed to be inferior to  $5ms$ , defining the atom  $\gamma$ . The time index  $t_0$  of the maximum of the atom is detected. A dictionary of atoms  $\phi$  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

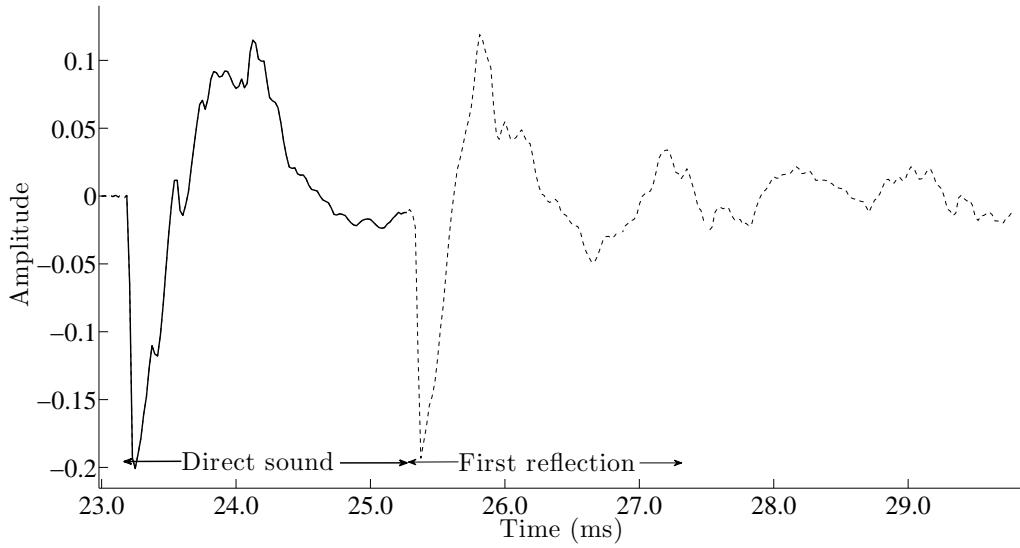


FIG. 5.2 – Experimental RIR for which determining the direct sound is obvious. Plain line : direct sound.

of [first index :last index], MP is ran. The temporal 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 [47].

The temporal boundaries of the direct sound depend on the geometry of the situation. In room acoustical measurements, the direct sound is often followed by the first reflection. Thus, the estimation of its temporal boundaries is approximative and leads to inaccurate estimation of arrivals in the signal of the RIR. However, this paper aims at using Matching Pursuit and the method used for constituting the dictionary as tools to estimate the mixing time. Therefore, investigating the impact of cutting the direct sound is outside the scope of the present paper. Nevertheless, this should be the topic of a future work.

## 5.2.5 Detection of arrivals

### 5.2.5.1 Validity of MP 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 [27].

In this model, arrivals are distributed in time according to a Poisson process, with a parameter which is time dependent, that is, the mean number of arrivals is given by Eq.(6.2). The cumulative number of arrivals (*CNA*) is a cubic function of time (Fig.5.6). Using input parameters, such as the reverberation time ( $RT_{30}$ ), the mean absorption ( $\bar{\alpha}$ ), and 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. This model has been validated on a set of 8 American concert halls for which a set of room acoustical indices were recovered [7].

Figure 5.6 compares the theoretical *CNA* of a single RIR synthesized by the model to the ones estimated by MP. One may notice the differences of estimation of arrivals when the dictionary is either constituted of a Dirac (the set of arrivals is not convolved by any atom), or

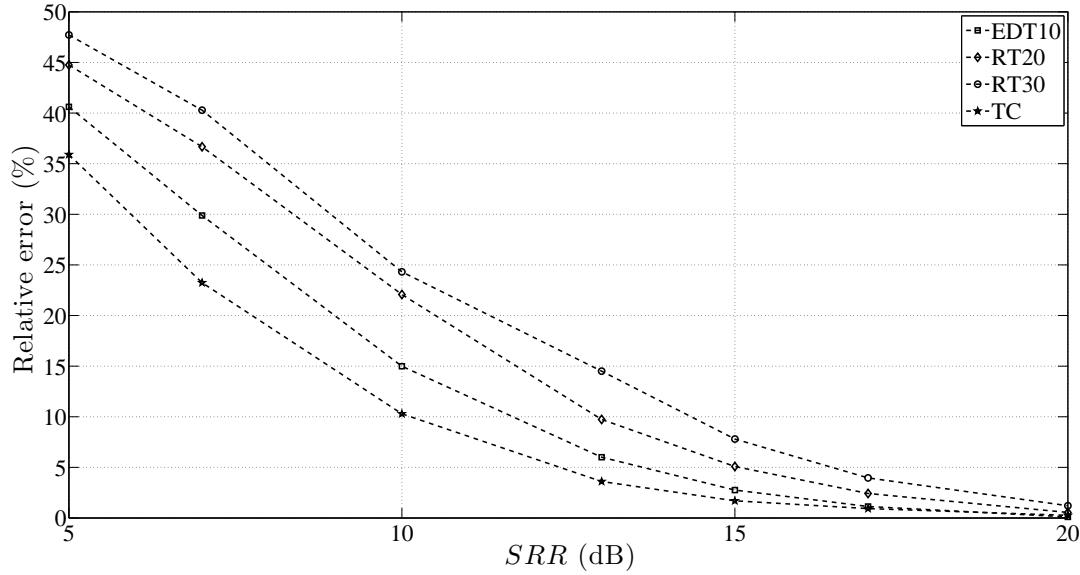


FIG. 5.3 – Signed variations in % of  $EDT_{10}$ ,  $RT_{20}$ ,  $RT_{30}$  and  $T_C$  versus  $SRR$  in dB.

of a pistol shot (the set of arrivals is convolved by this atom). These differences of estimation regarding the type of sound source are explained in the following.

#### 5.2.5.2 Towards a validation of the statistical theory

The presented results are derived from RIRs measured in Salle Pleyel carried out with pistol shots [52], for 21 different source-receiver positions. Therefore 21 experimental RIRs are under consideration. Based on section 6.10, the linear set of coefficients (Fig.5.4), derived from Matching Pursuit run on experimental RIRs, are assumed to represent the temporal distribution of arrivals.

**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.5.7. In other words, the  $CDF$  describes the time evolution of the probability to detect arrivals in the RIR.

Figure 5.7 underlines the decreasing of probability to detect arrivals at the end of the RIR. Mallat et al. [62] 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.5.4). 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.5.8), the mean reverberation decay of the room ( $RT \approx 2.0s$ ) is recovered in comparison to what has been calculated in [52]. This observation highlights that the  $CDF$  is linked to the energy of the signal.

**Compensating for the energy decay** As seen in section 5.2.5.2, 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 5.9, where it corresponds to the almost constant slope of the  $CDF$ . Note that the energy compensation is achieved by applying

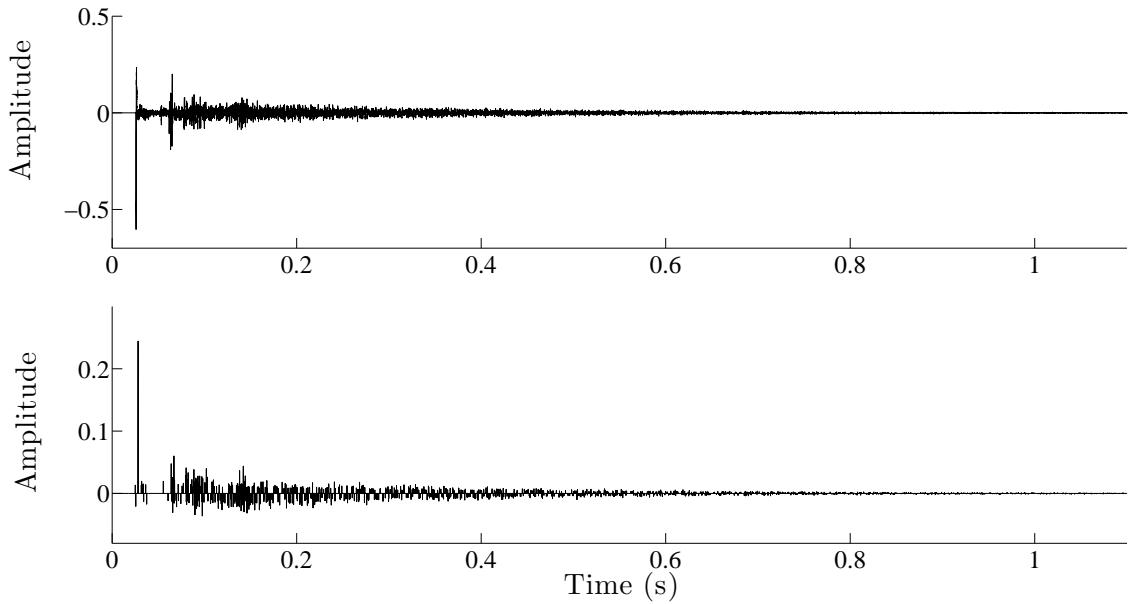


FIG. 5.4 – Matching Pursuit ran on an experimental RIR ( $SRR = 20dB$ ). -(top) : Experimental RIR -(bottom) : Linear set of coefficients, which correspond to the arrivals.

an inverse exponential, whose argument is proportional to the reverberation time and to the mean absorption [12].

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 [8]. 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.5.10), as it is explained in Section 5.2.6.

Matching Pursuit is ran on 21 RIRs measured in Salle Pleyel [52], compensating the energy decay. *CDF's* all show a breaking point (Fig.5.9) which divides the curve into two parts :

- $t < T_M$  : a few arrivals are detected after the direct sound. A cubic fit (Fig.5.11) of this part of the curve permits to verify that the number of arrivals is a function of  $t^3$ , as seen in Eq.(6.2). 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.5.12).

### 5.2.6 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 defined in the same way as the equivalent statistical bandwidth of the impulse, defined by [79], but replacing frequency spectrum with time signal. The equivalent statistical bandwidth is the bandwidth of a hypothetical rectangular filter which would pass a signal with the same

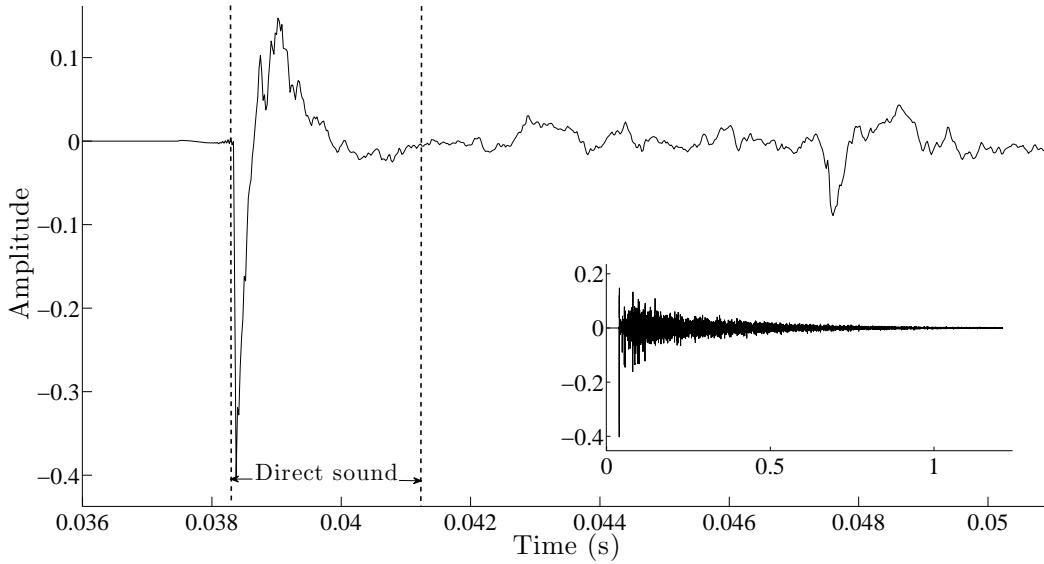


FIG. 5.5 – Detail of an experimental RIR for which determining the direct sound is not obvious. The direct sound estimated by MP is delimited by dashed lines.

mean-square statistical error as the actual filter when the input is white noise [79] [80].

The statistical 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 5.13-5.14 show statistical 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.5.13-5.14), where the RIR can be defined as behaving stochastically, rather than deterministically. In that sense, statistical times are related to Krylov's definition of mixing time (Section 5.2.5.2). 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.

#### 5.2.6.1 Dependence upon the value of 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 accurately by MP than without compensation (see Section 5.2.5.2). 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 5.2.4.2. Second, as studied in [81] 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 5.2.7. Therefore, one may expect a strong dependency of the accuracy of the estimation of the mixing time upon the  $SRR$ , since the higher the  $SRR$ , the larger the number of arrivals is.

Figure 5.15 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  $4 \leq SRR < 5dB$ . Furthermore, 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  $SRR \geq 10dB$ .

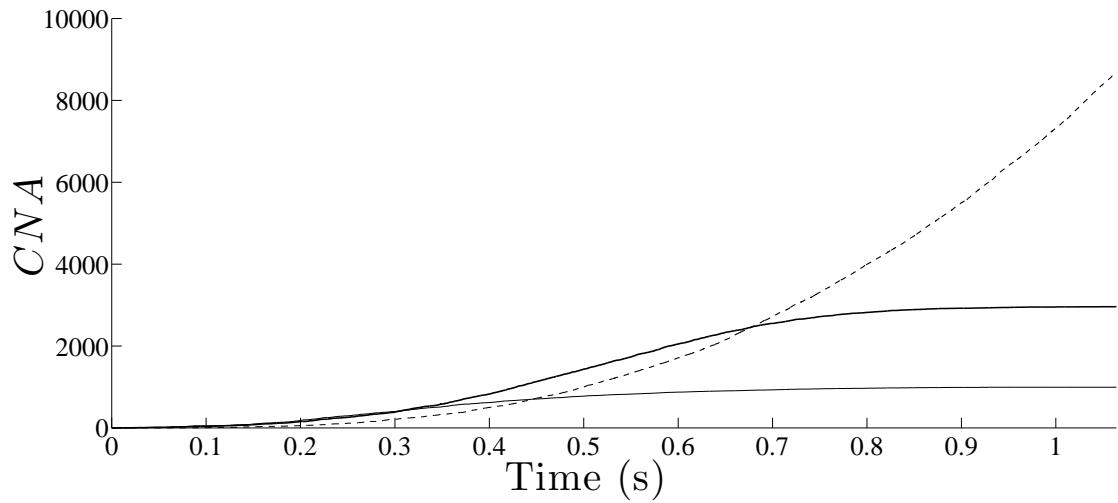


FIG. 5.6 – Cumulative numbers of the arrivals ( $CNA$ ) generated by the stochastical model of RIR (without compensation of energy decay). -dashed : model -plain : the atom is a pistol shot -bold : the atom is a Dirac.

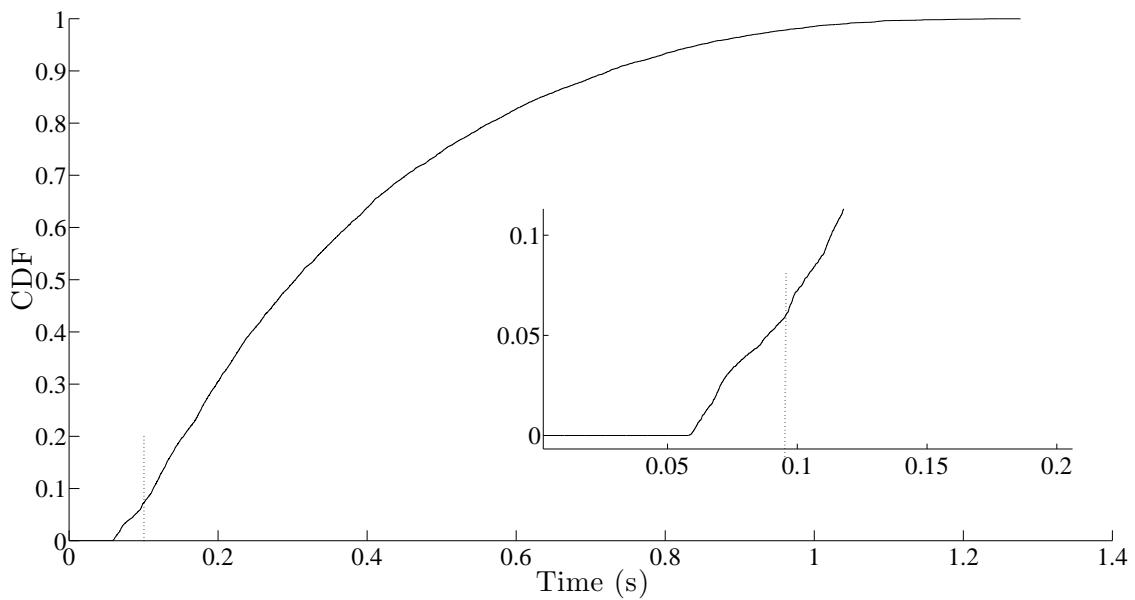


FIG. 5.7 –  $CDF$ , for an experimental RIR (without compensation of the energy decay). Dashed line : a kink in the distribution of arrivals is noticeable. The encapsuled graphic is a zoom on the early times of the  $CDF$ .

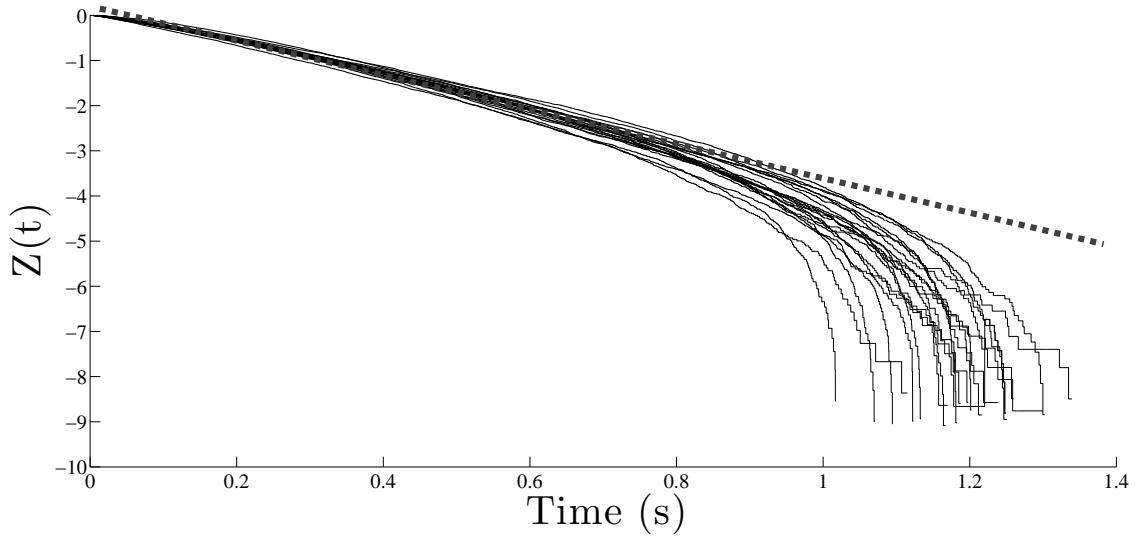


FIG. 5.8 – Plain : Logarithm of  $(1 - CDF)$  of the arrivals of 21 experimental RIRs (without compensation of the energy decay). Dash : mean logarithmic decay of the RIRs.

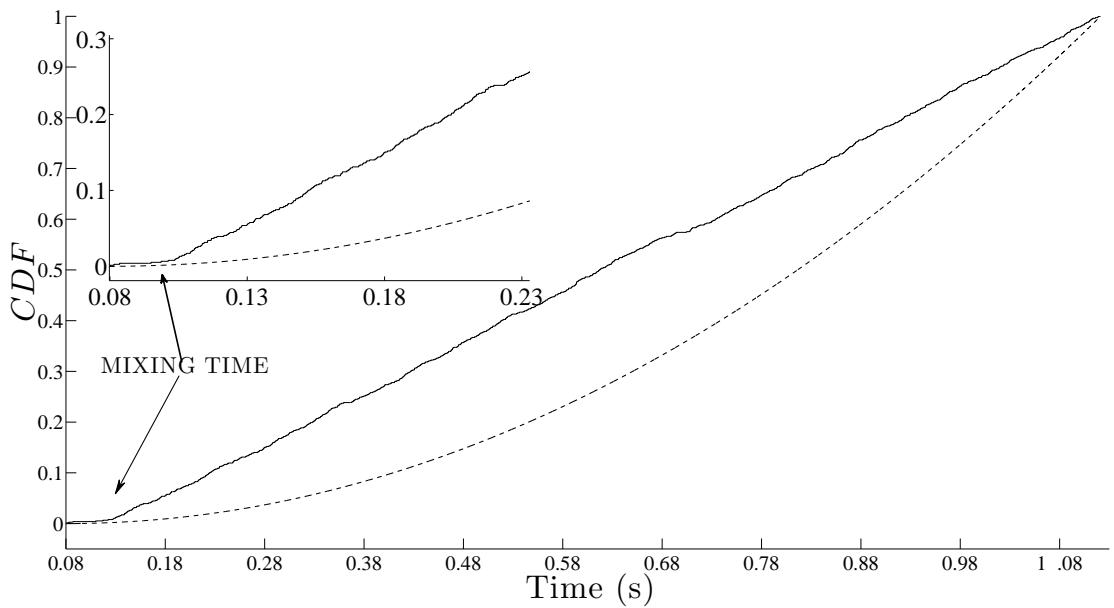


FIG. 5.9 –  $CDFs$ . Dash : theoretical  $CDF (\approx t^3)$ . Plain : Average of  $CDF$  for 21 RIRs (with compensation of the energy decay). The encapsulated graphic is a detail of the early part of the  $CDF$ .

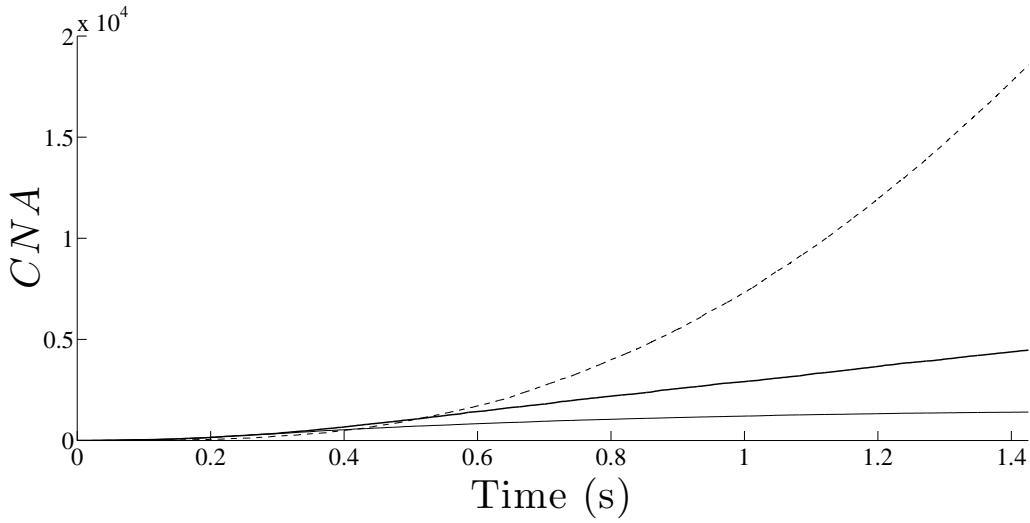


FIG. 5.10 – *CNAs* of a synthesized RIR. Dash : theoretical *CNA*. Plain : estimated *CNA* (the atom is a pistol shot) -bold : estimated *CNA* (the atom is a Dirac) (with compensation of the energy decay).

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 5.2.4.2, for the 21 RIRs. From these results (Fig.5.16), and according to [76], 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 6.10) 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.

### 5.2.6.2 Experimental RIRs

For each of the 21 experimental RIRs, the equivalent duration of the impulse ( $\tilde{d}$ ) is calculated, according to [79], since the direct sound is estimated by learning the dictionary of atoms (Section 5.2.4.3). Mixing times are estimated from the sets of arrivals (Fig.5.13), as described in Section 5.2.6.1. 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.5.17).

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.5.18), which is not predicted by the theory (Eq.6.4). The relationship is given by :

$$T_M = 0.0026.d + 0.026 \quad (5.10)$$

$$r = 0.82 \quad (5.11)$$

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

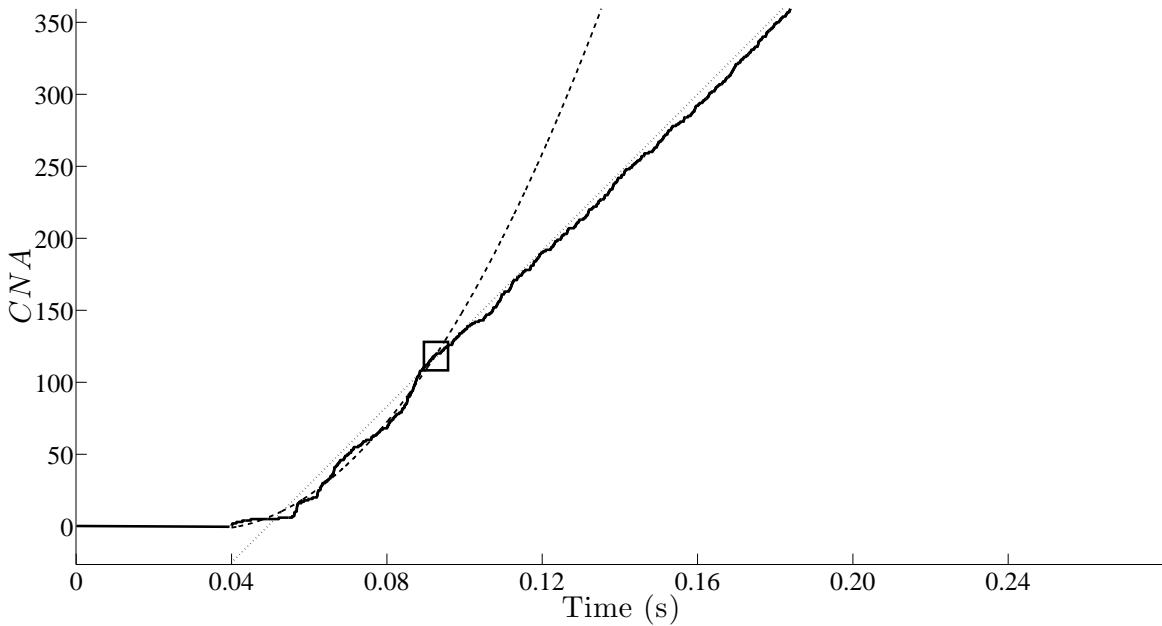


FIG. 5.11 – Detail of a cubic ( $t < T_M$ ) and a linear ( $t > T_M$ ) fits made on a  $CNA$ . Plain bold line :  $CNA$ . Dashed bold line : cubic fit. Dashed line : linear fit. Square : mixing time.

### 5.2.6.3 Synthesized RIRs

This section aims at estimating statistical times on the set of arrivals, obtained with MP, from RIRs synthesized with the model presented in Section 6.10. This is achieved by synthesizing a large number of times RIRs for the same input parameters. This is inspired by the Monte Carlo methods, often used when simulating physical and mathematical systems [82].

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 6.10). 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 5.14 shows the statistical time detected into a set of coefficients of MP. Figure 5.19 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 :

$$T_S = 0.0017.d + 0.057 \quad (5.12)$$

$$r = 0.97 \quad (5.13)$$

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 source/receiver distance is given by :

$$T_S = 0.0029.d \quad (5.14)$$

$$r = 0.98 \quad (5.15)$$

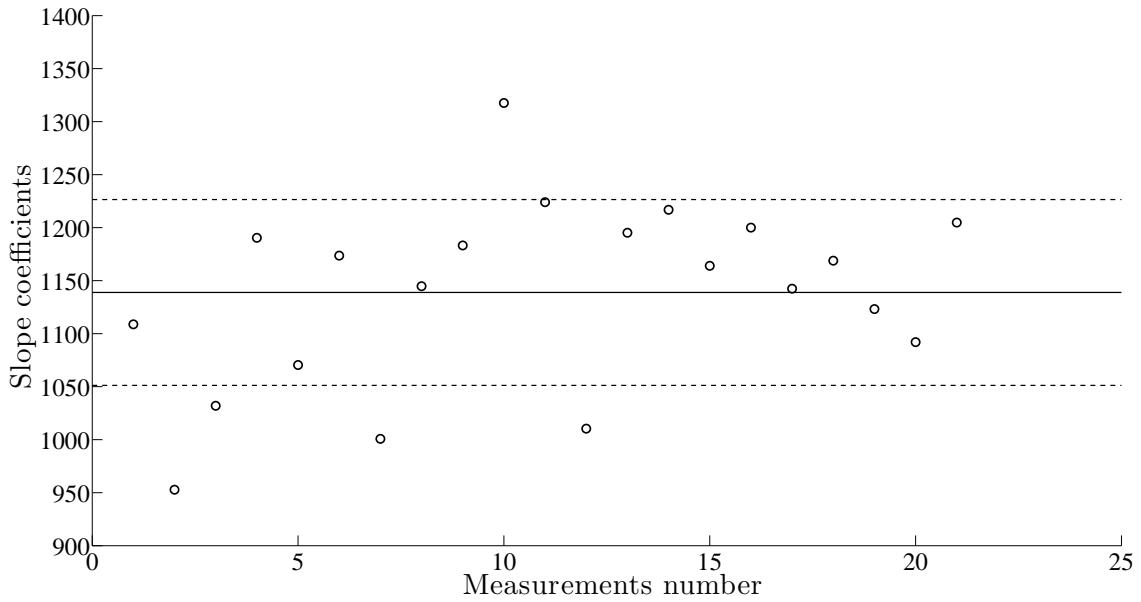


FIG. 5.12 – Circles : Slope coefficients of the linear fits made on the 21 CNAs ( $\text{mean} = 1160$  arrivals/second,  $\text{std} = 5\%$ ). Plain line : mean slope coefficients. Dashed lines : standard deviation.

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.5.14), then  $1/a = 1/0.0029 = 340 \text{ m.s}^{-1} \approx c_0$ . 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.

### 5.2.7 Differences between data from the experiment and the model

In Salle Pleyel, the volume is  $V = 19000 \text{ m}^3$  and the theoretical upper bound of the mixing time equals  $T_{M_{th}} = 137 \text{ ms}$ , according to Eq.(6.4.) Mean mixing times estimated from experimental RIRs ( $\bar{T}_M = 85 \text{ ms}$ ) on the one hand, and from synthesized RIRs ( $\tilde{T}_S = 101 \text{ ms}$ ) on the other hand, recover the theoretical assumption that states that  $T_{M_{th}} \approx \sqrt{V}(\text{ms})$  is an upper bound of mixing time (Section 5.2.3).

If the mean mixing times of experimental and synthesized RIRs are not strictly equal, one may notice that the relationships that link the mixing time to the distance are very similar (Fig.5.20-5.21). Actually, the diffusion phenomenon (which is related to mixing -see Section 5.2.3) is assumed to explain the difference between the model and the experiments (Fig.5.21).

On the one hand, as seen previously in Section 5.2.6, mixing time is a function of distance, as is the time delay between the direct sound and the first arrival of the RIR [42]. This time delay 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.

On the other hand, the number of arrivals estimated by MP is a function of the SRR (Section 5.2.6.1). As a consequence, MP estimates arrivals and also diffusion in experimental RIRs. Sturm et al. present, in [81], an artefact of MP : constructive and destructive interferences. When MP finds a correlation between an atom and the signal, it subtracts the contribution of this atom

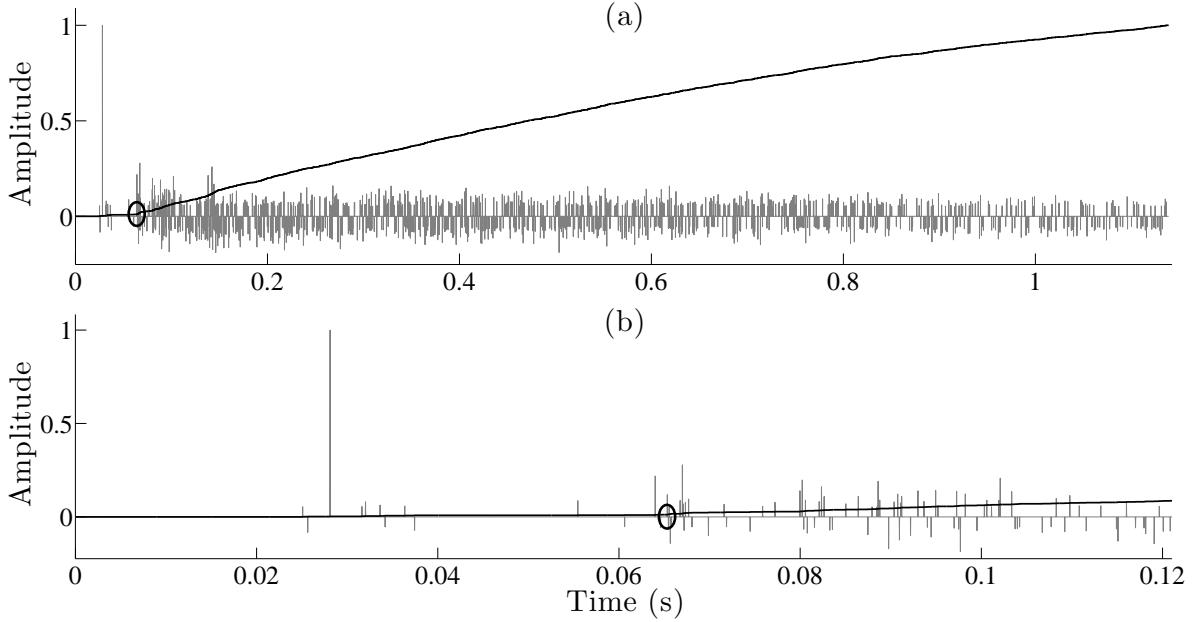


FIG. 5.13 – a) : Mixing time (circle) detected on a *CDF* (bold line) for an experimental RIR and the set of estimated arrivals. ( $SRR = 5dB$ ). Plot (b) details the beginning of plot (a).

from the signal. But in the case of experimental RIRs, the mother atom (the direct sound) can only accurately 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 subtracting 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 5.21.

### 5.2.8 Discussion

Studying the distribution of arrivals is not a new idea in room acoustics. As early as 1958, Schodder [35] 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 5.7.

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,

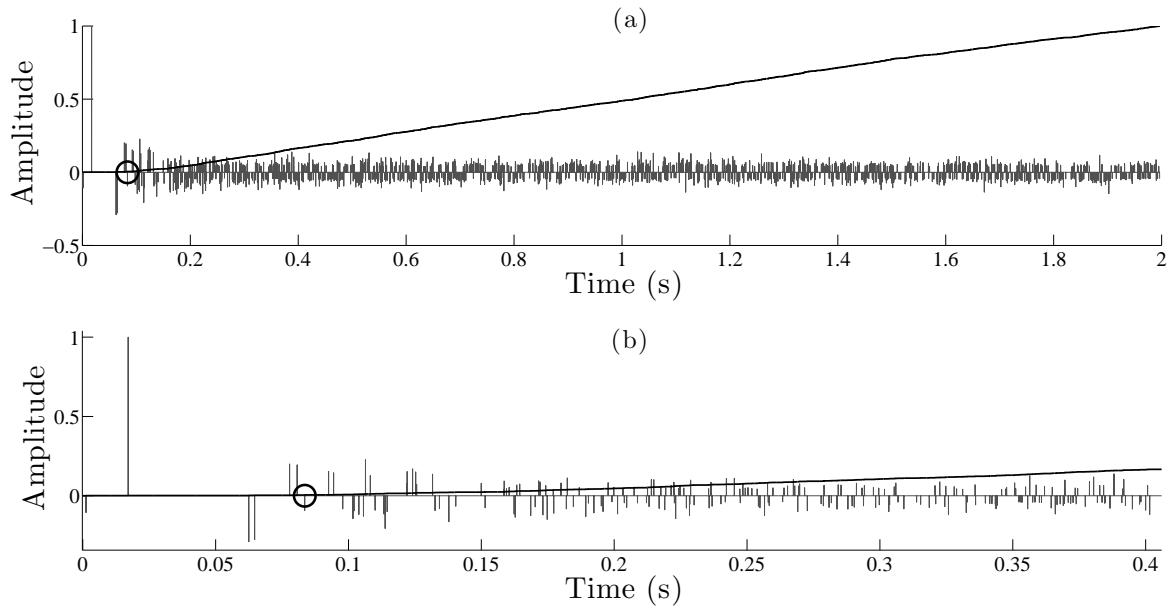


FIG. 5.14 – a) : Mixing time (circle) detected on a *CDF* (bold line) for a synthesized RIR convolved with a pistol shot and the set of estimated arrivals ( $SRR = 5dB$ ). Plot (b) details the beginning of plot (a).

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 in octave bands is thought to be an original manner to discriminate the phenomenon 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.5.10). 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 [7]. However, a future work would consist in adding diffusion to the model and in comparing the obtained mixing times to the experimental ones.

### 5.2.9 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 chosen 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

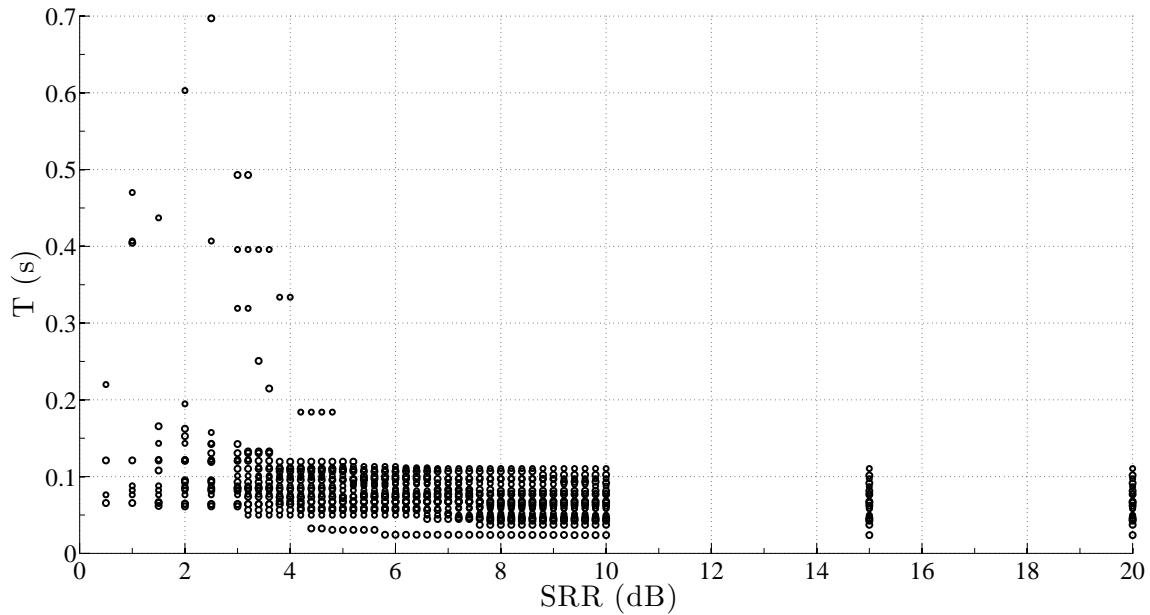


FIG. 5.15 – Mixing times estimated on experimental RIRs using different values of  $SRR$  ( $dB$ ) (with compensation of energy decay).

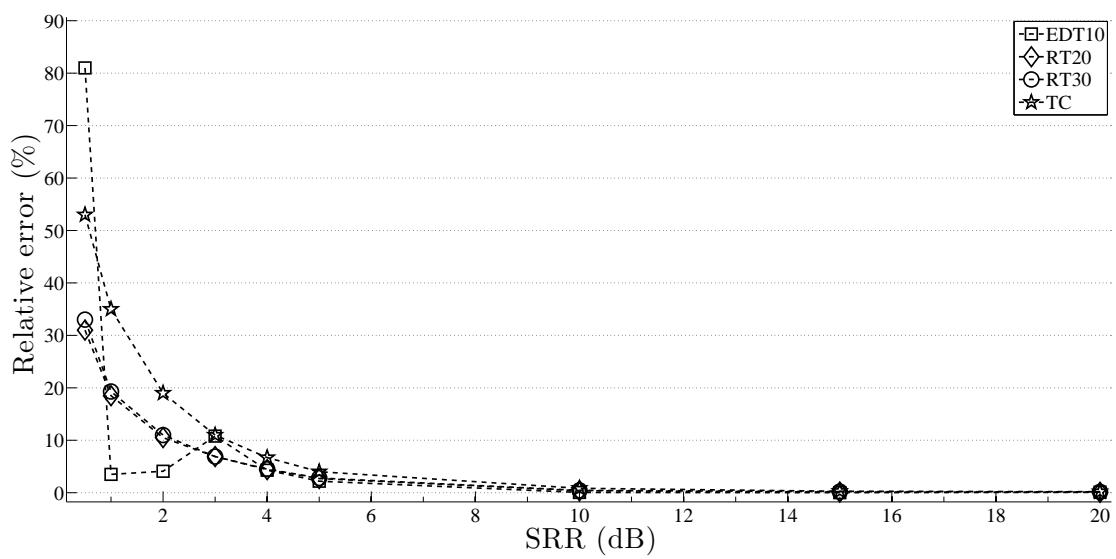


FIG. 5.16 – Variations in % of  $EDT_{10}$ ,  $RT_{20}$ ,  $RT_{30}$ ,  $T_C$  versus the  $SRR$  in  $dB$  (with compensation of the energy decay).

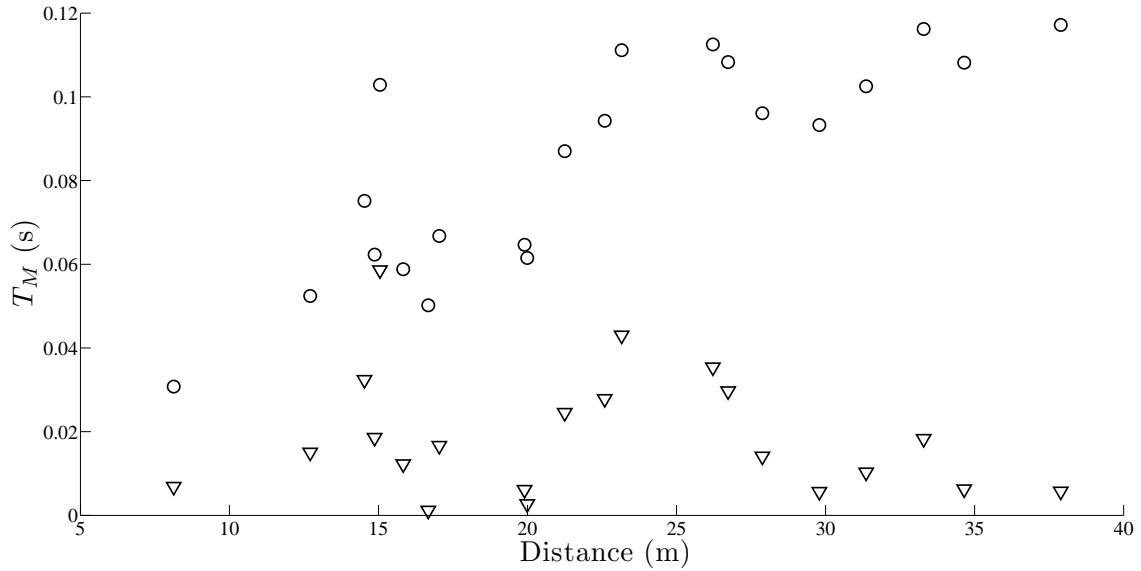


FIG. 5.17 – Estimated mixing times as a function of distance (without (triangles) and with (circles) time propagation between the source and the receiver).

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 [47]. A future work would consist in studying the impact of an inaccurate detection of the direct sound on the estimation of arrivals and of mixing time.

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 adequacy 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.

## 5.3 Développements ultérieurs

### 5.3.1 Taux d'accroissement du nombre de retours

Dans cette Section, nous tentons d'expliquer le lien entre la durée équivalente de l'impulsion émise par la source et la pente moyenne des CNAs sur la Salle Pleyel (Section 5.2.5.2). Nous

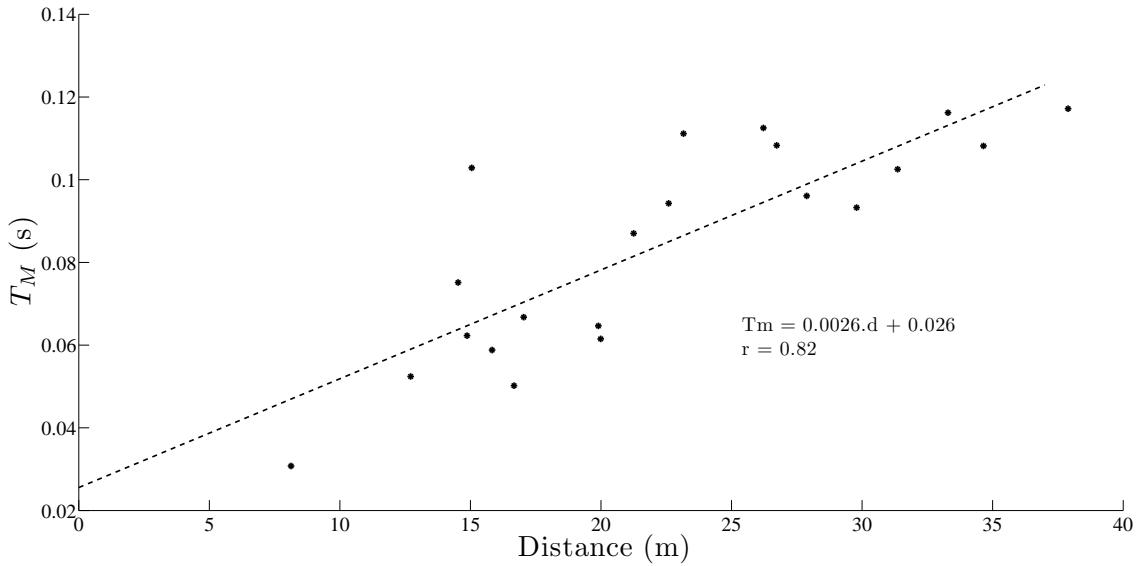


FIG. 5.18 – Estimated mixing times (dots) as a function of distance (with time propagation between the source and the receiver). Dashed line : linear regression.

montrons que ces deux grandeurs sont liées en calculant le nombre de maxima d'une réponse impulsionnelle.

Considérons un modèle de réponse impulsionnelle de salle donné par un signal Gaussien. Comme nous l'avons fait précédemment (Section 5.2.5.2), nous ne tenons pas compte de la décroissance de l'énergie ; nous considérons donc un signal stationnaire. Nous nous intéressons aux maxima du signal car Matching Pursuit détecte des maxima de corrélations aux instants où l'énergie locale du signal est la plus élevée (Section 5.2.5.2). Afin de d'estimer le nombre de retours que MP détecte dans le champ diffus, nous cherchons le nombre de maxima d'un signal Gaussien. Pour cela, nous nous référons aux travaux de Rice [83] (pp.77-81) qui sont d'une grande aide dans ce cas. En effet, les détails des calculs et les résultats sont présentés clairement. Par conséquent, nous ne présentons que les résultats.

A partir de la fonction de corrélation d'un processus aléatoire normal et de la densité de probabilité d'observer un maximum dans l'intervalle de temps  $t; t + dt$ , Rice aboutit à la formule donnant le nombre moyen de maxima par seconde entre une fréquence  $f_a$  et  $f_b$  :

$$N = \left[ \frac{3 \times (f_b^5 - f_a^5)}{5 \times (f_b^3 - f_a^3)} \right]^{1/2}. \quad (5.16)$$

Le microphone utilisé pour les mesures dans la Salle Pleyel [52] possède une bande-passante allant de  $20Hz$  à  $10kHz$ . Le nombre de maxima est alors égale à 7745 pour une durée d'observation de 1 seconde. De plus, la durée équivalente de l'impulsion émise par la source est d'environ  $0.6ms$ , soit la durée d'observation de 5 maxima. En d'autres termes, à partir du moment où le système est diffus, les retours se superposent et se recouvrent (Section 5.2.6) et Matching Pursuit détecte un retour au lieu de 5 dans le champ diffus. Au total, MP estime  $7750/5 = 1150$  retours/seconde dans le champ diffus ce qui correspond environ à la pente moyenne des CNAs obtenus dans la Section 5.2.5.2. Autrement dit, le taux d'accroissement du nombre de retours dans le champ diffus est fonction de la largeur de bande équivalente du son direct, donc de l'atome-mère du dictionnaire de MP.

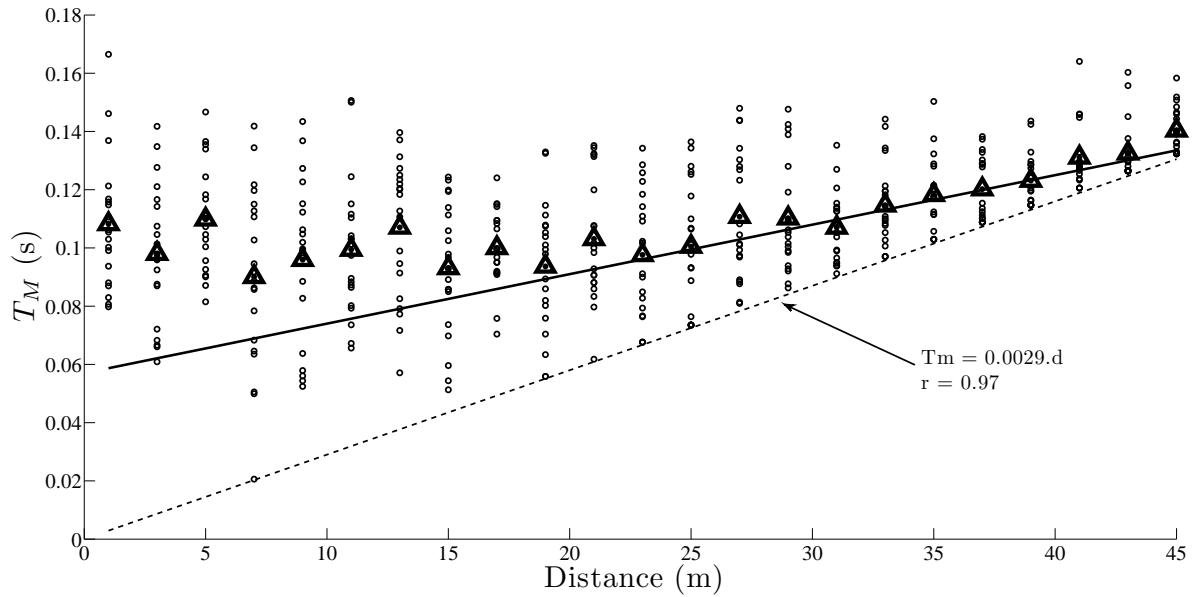


FIG. 5.19 – Mixing times (dots) detected for 20 RIRs synthesized at different distances (the time of propagation is taken into account) ( $SRR = 5dB$ ). Triangles : mean times. Dashed line : linear fit on the minima mixing times (for  $d \geq 19m$ ). Plain line : linear fit on the mean mixing times (for  $d \geq 25m$ ).

Nous entrevoyons ici la possibilité de définir autrement le critère d'arrêt de la poursuite. Connaissant la bande passante du signal et la durée équivalente de l'atome-mère, nous pouvons estimer le nombre de retours moyens à détecter dans le champ diffus grâce aux travaux de Rice. Tant que ce nombre n'est pas atteint, la poursuite doit continuer.

### 5.3.2 Améliorations du modèle - Impacts

#### 5.3.2.1 Ajout de la diffusion au modèle stochastique

Ce qui précède montre une certaine adéquation entre les résultats obtenus depuis les mesures de réponses de salles et le modèle stochastique. Cependant, cette adéquation est toute relative, et réside principalement dans l'observation du même comportement du modèle et de la mesure : le temps de mélange est une fonction de la distance source/récepteur. De plus, si le modèle (les temps de mélange moyens plus exactement) se superpose aux valeurs expérimentales au-delà d'une certaine distance source/récepteur, il n'en va pas de même pour les faibles distances (Fig. 5.20). La Fig. 5.22 superpose les résultats obtenus précédemment (modèle et mesure). Les raisons de cette différence sont développées précédemment, et nous amènent à conclure quant à la nécessité d'ajouter le phénomène de diffusion au modèle (Sections 5.2.7-6.11).

La Section 2.5 introduit en détail le modèle stochastique dont nous nous sommes servi dans la précédente étude. La synthèse de la diffusion, basée sur l'équation de la diffusion, est également présentée et est ajoutée au modèle. Si nous reproduisons l'expérience décrite dans la Section 5.2.6.3 et que nous traçons les temps de mélange estimés en fonction de la distance source/récepteur (Fig. 5.23), nous observons que les variations de temps de mélange sont plus faibles. De plus, les temps de mélange moyens sont une fonction linéaire croissante de la distance pour  $d \geq 20m$  (sans diffusion, cette relation était vérifiée pour  $d \geq 25m$ ). Notons que les relations qui lient le temps de mélange à la distance pour l'expérience et le modèle sont maintenant

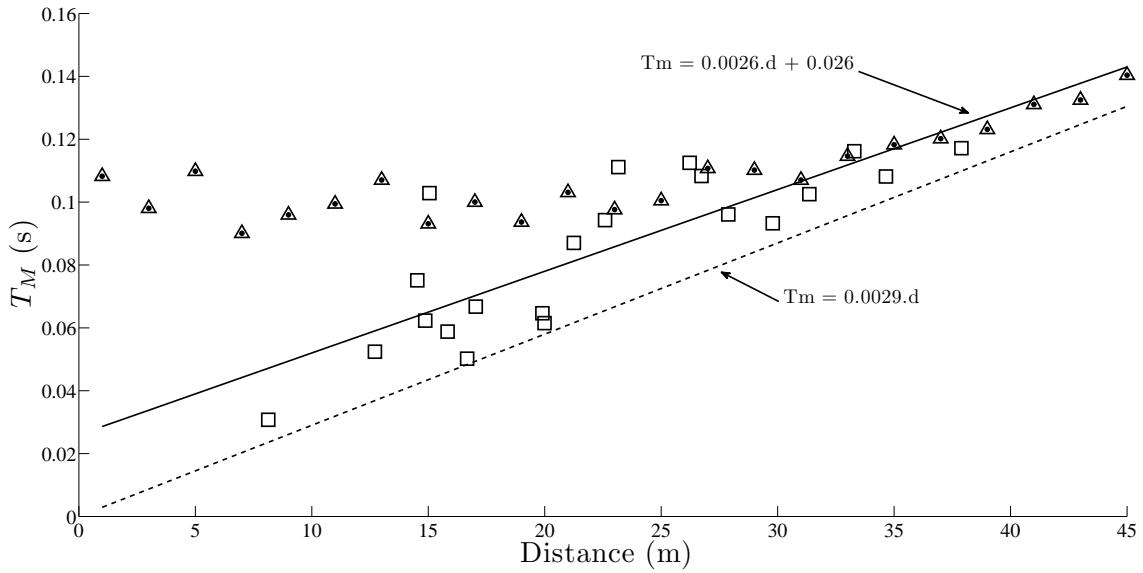


FIG. 5.20 – Mean mixing times of the model (triangles) and of the experiments (squares). Dashed line : linear regression made on the minima mixing times of the model. Plain line : linear regression of experimental RIRs.

Configuration	Relation temps de mélange/distance
Modèle sans diffusion	$T_m = 0.0017 \times d + 0.06, \forall d \geq 25m$
Modèle avec diffusion	$T_m = 0.0027 \times d + 0.01, \forall d \geq 20m$
Salle Pleyel	$T_m = 0.0026 \times d + 0.026, \forall d$

TAB. 5.1 – Relation entre temps de mélange et distance lorsque la diffusion est synthétisée par le modèle

semblables, mais restent légèrement décalées (Table 5.1, Fig. 5.23). Avec l'ajout de la diffusion, les résultats expérimentaux et du modèle se confondent aux faibles distances. Ce résultat était attendu. En effet, la diffusion ayant lieu après la première réflexion -et donc le premier retour- le modèle voit son nombre d'événements augmenté après le premier retour comparativement au modèle sans diffusion. Dans la Section 5.2.7, nous avons montré que les temps de mélange sont reliés à l'Inter Time Delay Gap [42] (ITDG). Nous retrouvons ici le même raisonnement expliquant le lien entre le temps de mélange, les faibles distances source/récepteur et la diffusion. Notons enfin, que lorsque la source et le point de réception sont confondus, le mélange est atteint lorsque les rayons ont parcouru environ un demi libre parcours moyen, soit environ 6m (le modèle calcule un libre parcours moyen de 13m). Ce constat nous autorise à penser que même si la synthèse de la diffusion permet de se rapprocher des données expérimentales, il existe encore une différence qui trouverait peut-être son origine dans le filtrage de la salle.

### 5.3.2.2 Ajout du filtrage de la salle au modèle stochastique

Dans le Chapitre 2.5, nous avons présenté trois méthodes aléatoires de filtrage de la salle, basées sur le nombre de réflexions subi par chaque rayon sonore. Rappelons que ce nombre est calculé de manière probabiliste par le modèle. Nous nous intéressons ici à ajouter au modèle sans diffusion successivement les trois types de filtrages et à reproduire l'expérience faite dans

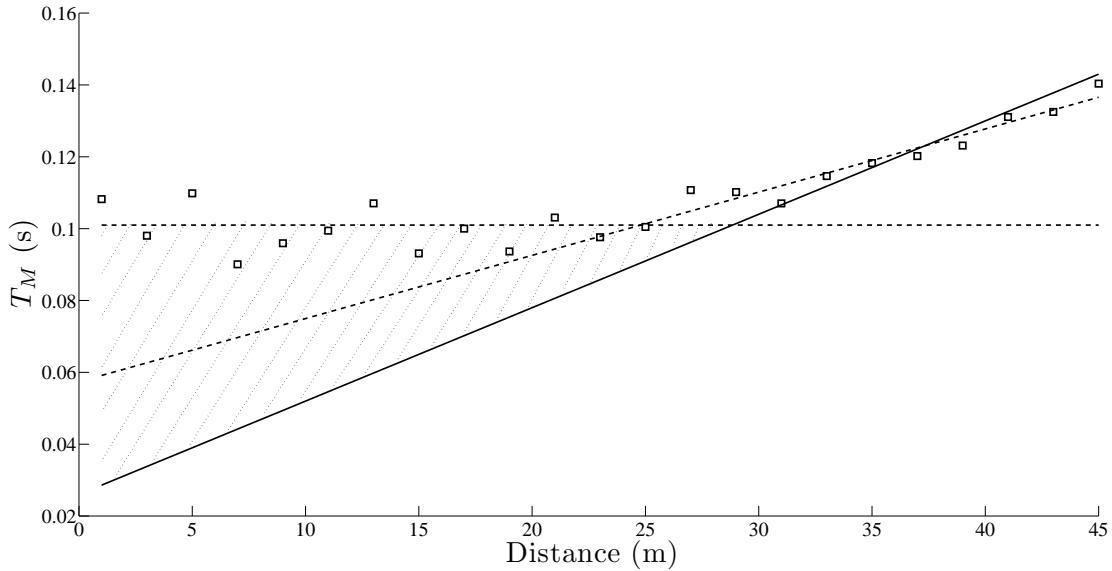


FIG. 5.21 – Mean mixing times from the model (squares). Dashed lines : linear regressions made on the model (for  $d \geq 25m$ ). Plain line : linear regression of experimental RIRs. Hatched part : missing contribution of diffusion and room filtering in the model.

la Section 5.2.6.3. Le dictionnaire d'atomes de MP est constitué du son direct, plus précisément d'une seule et même impulsion de revolver. Par ailleurs, l'emploi d'un dictionnaire plus complet d'atomes qui seraient des versions filtrées d'une même impulsion source permettrait sans aucun doute une meilleure approximation du signal. Ceci est discuté dans la Section 5.3.5 (voir aussi Fig. 5.39).

Les trois types de filtres utilisés sont :

- filtres en bandes d'octave ;
- filtres de trois matériaux ;
- filtre de Variations Spectrales Aléatoires (VSA).

Les Fig. 5.24-5.26 présentent les temps de mélange estimés sur le modèle, et permettent la comparaison avec les valeurs expérimentales. En premier lieu, nous constatons que les variations des temps de mélange sont amoindries, comme dans le cas de l'ajout de la diffusion (Section 5.3.2.1). Le premier retour étant une version filtrée du son direct, MP crée des interférences constructives et destructives autour de ce retour. Par conséquent, le temps statistique a de fortes chances d'être détecté aux environs de ce retour. Mais ceci n'est pas toujours vérifié, comme l'indiquent les Fig. 5.24-5.26 où l'on voit des temps de mélange détectés à des temps supérieurs à l'arrivée du premier retour ( $T_M > 0.14s$ ).

Lorsque le modèle synthétise le filtrage de la salle, les temps de mélange moyens suivent tous une relation directement proportionnelle au temps de propagation entre la position de la source et du récepteur (pour les raisons que nous avons expliquées précédemment -Section 5.2.7). De plus, les temps de mélange expérimentaux coïncident avec les nuages de points du modèle, pour chaque type de filtrage. D'autre part, la synthèse du filtrage d'une salle aléatoire permet -à l'exception du filtrage en bandes d'octave- d'obtenir une relation temps de mélange et distance source/récepteur semblables à celles mesurées expérimentalement (Table 5.2). Les différences entre les relations obtenues avec le modèle et celle expérimentales sont difficilement qualifiables

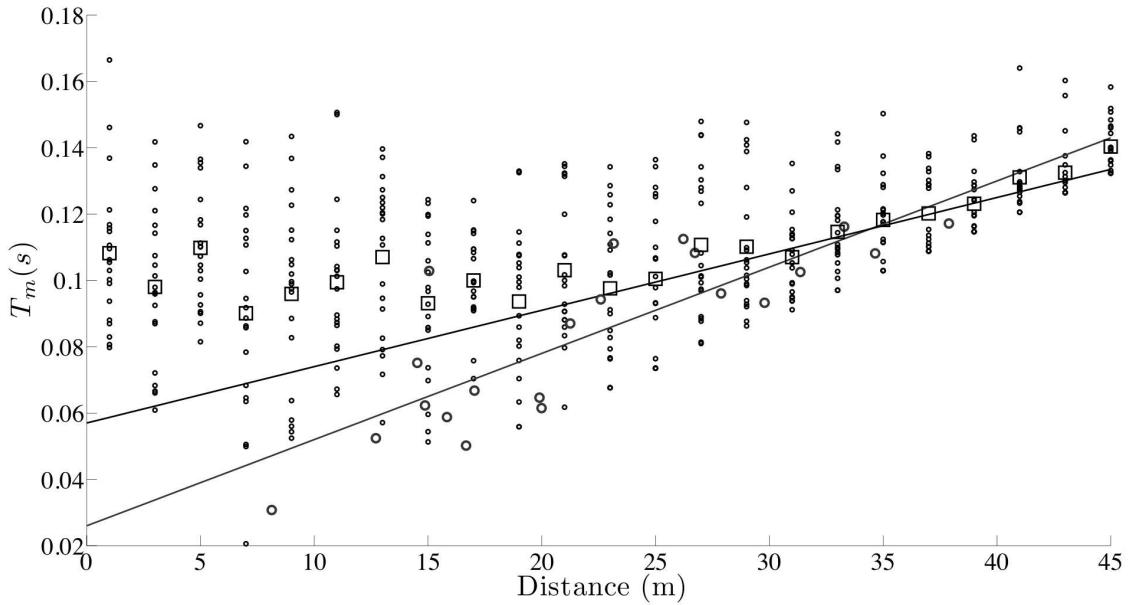


FIG. 5.22 – Superposition des résultats expérimentaux et numériques (et régressions linéaires). La diffusion n'est pas synthétisée par le modèle. Points : temps de mélange du modèle. Carrés : temps de mélange moyens du modèle. Cercles : temps de mélange expérimentaux ( $SRR = 5dB$ ).

Configuration	Relation temps de mélange/distance
Salle Pleyel	$Tm = 0.0026 \times d + 0.026, \forall d$
Filtrage en bandes d'octave	$Tm = 0.0031 \times d + 0.003, \forall d$
Filtrage matériaux	$Tm = 0.0022 \times d + 0.031, \forall d$
Filtrage VSA	$Tm = 0.0023 \times d + 0.027, \forall d$

TAB. 5.2 – Relation temps de mélange et distance pour trois types de filtrages aléatoires différents.

pour le moment<sup>4</sup>. Cependant, nous pouvons remarquer que le filtrage en bandes d'octave crée un modèle de salle immédiatement diffus, puisque le mélange est réalisé lorsque le son direct atteint le récepteur. Les données issues de l'expérimentation semblent indiquer qu'il existe un temps de relaxation du système, et que le mélange n'est donc pas immédiat. C'est ce que nous observons avec les deux autres types de filtrage : matériaux et VSA. Enfin, lorsque ces deux derniers types de filtres sont employés le mélange a lieu après environ 30ms (soit une distance de 10m) lorsque la source et le récepteur sont confondus. Alors que dans le cas du filtrage par bandes d'octave, ce temps est divisé par 10 (soit une distance de 1m). Rappelons que le libre parcours moyen estimé par le modèle est de 13m, ce qui tendrait à indiquer que seuls les filtrages type matériaux et VSA sont des modèles plausibles de filtrage global d'une salle.

Le tableau (Table 5.2) regroupe les différentes relations liant le temps de mélange à la distance source/récepteur.

### 5.3.2.3 Ajout de la diffusion et du filtrage de la salle au modèle stochastique

Lorsque le modèle synthétise tant la diffusion que le filtrage de la salle, la relation reliant le temps de mélange à la distance source/récepteur est proche de celle mesurée expérimentalement

<sup>4</sup>C'est-à-dire, en attendant d'autres estimations sur d'autres salles, avec un dictionnaire plus complet.

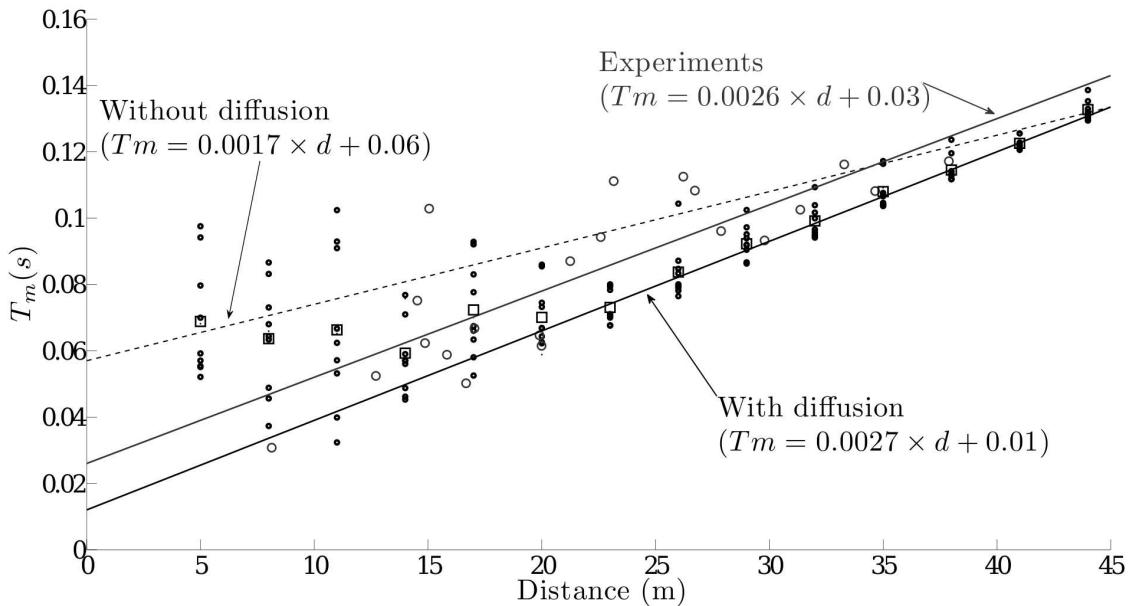


FIG. 5.23 – Superposition des résultats expérimentaux et numériques (et régressions linéaires). La diffusion est synthétisée par le modèle. Points : temps de mélange du modèle. Carrés : temps de mélange moyens du modèle. Cercles : temps de mélange expérimentaux ( $SRR = 5dB$ ).

(Fig. 5.27-5.29). Les raisons de cette similitude sont les mêmes que celles expliquées précédemment (interférences constructives et destructives autour du premier retour). Remarquons également, que à l'inverse de la synthèse du filtrage sans diffusion (Section 5.3.2.2), les trois types de filtres donnent sensiblement les mêmes relations (Table 5.3). De plus, les temps de mélange expérimentaux coïncident avec les nuages de points du modèle, pour chaque type de filtrage. Enfin, le temps au bout duquel le mélange est atteint pour une position identique de la source et du récepteur est le même quelque soit le type de filtrage, soit environ 10m. Cette valeur est en relativ accord avec celle calculée par le modèle (13m).

En conclusion, bien que les ajouts de la synthèse de la diffusion et d'un filtrage aléatoire d'une salle aient des impacts "positifs" sur la statistique des retours (nous retrouvons les données expérimentales de la Salle Pleyel), il semble nécessaire de rappeler que :

- cette étude ne concerne qu'une seule salle ;
- les mesures sont réalisées à l'aide d'impulsions de revolver dont la reproductibilité n'est pas encore établie de manière certaine ;
- le dictionnaire d'atomes servant à Matching Pursuit n'est composé que du son direct de chaque mesure, et non de plusieurs versions filtrées de celui-ci et de durées variables ;
- les temps de mélange calculés à partir du modèle et confrontés aux valeurs expérimentales sont obtenus à partir d'une seule et même impulsion de revolver ;
- il manque une étude expérimentale pour confirmer la valeur du temps de mélange lorsque la source et le récepteur sont confondus.

### 5.3.3 Les retours dans l'espace

Les Sections précédentes montrent comment il est possible d'estimer le temps d'occurrence des retours dans une réponse impulsionale à partir de la connaissance du son direct. Nous nous sommes donc concentrés sur la statistique des retours et sur l'estimation du temps de mélange. Il

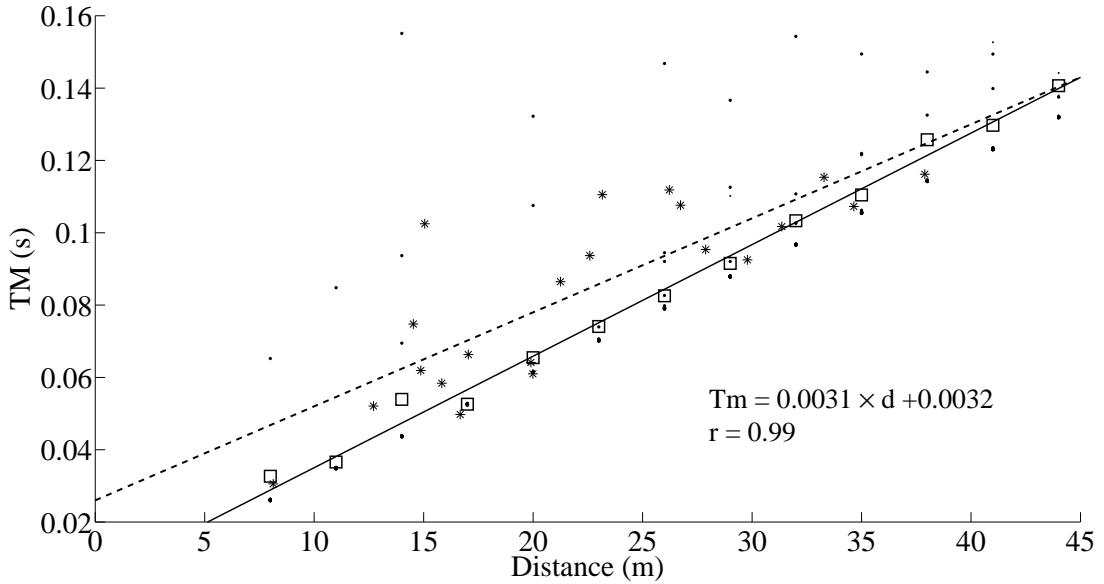


FIG. 5.24 – Résultats expérimentaux et numériques (et régressions linéaires). Le filtrage (en bandes d'octaves) est synthétisé par le modèle. Points : temps de mélange du modèle. Carrés : temps de mélange moyens du modèle. Etoiles : temps de mélange expérimentaux. Ligne pleine : régression linéaire sur les temps de mélange moyens du modèle. Ligne pointillée : régression linéaire sur les temps de mélange expérimentaux ( $T_m = 0.0026 \times d + 0.026$ ) ( $SRR = 5dB$ ).

Filtrage	Sans Diffusion	Avec Diffusion
Sans	$T_m = 0.0017 \times d + 0.06$	$T_m = 0.0027 \times d + 0.01$
Bandes d'octave	$T_m = 0.0031 \times d + 0.003$	$T_m = 0.0023 \times d + 0.028$
Matériaux	$T_m = 0.0022 \times d + 0.031$	$T_m = 0.0023 \times d + 0.029$
Variations Spectrales Aléatoires	$T_m = 0.0023 \times d + 0.027$	$T_m = 0.0023 \times d + 0.028$

TAB. 5.3 – Relations temps de mélange et distance pour trois types de filtrages aléatoires différents en présence -ou non- de diffusion.

est également possible de prêter attention à l'évolution du nombre de retours au cours du temps dans l'espace. Ainsi, nous pouvons distinguer une évolution des retours à certaines positions d'une salle, mais aussi selon certaines directions de l'espace<sup>5</sup>.

Dans la suite, nous présentons quelques résultats obtenus à partir des mesures dans la Salle Pleyel, et tentons d'identifier quelques pistes d'investigation futures.

### 5.3.3.1 Enregistrements omnidirectionnels

La Fig. 5.12 montre que l'évolution du nombre de retours au cours du temps est à peu près constante dans toute la salle (en moyenne, 1160 retours/sec.,  $\sigma = 5\%$ ). La Fig. 5.30 présente les mêmes données que la Fig. 5.12 mais regroupées selon la position de la source et des récepteurs dans la salle. Il est également intéressant de constater que le nombre de retours captés diffère selon la position du récepteur dans la salle (Fig. 5.30). Si cette observation ne dépendait ni de

<sup>5</sup>Cette section ne présente pas une étude complète, et n'en a pas la prétention. En effet, elle ne concerne qu'une seule salle de spectacle : la Salle Pleyel à Paris. Par conséquent, les résultats présentés ne font office que de pistes ouvertes à de futures recherches.

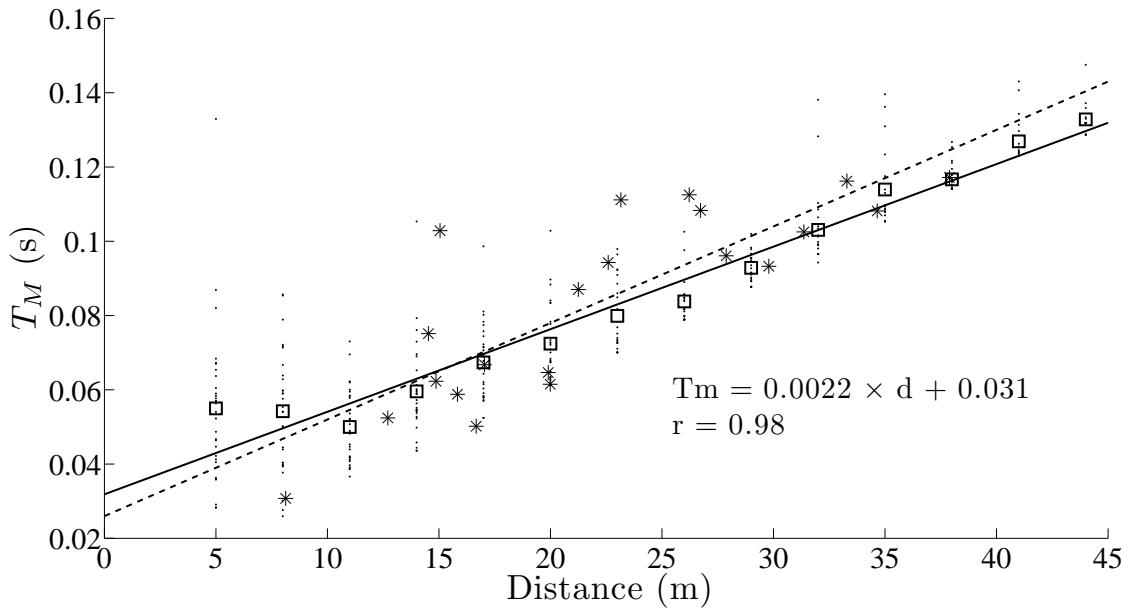


FIG. 5.25 – Résultats expérimentaux et numériques (et régressions linéaires). Le filtrage (type matériaux) est synthétisé par le modèle. Points : temps de mélange du modèle. Carrés : temps de mélange moyens du modèle. Etoiles : temps de mélange expérimentaux. Ligne pleine : régression linéaire sur les temps de mélange moyens du modèle. Ligne pointillée : régression linéaire sur les temps de mélange expérimentaux ( $T_m = 0.0026 \times d + 0.026$ ) ( $SRR = 5dB$ ).

l'estimateur et de son dictionnaire, ni de la reproductibilité de la source, alors nous pourrions conclure qu'il n'y a rigoureusement pas équpartition de l'énergie dans la salle. Néanmoins, il est préférable de réaliser plusieurs campagnes de mesures dans d'autres salles avant de trancher (Annexe C).

La Fig. 5.30 montre qu'il existe des similarités et des différences entre les coefficients directeurs des *CNA* selon les trois positions de la source sur scène en fonction de la position du récepteur. Ainsi, il semble qu'un auditeur assis sur l'axe de symétrie de la salle dans les premiers rangs (Source 1, réception en PF) est sujet à une évolution temporelle rapide du rayons sonores, à l'inverse d'un auditeur placé sur un balcon latéral (Source 1, réception BL1) dont la vue directe de la scène est cachée par les rebords du balcon. Tandis qu'un auditeur assis soit sur les côtés de la salle, soit sous un balcon ou le plafond, semble être sujet à une évolution temporelle plus lente du nombre de retours (Fig. 5.30).

Les différences de l'évolution du nombre de retours reçus par seconde ainsi que le nombre maximum de retours selon l'emplacement de la source et du récepteur peuvent avoir deux explications évidentes. La première concerne la nature de la source sonore utilisée pour réaliser la mesure de la réponse impulsionale. Il s'agit ici d'impulsions de revolver, dont la reproductibilité est étudiée et discutée dans le Chapitre 3 et dans [47]. La Section 5.2.4.3 présente la méthode utilisée pour définir le dictionnaire d'atomes utilisés par Matching Pursuit pour estimer les retours dans le signal. Cette définition de l'atome-mère (i.e. le son direct) admet naturellement des erreurs car elle se base sur une estimation. De plus, rappelons que l'estimation des retours dépend de la nature de la source employée (Fig. 5.10). La deuxième explication concerne l'emplacement de la source et du récepteur dans la salle lors de la mesure, ou autrement dit, la différence de marche des rayons sonores. Ces derniers, selon la position de la source rencontrent des obstacles, créé de la diffusion, sont absorbés, diffractés, etc., comme cela a été vu au Chapitre 2.

Explorer la relation possible entre la qualité d'écoute et l'évolution temporelle du nombre de

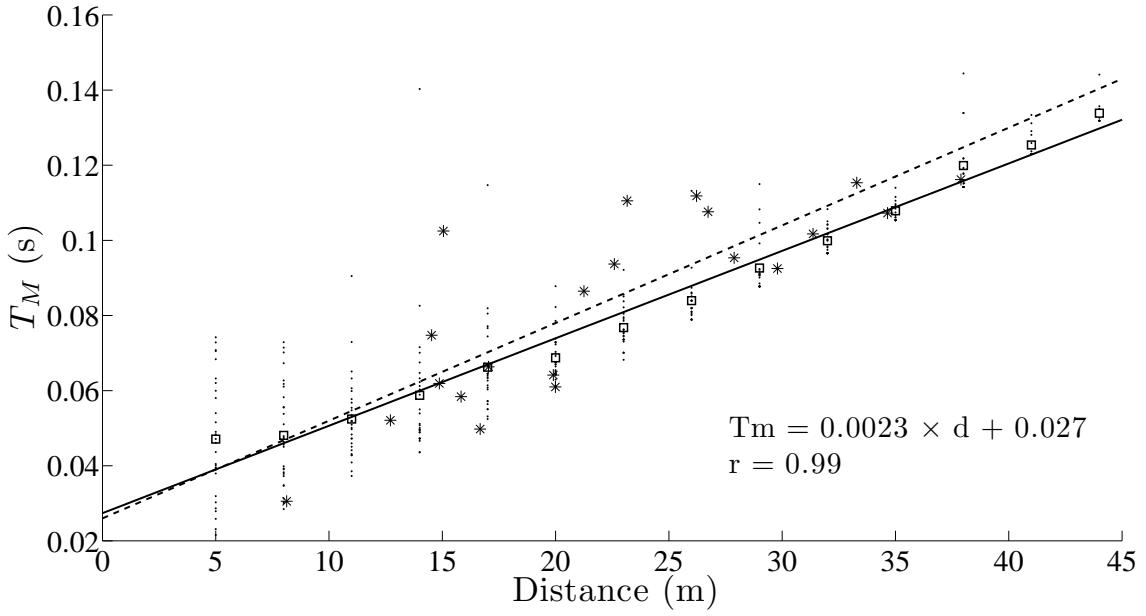


FIG. 5.26 – Résultats expérimentaux et numériques (et régressions linéaires). Le filtrage (Variations Spectrales Aléatoires -VSA) est synthétisé par le modèle. Points : temps de mélange du modèle. Carrés : temps de mélange moyens du modèle. Etoiles : temps de mélange expérimentaux. Ligne pleine : régression linéaire sur les temps de mélange moyens du modèle. Ligne pointillée : régression linéaire sur les temps de mélange expérimentaux ( $T_m = 0.0026 \times d + 0.026$ ) ( $SRR = 5dB$ ).

retours pourrait constituer une voie de recherche intéressante. En effet, elle permettrait peut-être de définir de nouveaux indices acoustiques.

### 5.3.3.2 Enregistrements B-format

Des mesures de réponses impulsionales à l'aide d'un microphone Soundfield sont réalisées en 21 positions de l'espace dans la Salle Pleyel. La source utilisée est un revolver. Le microphone Soundfiled utilise une combinaison de capteurs appelée B-format. Ce format fut imaginé par Michael Gerzon en 1975 [84] qui proposait de placer 3 capsules de directivité cardioïde sur les quatre faces d'un tétraèdre régulier. Par un simple jeu d'addition et de soustraction, nous obtenons quatre composantes fondamentales : W, X, Y et Z. Une mesure B-format permet une captation du champ de pression en un point de l'espace en 3 dimensions. Les lobes cardioïdes sont également appelés fonctions de directivités du premier ordre, car ils correspondent au terme du premier ordre dans la décomposition en harmoniques sphériques du champ acoustique. Dans la Salle Pleyel, les mesures sont réalisées en dirigeant l'axe X vers la place du chef d'orchestre. De telles mesures permettent d'étudier les différentes distributions temporelles des retours selon les directivités des mesures que l'on considère. Les Fig. 5.32-5.33 présentent les *CNA* pour les 4 directivités du B-format en deux endroits différents de la Salle Pleyel. Les différentes statistiques des retours selon les directivités sont mises en évidence. La Fig. 5.32 présente les *CNA*s d'une réponse impulsionnelle mesurée en champ direct et proche de la scène. La Fig. 5.33 présente les *CNA*s d'une réponse impulsionnelle mesurée sur un balcon. Précisons aussi que la source était invisible de cet emplacement. Il apparaît clairement que le nombre cumulé de retours est plus faible pour le canal Y (rayons sonores latéraux) que pour les autres canaux. De plus, le nombre de retours est plus important en provenance de la place du chef d'orchestre que des côtés ou

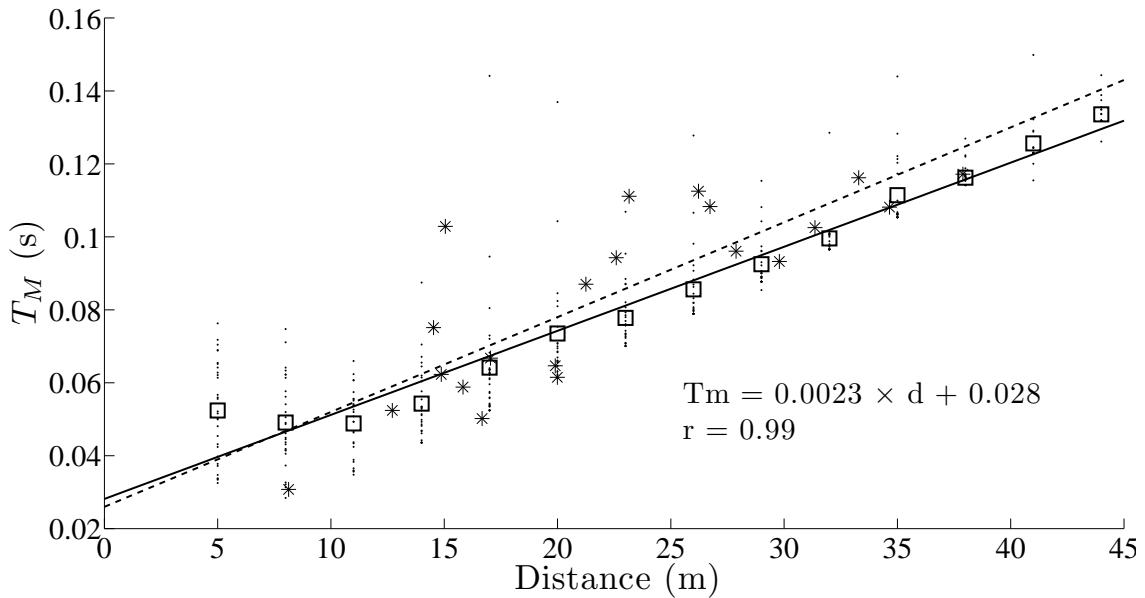


FIG. 5.27 – Résultats expérimentaux et numériques (et régressions linéaires). Le filtrage (en bandes d'octaves) et la diffusion sont synthétisés par le modèle. Points : temps de mélange du modèle. Carrés : temps de mélange moyens du modèle. Etoiles : temps de mélange expérimentaux. Ligne pleine : régression linéaire sur les temps de mélange moyens du modèle. Ligne pointillée : régression linéaire sur les temps de mélange expérimentaux ( $T_m = 0.0026 \times d + 0.026$ ) ( $SRR = 5dB$ ).

même du plafond. En revanche, l'évolution temporelle du nombre de retours est sensiblement la même selon les 4 directivités (les 4 CNAs ont environ tous le même coefficients directeur).

Si nous nous intéressons aux premiers instants de la réponse mesurée (Fig. 5.33), nous notons qu'il existe un moment à partir duquel les CNAs se différencient nettement les uns des autres. Cet instant correspond à un changement dans le comportement statistique de la réponse mesurée. Des observations préliminaires montrent que cet instant est différent du temps de mélange. Notons cependant qu'une plus large étude, comprenant plusieurs salles et d'autres sources sonores, est nécessaire à la validation de cette première observation. Enfin, une prochaine étape dans l'étude de la statistique des retours concerne l'analyse et la compréhension de ces différentes statistiques selon les directivités.

La Fig. 5.31 montre clairement qu'il existe des emplacements pour lesquels l'évolution du nombre de retours reçus est plus rapide que pour d'autres. Outre l'erreur d'estimation des retours due au Matching Pursuit, et le manque de reproductibilité de la source, il apparaît clairement qu'il y a des directions privilégiées dans lesquelles se crée un plus grand nombre de rayons sonores une fois le régime diffus établi. Toutefois, attention aux conclusions hâtives ! En effet, nous ne disposons pas aujourd'hui des connaissances et des outils suffisants pour affirmer avec certitude que 900 retours/sec. est très inférieur à 1100 retours/sec. Nous pouvons aussi nous interroger : cette différence du point de vue physique existe-t-elle du point de vue perceptif ? Sans pour autant apporter une réponse précise à cette question, la Section 5.3.4 aborde ce problème sans quantifier le nombre de retours mais simplement en cherchant s'il existe une différence perceptible.

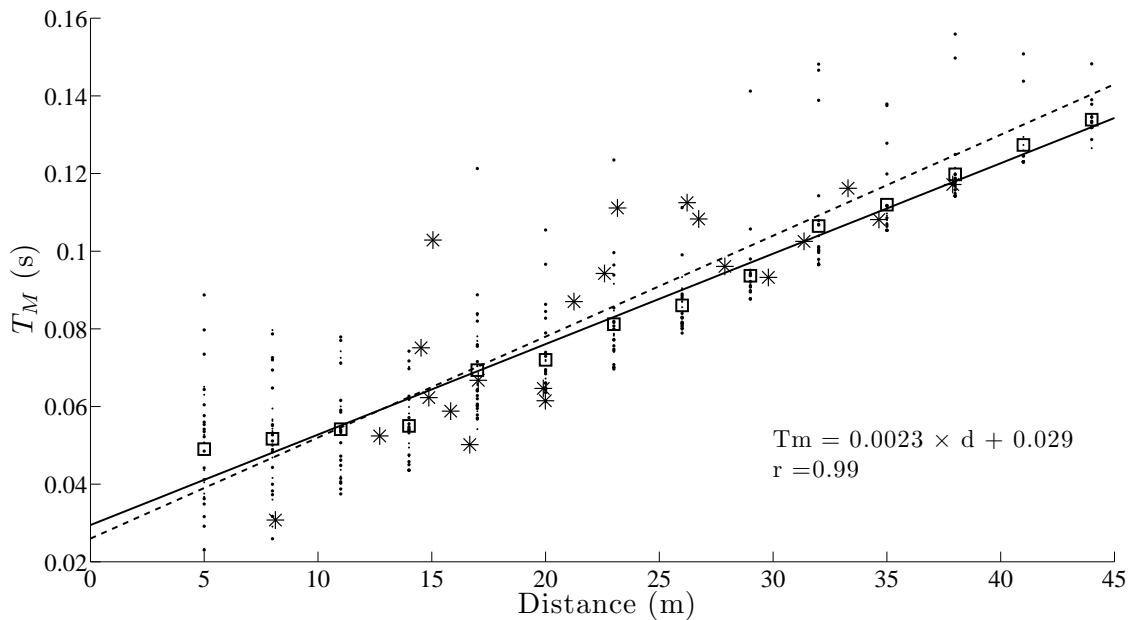


FIG. 5.28 – Résultats expérimentaux et numériques (et régressions linéaires). Le filtrage (type matériaux) et la diffusion sont synthétisés par le modèle. Points : temps de mélange du modèle. Carrés : temps de mélange moyens du modèle. Etoiles : temps de mélange expérimentaux. Ligne pleine : régression linéaire sur les temps de mélange moyens du modèle. Ligne pointillée : régression linéaire sur les temps de mélange expérimentaux ( $T_m = 0.0026 \times d + 0.026$ ) ( $SRR = 5dB$ ).

### 5.3.4 Estimation perceptive du critère d'arrêt de Matching Pursuit

L'article présenté ci-dessus met en avant l'importance du choix du critère d'arrêt ( $SRR$ ) de la poursuite [33]. En effet, il est montré que le nombre d'événements estimés dans le signal (ceci comprend les retours, la diffusion et le bruit de fond) est une fonction du  $SRR$ . La Fig. 5.15 souligne l'importance et l'impact du choix du critère d'arrêt. Le temps de mélange estimé est donc, nous l'avons vu, une fonction du  $SRR$ .

Ici, nous considérons une autre manière d'estimer le  $SRR$  le plus *approprié*. D'ailleurs, c'est là toute la question. En effet, déterminer le  $SRR$  du point de vue des indices acoustiques (et donc de la distribution temporelle énergétique de la RI<sup>6</sup>) est une méthode que l'on peut admettre. Nous pouvons tout aussi bien nous intéresser à une méthode d'estimation du  $SRR$  basée sur la perception du signal reconstruit après la poursuite.

Considérons trois emplacements dans la Salle Pleyel, où les réponses impulsionales sont mesurées à l'aide du système Ambisonics :

- en face de la scène, parterre, en vue directe de la source, et à 8m de distance de la place du chef d'orchestre ;
- au premier balcon tout au fond de celui-ci, en vue directe de la source, et à 30m de distance de la place du chef d'orchestre ;
- au second balcon latéral, en vue cachée de la source, et à 20m de distance de la place du chef d'orchestre.

Les quatre canaux ( $W$ ,  $X$ ,  $Y$ ,  $Z$ ) de chacune des réponses sont décomposés à l'aide de Matching Pursuit comme dans [33] pour différentes valeurs du  $SRR$  (0.5, 1.0, 1.5, 2.0, 2.5,

<sup>6</sup>RI : Réponse Impulsionnelle.

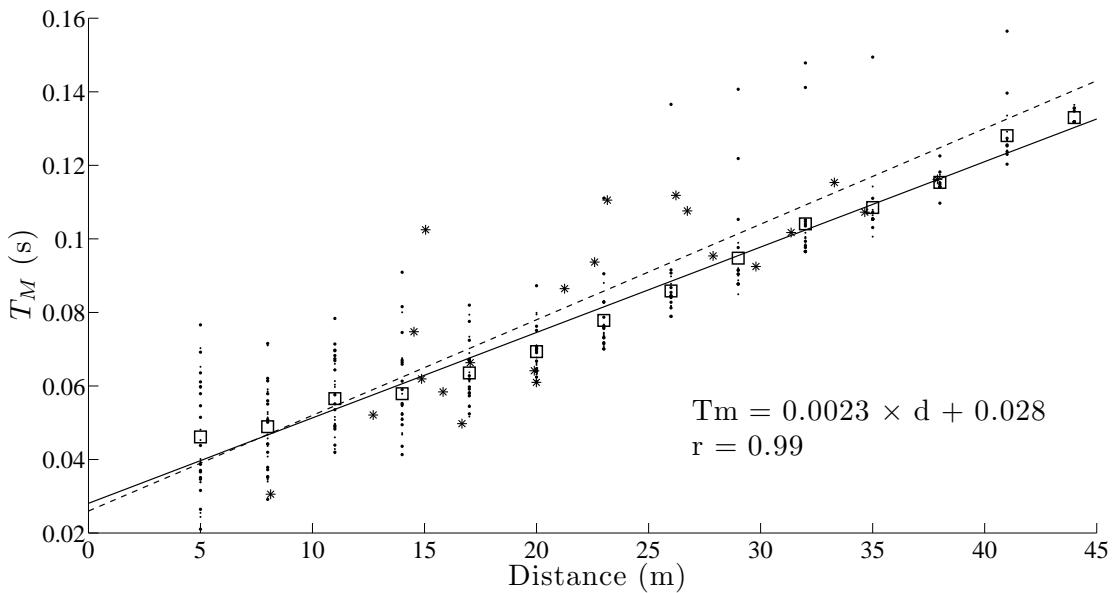


FIG. 5.29 – Résultats expérimentaux et numériques (et régressions linéaires). Le filtrage (VSA) et la diffusion sont synthétisés par le modèle. Points : temps de mélange du modèle. Carrés : temps de mélange moyens du modèle. Etoiles : temps de mélange expérimentaux. Ligne pleine : régression linéaire sur les temps de mélange moyens du modèle. Ligne pointillée : régression linéaire sur les temps de mélange expérimentaux ( $T_m = 0.0026 \times d + 0.026$ ) ( $SRR = 5dB$ ).

Séquence n°	Caractéristiques
1	Grand orchestre / ensemble de violons / peu percussif
2	Harpe seule / percussif
3	Petit orchestre / ensemble de violons / assez percussif

TAB. 5.4 – Caractéristiques générales des morceaux de musiques anéchoïques.

3.0, 3.5, 4.0, 4.5, 5.0, 6.0, 7.0, 8.0, 9.0, 10.0, 15.0, 20.0 dB). Les signaux reconstruits après chaque poursuite sont convolus par trois morceaux de musique classique enregistrés en chambre anéchoïque<sup>7</sup> (Table 5.4).

Chaque séquence sonore est diffusée sur le système de reproduction Ambisonics de la salle sèche du LAM [85]. La base de cette salle a une forme hexagonale et comprend douze haut-parleurs plus un caisson basses-fréquences. Les douze enceintes de proximité sont suspendues sur une armature métallique cachée de la vue des sujets par des plaques de toile grise opaques mais acoustiquement transparentes. Six enceintes sont à mi-hauteur sur les sommets du grand hexagone, six au milieu des côtés de l'hexagone répartis en deux groupes de trois (3 à hauteur du plafond sur un triangle équilatéral, et 2 au sol sur un autre triangle équilatéral alterné). Ainsi les plans horizontal et vertical sont couverts. Le niveau des haut-parleurs et du caisson basses fréquences sont calibré, par une mesure du niveau au centre de la salle (i.e., le point d'écoute) en niveau avec un son de référence de 1kHz. La fréquence de coupure est entre les enceintes de

<sup>7</sup>Séquences sonores tirées de : *Anechoic orchestral music recording*, Pure Gold Collection. Séquence 1 : W.A Mozart, Les Noces des Figaro, (80 premières secondes). Séquence 2 : Johann et Joseph STRAUSS, Pizzicata Polka (80 premières secondes). Séquence 3 : J. Brahms, Premier mouvement, Symphonie n°4, mesures 386-407 (35 premières secondes).

Séquence n°	Caractéristiques	$SRR_{lim}$ (dB)	$\sigma$ (%)
1	Grand orchestre / ensemble de violons / peu percussif	5.0	34
2	Harpe seule / percussif	8.0	17
3	Petit orchestre / ensemble de violons / assez percussif	7.0	22

TAB. 5.5 – Caractéristiques générales des morceaux de musiques anéchoïques et le  $SRR$  limite moyen associé et l'écart-type associé.

proximité et le caisson est fixée à  $100Hz$  [85]. Le caisson est installé dans un coin de la salle afin que le maximum de modes de la salle soient excités et aussi afin de procurer une sensation d'enveloppement optimal [85]. Les réponses impulsionales Ambisonics testées sont décomposées en 13 canaux de restitution (il y a 12 enceintes de proximité plus un caisson basses fréquences) grâce au logiciel AudioMulch.

Les sujets passant le test d'écoute doivent répondre à une question à partir d'une interface graphique créée sous MatLab et inspirée de tests MUSHRAM [86] (Fig. 5.34) utilisés dans le domaine de la compression audio. L'exercice proposé est le suivant : le sujet doit reconnaître une séquence sonore qui lui est donnée comme référence parmi deux séquences sonores  $A$  et  $B$ . Dans ce test, la séquence de référence correspond à une décomposition des réponses impulsionales avec un  $SRR = 20dB$ . En d'autres termes, la séquence de référence correspond à la meilleure approximation de la réponse impulsionnelle considérée. La séquence  $A$  (ou  $B$ ) est soit la séquence de référence, soit une approximation de la RI obtenue pour un  $SRR$  différent. La séquence de référence est soit  $A$  soit  $B$ , obligatoirement, mais jamais les deux en même temps. L'idée du test est de différencier la séquence de référence d'une séquence obtenue avec un  $SRR$  qui va croissant au fur et à mesure des bonnes réponses du sujet. Autrement dit, au fur et à mesure que le sujet reconnaît bien la séquence de référence, il se voit proposer un exercice de plus en plus difficile, car le  $SRR$  augmente et se rapproche donc de celui utilisé pour la séquence de référence. Dans la pratique, le test commence avec le  $SRR$  le plus faible, c'est-à-dire  $0.5dB$ . A chaque bonne réponse, nous demandons au sujet de confirmer sa réponse en lui proposant (sans qu'il le sache) les mêmes séquences sonores, soit le même niveau de difficulté. Si la bonne réponse est confirmée à  $SRR = 0.5dB$  par exemple, le sujet passe à la difficulté supérieure ( $SRR = 1.0dB$ ), et ainsi de suite. Lorsque le sujet ne reconnaît pas la référence parmi les séquences  $A$  et  $B$ , le test lui propose de descendre d'un niveau la difficulté (par exemple, passage de  $SRR = 9dB$  à  $SRR = 8dB$ ). L'intérêt de ce test est de trouver les valeurs du  $SRR$  pour lesquels les réponses du sujet vont osciller autour d'une valeur limite. Cette valeur limite est telle que le sujet ne fait plus la distinction entre la séquence de référence et l'autre séquence. Perceptivement, les deux séquences sont alors identiques. Nous retrouvons là en quelque sorte la méthode adoptée dans la Section 5.2.4.2 pour déterminer la valeur de  $SRR$  à l'aide des indices acoustiques. Ce test, appelé "Adaptive staircase technique", est proposé par Levitt [87] et étudié par Kollmeier *et al.* [88].

Vingt-six sujets ont passé le test sur les trois séquences sonores et pour les trois réponses impulsionales différentes. La limite de différence perceptible varie en fonction du morceau anéchoïque. Ainsi, il est plus aisé de reconnaître la séquence de référence lorsque le morceau de musique est percussif. Le  $SRR$  limite ( $SRR_{lim}$ ) dans ce cas-là est plus élevé que lorsque la séquence ne comporte que des ensembles de violons (Table 5.5).

Au regard des résultats (Table 5.5), nous pouvons considérer que le *meilleur* critère d'arrêt, d'un point de vue perceptif, est celui maximum, c'est-à-dire  $SRR = 8dB$ . Pour une telle valeur du  $SRR$ , la relation du temps de mélange est directement liée au temps d'arrivée du son direct à la

$SRR$ (dB)	Relation temps de mélange/distance
5	$Tm = 0.0026 \times d + 0.026$ ( $r = 0.82$ )
6	$Tm = 0.0027 \times d + 0.017$ ( $r = 0.90$ )
8	$Tm = 0.0030 \times d - 0.001$ ( $r = 0.99$ )

TAB. 5.6 – Relation entre le temps de mélange et la distance pour différentes valeurs du critère d’arrêt  $SRR$ . Note : l’ordonnée à l’origine de la relation du temps de mélange est négative ( $-0.001s$ ). Cette valeur n’est naturellement pas recevable car le mélange s’établirait à un temps négatif pour une position identique de la source et du récepteur. En toute rigueur, nous devrions écrire :  $Tm = 0.0030 \times d$ .

$SRR$ (dB)	Mesure n°1	Mesure n°5	Mesure n°6
5	<b>9</b>	38	32
6	<b>11</b>	32	25
7	17	57	61
8	25	74	75

TAB. 5.7 – Nombre de retours estimés dans une fenêtre de  $24ms$  pour trois points de mesures dans la Salle Pleyel et pour différentes valeurs du critère d’arrêt ( $SRR$ ).

position du récepteur dans la salle (Fig. 5.15 et 5.35). Autrement dit, le système est quasiment immédiatement diffus. En revanche, si nous considérons la valeur limite moyenne du critère d’arrêt sur les trois séquences sonores ( $S\bar{R}R = 6dB$ ), la relation du temps de mélange à la distance est proche de celle estimée dans la Section 5.2.6.2 (Fig. 5.36). Le tableau 5.7 regroupe ces différents résultats.

Nous ne pouvons naturellement pas conclure d’une manière définitive au sujet du *meilleur* critère d’arrêt de Matching Pursuit (si tant est qu’il en existe un!). Cependant, les précédents résultats, qui, il faut le rappeler, ne valent que pour une seule salle, montrent que les valeurs du critère d’arrêt calculées sur la base des indices acoustiques et sur la base de tests perceptifs ( $SRR$  moyen) sont similaires. Dans ce cas-là, les temps de mélange estimés sont également similaires. En revanche, si le critère d’arrêt est celui maximal obtenu lors des tests perceptifs, alors les temps de mélange de la salle sont directement liés au temps d’arrivée du premier front d’onde à une position dans l’espace : la salle de concert est immédiatement diffuse. En conclusion, estimer le critère d’arrêt perceptivement revient à estimer le temps de mélange.

D’autre part, grâce à MP, nous sommes à même de vérifier si le critère proposé par Polack [9] (Section 5.2.3) est validé ou non. En d’autres termes, il est possible de dénombrer les retours estimés, dans une fenêtre de  $24ms$ , pour chaque réponse mesurée et pour chaque valeur du critère d’arrêt ( $SRR$ ) (Fig. 5.37). Considérons les RIIs qui nous ont servi aux tests d’écoute. Le nombre de retours estimés pour chacune de ces mesures pour les critères d’arrêt 5, 6, 7, 8 dB sont regroupés dans le tableau ci-après (Table 5.7).

En conclusion, le nombre de retours dans une fenêtre de  $24ms$  estimés selon les différentes valeurs du critère d’arrêt estimés soit lors de tests perceptifs, soit à l’aide d’une étude basée sur la variations d’indices acoustiques, ne sont pas en accord avec ceux prédis par Polack et Cremer [9, 12]. Par ailleurs, il aurait été illusoire de s’attendre à ce que le nombre de retours soit de 10 dans une fenêtre de  $24ms$  pour différents  $SRR$  et différents emplacements de mesure. En effet, rappelons que chaque réponse est mesurée dans des conditions qui sont toujours particulières (proximité d’une paroi, rebord d’un balcon, plafond, etc.). Notons cependant que la mesure n°1,

mesurée en champ proche (Table 5.7), est en accord avec les paramètres proposés par Cremer, pour  $5 \leq SRR < 8dB$ . Encore une fois, une étude concernant un grand nombre de salles et des sources plus reproductibles que le revolver [47] sont nécessaires avant de tirer quelque conclusion quant à la validité du critère du temps de mélange proposé par Polack.

### 5.3.5 Ouvertures possibles

L'estimation des retours d'une réponse impulsionnelle de salle laisse envisager plusieurs pistes d'applications.

#### 5.3.5.1 Angles d'incidence des retours

A partir des enregistrements B-format, il est possible d'estimer les angles d'incidence des retours au point de réception. La mesure Ambisonics au premier ordre est constituée de quatre canaux  $W$ ,  $X$ ,  $Y$  et  $Z$  qui sont reliés aux angles d'incidence (azimut et site) par les formules suivantes :

$$W = x \quad (5.17)$$

$$X = x.\sqrt{2}.\cos(\theta).\cos(\phi) \quad (5.18)$$

$$Y = x.\sqrt{2}.\sin(\theta).\cos(\phi) \quad (5.19)$$

$$Z = x.\sqrt{2}.\sin(\phi), \quad (5.20)$$

où  $x$  est le signal omnidirectionnel de la réponse impulsionnelle,  $\theta$  l'azimut et  $\phi$  le site.

L'azimut  $\theta$  est estimée par :

$$\tilde{\theta} = \text{atan} \left( \frac{Y}{X} \right). \quad (5.21)$$

Le site  $\phi$  est estimé par :

$$\tilde{\phi} = \text{asin} \left( \frac{Z}{W.\sqrt{2}} \right). \quad (5.22)$$

Grâce à l'estimation des retours du signal  $\tilde{x}$ , il est désormais possible non plus d'estimer les angles d'incidence de l'énergie en un point de l'espace [34], mais bien de s'intéresser aux retours-mêmes. Il est néanmoins nécessaire d'identifier la diffusion dans le signal  $\tilde{x}$  afin de s'affranchir de la directivité d'incidence de celle-ci (Section 5.3.5). Les relations (5.21) et (5.22) permettent d'estimer  $\tilde{\theta}$  et  $\tilde{\phi}$  à  $\pi$  près. Afin de lever l'ambiguité, une étude de signe est nécessaire. A  $\tilde{\phi}$  fixé, le signe de  $X$  et  $Y$  dépend de  $\tilde{\theta}$  et de  $x$ . Il faut par conséquent s'intéresser aux signes de  $\cos(\theta)$  et  $\sin(\phi)$ , soit aux signes de  $\frac{X}{W}$  et de  $\frac{Y}{W}$ . Dans la pratique, on s'intéresse aux signes des produits  $W.X$  et  $W.Y$ , afin de se prévenir des zéros. Les produits  $W.X$  et  $W.Y$  sont proportionnels à  $x^2$ , toujours positif.

Des méthodes plus élaborées (analyses par bandes d'octaves, analyse en composantes principales, etc.), mais qui comportent elles aussi leurs erreurs d'estimation (notamment en basses fréquences), ont été employées afin d'estimer l'angle d'incidence en un point de l'espace de sources sonores mouvantes, avec un enregistrement B-format [89, 90].

A partir de la mesure de 21 réponses impulsions réalisées dans la Salle Pleyel à l'aide d'un revolver et d'un microphone Ambisonics, et à l'aide des Eq. 5.21-5.22 et de Matching Pursuit, nous pouvons estimer les angles d'incidence des rayons sonores au point de réception. La Fig. 5.38

représente les azimuts et les sites des retours estimés par MP sur une réponse impulsionale au premier rang dans l'axe de symétrie de la Salle Pleyel. Nous remarquons que le nombre maximum de retours ne provient pas d'un angle de  $0$  radians ; cela est certainement dû au mauvais alignement de l'axe  $X$  du microphone Ambisonics avec la place du chef d'orchestre. La distribution de ces angles d'incidence n'a pas été caractérisée et étudiée en détails comme nous l'avons déjà dit. Cependant, en première approximation, les observations sur toutes les réponses mesurées nous amènent à dire que ces distributions seraient peut-être normales et centrées en  $0$ . Enfin, notons que les résultats présentés ici concernent uniquement 21 points de mesures ; la source sonore était un revolver dont la reproductibilité de rayonnement n'a à ce jour pas encore été caractérisée.

**Remarque** Notons que les mesures Ambisonics permettent également de vérifier la présence de la diffusion. C'est ce qu'a montré et utilisé Hurtado [39]. Cette présence est mise en évidence grâce aux relations qui lient l'intensité du signal  $x$  et les quatre canaux  $W$ ,  $X$ ,  $Y$  et  $Z$  :

$$W^2 = x^2 \quad (5.23)$$

$$X^2 + Y^2 + Z^2 = 2x^2. \quad (5.24)$$

D'où, en absence de diffusion :

$$W^2 - \frac{1}{2}(X^2 + Y^2 + Z^2) = 0. \quad (5.25)$$

Par conséquent, si la relation (5.25) n'est pas vérifiée, alors nous pouvons conclure quant à la présence de diffusion dans le signal  $x$  mesuré. Signalons que cette approche ne permet que d'estimer les instants auxquels la diffusion est prépondérante, mais pas obligatoirement le début du champ diffus. En effet, comme nous l'avons dit précédemment (Section 5.2.7), la diffusion est captée entre deux retours successifs et ce dès l'arrivée du premier retour. Outre le cas des salles immédiatement mélangeantes, l'étude des différentes directivités du microphone B-format n'est pas un outil suffisamment robuste et fiable pour estimer le début du champ diffus.

### 5.3.5.2 Estimation du filtrage global de la salle

L'estimation des temps d'occurrence des retours d'une réponse impulsionale de salle permet, en théorie, d'estimer le filtrage qu'a subi le rayon sonore durant son trajet entre le point source et le point de réception. En effet, nous pouvons émettre l'hypothèse que le filtrage de ce rayon peut être estimé en calculant le rapport des spectres d'un retour et du son direct. Cependant, la difficulté réside maintenant dans l'estimation des bornes temporelles du dit retour. En première approximation, nous pouvons supposer que les retours du signal ont tous une durée égale à celle du son direct. Le Chapitre 2 nous rappelle néanmoins qu'il s'agit là d'une approximation grossière. En effet, au cours du temps, les formes d'onde des retours interceptés au point de réception sont de plus en plus étalées dans le temps, mais ont lieu également à des temps de plus en plus rapprochés, pour se superposer après le temps de mélange (voir aussi [33]).

Une possibilité à explorer est de constituer un dictionnaire dans lequel chacun des atomes est une version filtrée du son direct et de durée différente. Se posent alors les questions du filtrage à utiliser, et des durées de chaque atome. La Fig. 5.39 est un exemple de décomposition d'une réponse expérimentale à l'aide d'un dictionnaire composé du son direct filtré en tiers de bandes d'octave. Notons que tous les atomes ont la même durée et que la représentation temps/fréquence obtenue permet d'une part d'estimer la coloration de la salle, et d'autre part, de retrouver la coloration de l'impulsion émise par la source (ici, un revolver).

Estimer le filtrage de la salle comme une fonction du temps revient à estimer la perte de mémoire progressive du système, c'est-à-dire son entropie, d'une autre manière qu'en appliquant la formule de l'entropie de Shannon issue de la théorie de l'information [68] (Chapitre 4).

### 5.3.5.3 Estimation de la diffusion

Cette Section présente une manière différente de celle décrite précédemment d'estimer la diffusion (Section 5.3.5.1). En effet, cette estimation est possible à partir d'enregistrement omnidirectionnels. L'estimation de la diffusion nécessite en premier lieu d'identifier les retours, puis de soustraire leur contribution au signal. En second lieu, il convient de filtrer par bande d'octaves hautes fréquences le signal, ainsi que le dictionnaire d'atomes. Enfin, Matching Pursuit permet une estimation des événements restant dans le signal, c'est-à-dire la diffusion.

C'est en quelque sorte le résultat obtenu précédemment, lorsque MP, dont le dictionnaire est uniquement composé du son direct, est appliqué aux RIs. Le signal restant, après soustraction de la contribution des retours estimés, est ce que nous avons appelé le résidu. Il est par conséquent possible d'émettre l'hypothèse suivante : le résidu de la poursuite est une estimation de la diffusion. Cette hypothèse sera d'autant plus vraisemblable que le dictionnaire sera constitué d'une grande variété d'atomes, chacun étant une version filtrée du son direct. Comme pour l'estimation du filtrage global de la salle (Section 5.3.5.2), l'estimation de la diffusion doit être réalisée à l'aide d'un dictionnaire d'atomes de durées variables. Enfin, des mesures spatiales de RIs permettraient de caractériser la direction d'incidence de la diffusion au cours du temps [91, 92].

### 5.3.6 Conclusion

L'application de l'algorithme de Matching Pursuit aux réponses impulsionales de salles ouvre de nombreuses perspectives. Parmi les plus intéressantes du point de vue des systèmes dynamiques, nous retiendrons certainement celle de la détection des retours et l'étude de leur statistique qui débouche sur l'estimation du temps de mélange. De plus, l'emploi de MP rend possible l'étude de la distribution angulaire au cours du temps des rayons dans une salle, la caractérisation angulaire de la diffusion et l'estimation du filtrage global de la salle comme une fonction du temps.

Une autre voie d'exploration de la statistique des retours, celle des retours "positifs" et "négatifs" a été envisagée sans que les résultats soient pour autant reportés dans ce mémoire. En effet, l'observation de nombreuses réponses impulsionales de grandes salles semblent montrer que la distribution des retours d'amplitude positive et négative évolue au cours du temps. En effet, nous observons que les premiers temps sont constitués principalement de retours d'amplitude positive qui tendent à se mélanger de manière plus ou moins homogène avec les retours d'amplitude négative.

L'estimation des retours permet également d'estimer le phénomène de masquage entre les retours [93], à partir du temps qui sépare deux retours consécutifs et du niveau de chaque retour. Ceci ouvre sans doute la voie à la définition de nouveaux indices d'acoustique des salles. Enfin, l'estimation des temps d'arrivée des retours permettrait d'estimer le volume d'une salle uniquement à partir de la connaissance d'une ou plusieurs réponses impulsionales (Eq. 6.2) [69]. Peut-être s'agit-il là d'un critère d'arrêt de Matching Pursuit pertinent ? En effet, la poursuite s'arrêterait lorsque le nombre de retours estimés permet de retrouver le volume de la salle dont nous étudions une réponse impulsionale.

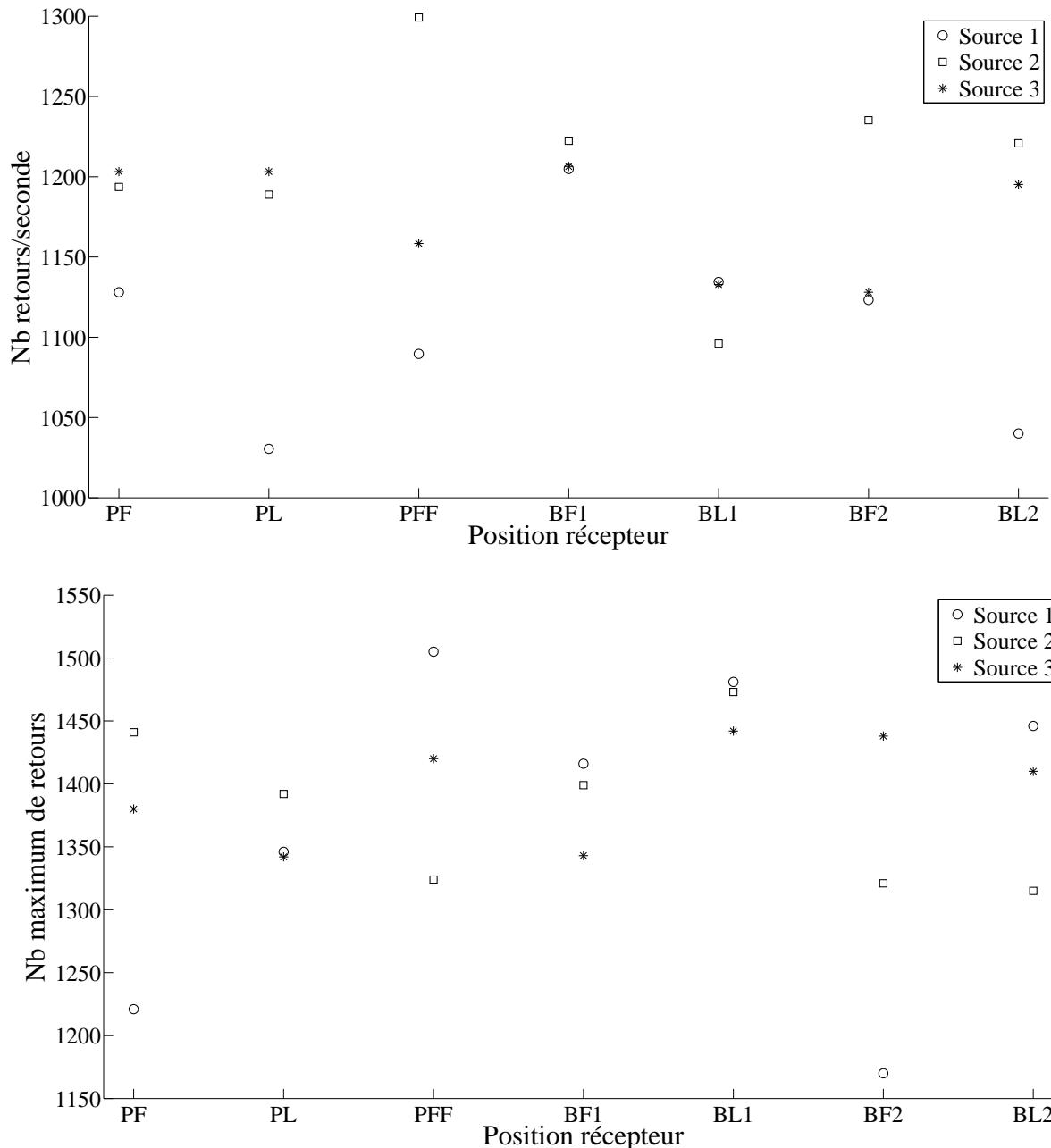


FIG. 5.30 – Haut : nombre de retours par seconde en fonction de l'emplacement de réception choisi dans la Salle Pleyel (mesures omnidirectionnelles) pour trois points d'émission sur la scène. Bas : nombre maximum de retours captés au point de réception pour trois points d'émission sur la scène. Note : PF= Parterre-Face ; PL=Parterre Latéral ; PFF=Parterre Face au Fond ; BF1=Balcon Face 1er étage; BL1=Balcon Latéral 1er étage ; BF2=Balcon Face 2nd étage ; BL2=Balcon Latéral 2nd étage.

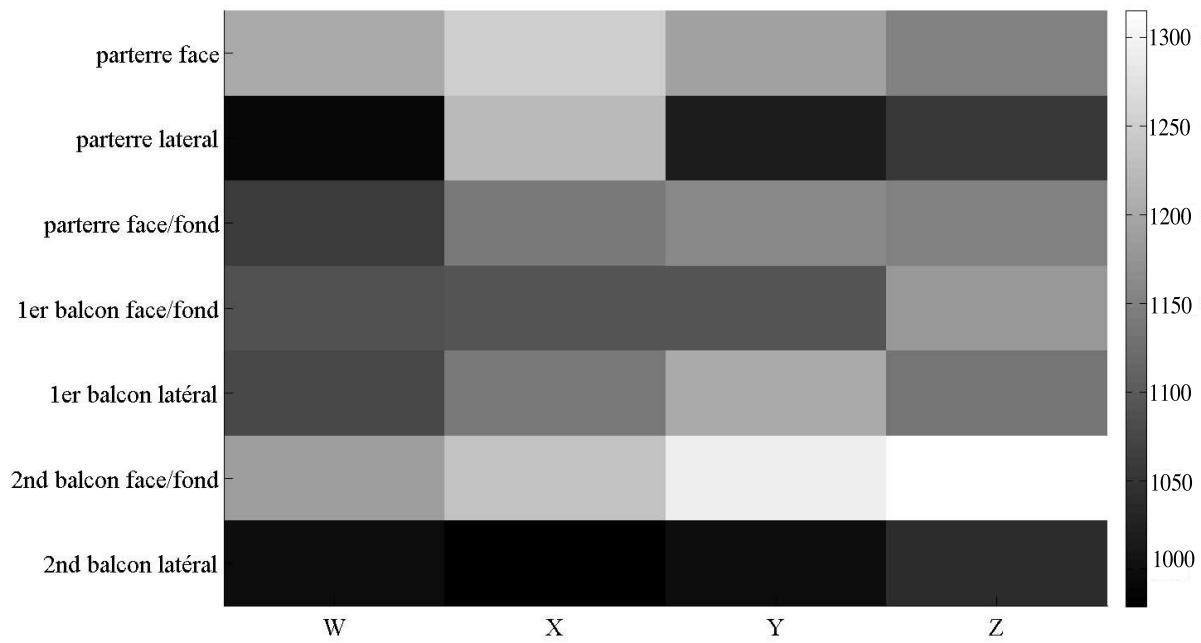


FIG. 5.31 – Nombre de retours par seconde (dans le champ diffus) en fonction de l'emplacement de réception choisi dans la Salle Pleyel et de la direction d'incidence des rayons (mesures B-format). Ici, une seule position de source est considérée. Note :  $W$  : captation omnidirectionnelle ;  $X$  : directivité avant/arrière ;  $Y$  : directivité latérale ;  $Z$  : directivité dessus/dessous. On rappelle que lors de mesures, l'axe  $X$  est dirigé vers la source située à la place du chef d'orchestre.

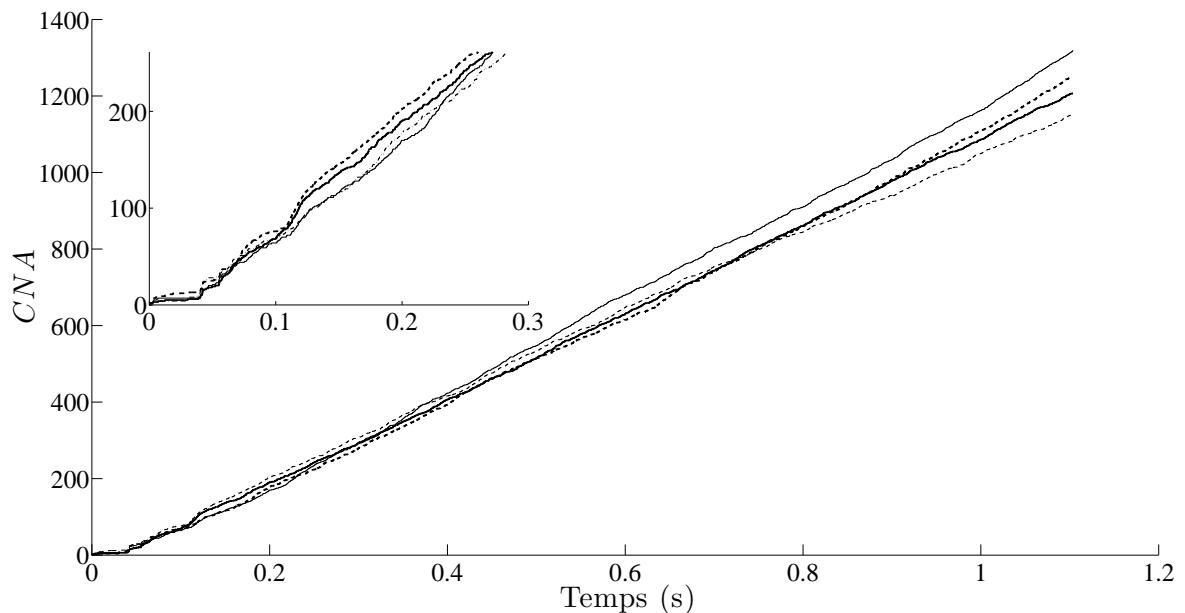


FIG. 5.32 – Les 4 CNAs correspondant aux 4 canaux de B-format. Mesure au premier rang du parterre (source  $S1$ , réception  $PF$ ). La source est en champ direct à 8m. Trait plein gras : canal  $W$ . Trait pointillé gras : canal  $X$ . Trait pointillé : canal  $Y$ . Trait plein : canal  $Z$ . Note : le temps de propagation entre la source et le récepteur n'est pas pris en compte.

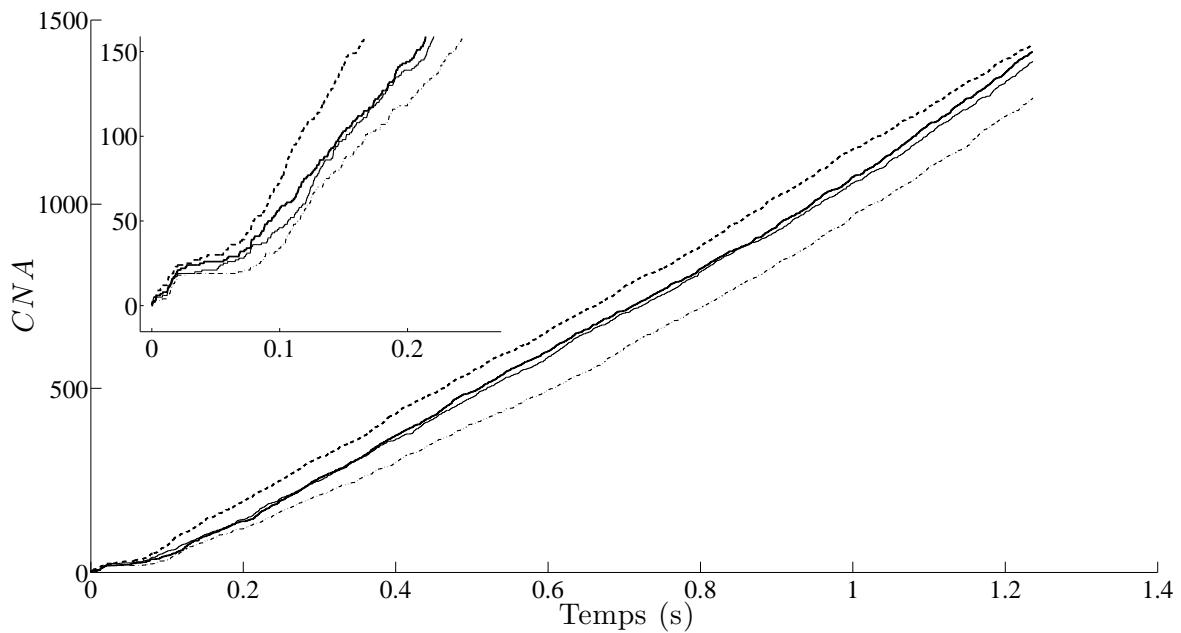


FIG. 5.33 – Les 4 CNAs correspondant aux 4 canaux de B-format. Mesure (source S3, réception BL1) sur un balcon à 18m de la source qui est cachée. Trait plein gras : canal W. Trait pointillé gras : canal X. Trait pointillé : canal Y. Trait plein : canal Z. Note : le temps de propagation entre la source et le récepteur n'est pas pris en compte.

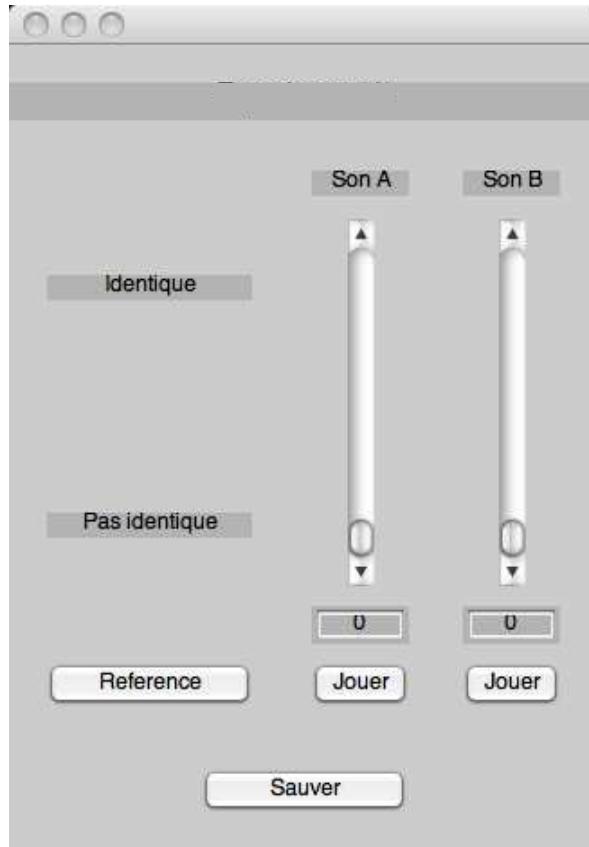


FIG. 5.34 – Interface de test inspiré des tests MUSHRAM [86].

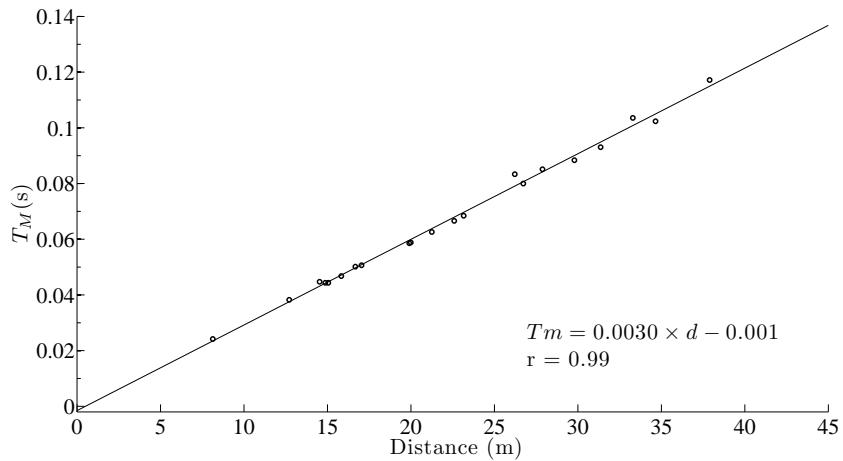


FIG. 5.35 – Temps de mélange estimé dans la Salle Pleyel à l'aide Matching Pursuit pour un critère d'arrêt  $SRR = 8dB$ .

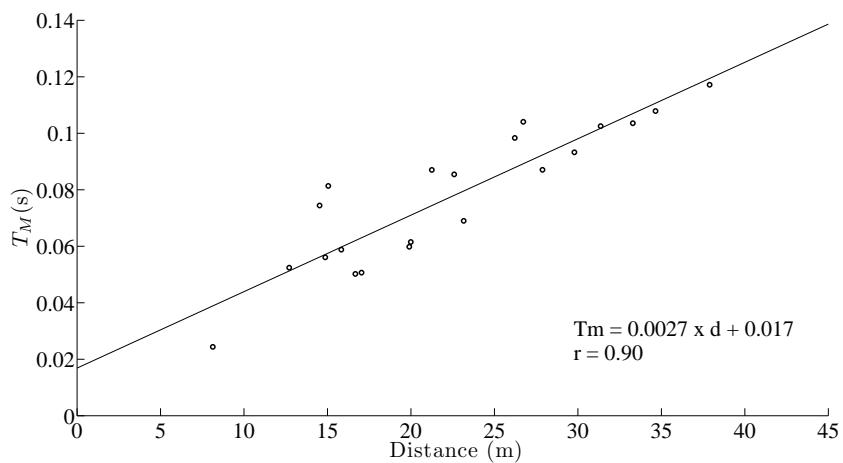


FIG. 5.36 – Temps de mélange estimé dans la Salle Pleyel à l'aide Matching Pursuit pour un critère d'arrêt  $SRR = 6dB$ .

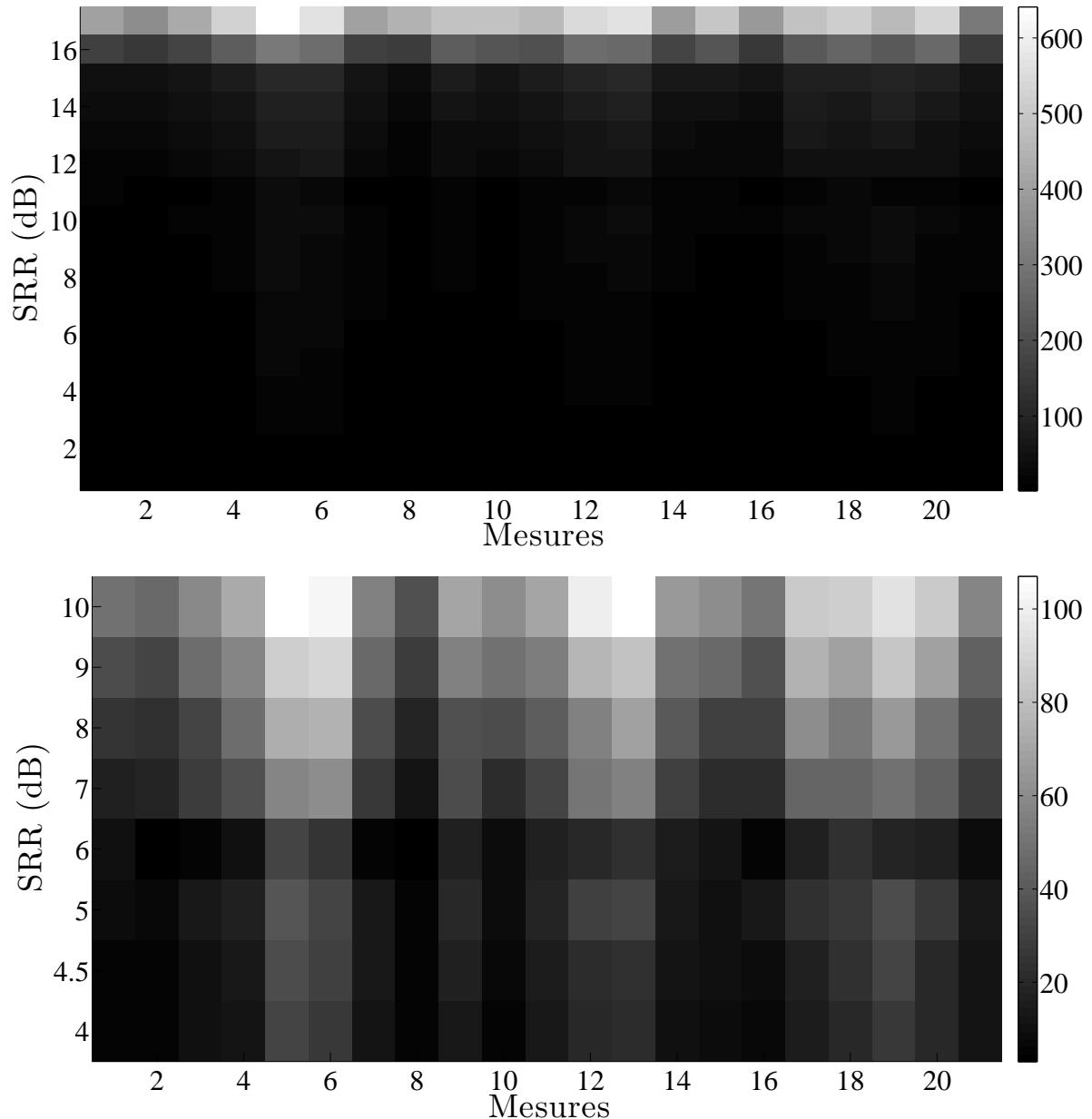


FIG. 5.37 – En haut : nombre de retours estimés (fenêtre de 24ms) dans la Salle Pleyel à l'aide Matching Pursuit pour des critères d'arrêt allant de 0.5 à 20dB. En bas : nombre de retours estimés dans la Salle Pleyel à l'aide Matching Pursuit pour des critères d'arrêt allant de 4 à 10dB.

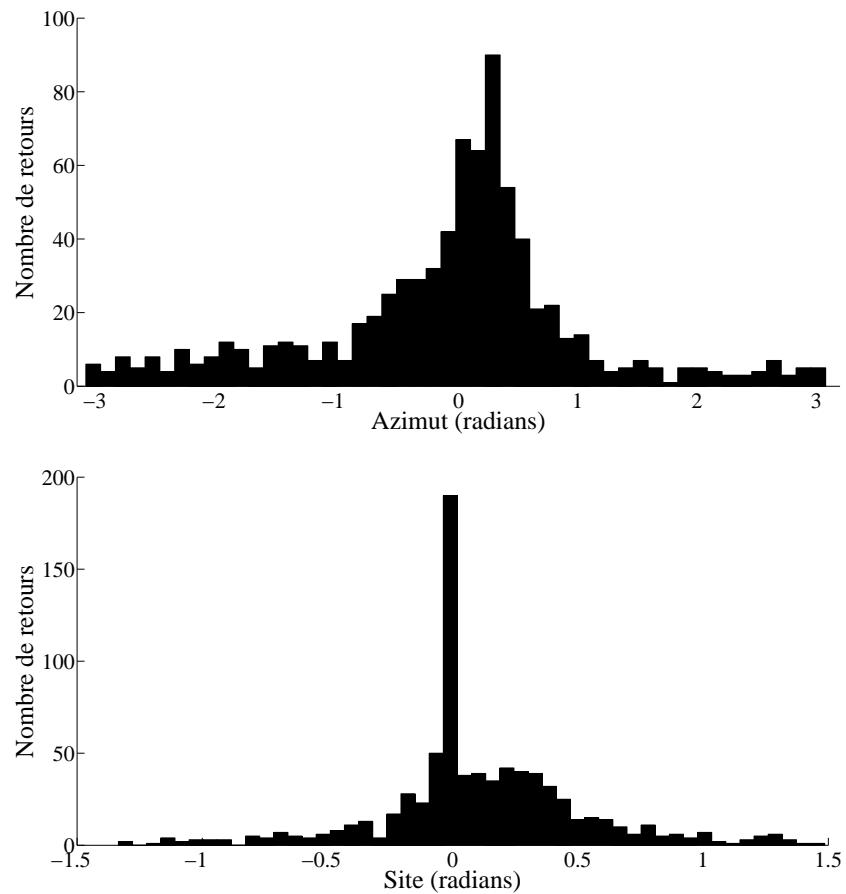


FIG. 5.38 – Haut : azimuts des retours estimés dans la Salle Pleyel à l'aide Matching Pursuit pour un critère d'arrêt  $SRR = 5dB$  au premier rang. Bas : sites des retours estimés dans la Salle Pleyel à l'aide Matching Pursuit pour un critère d'arrêt  $SRR = 5dB$  au premier rang.

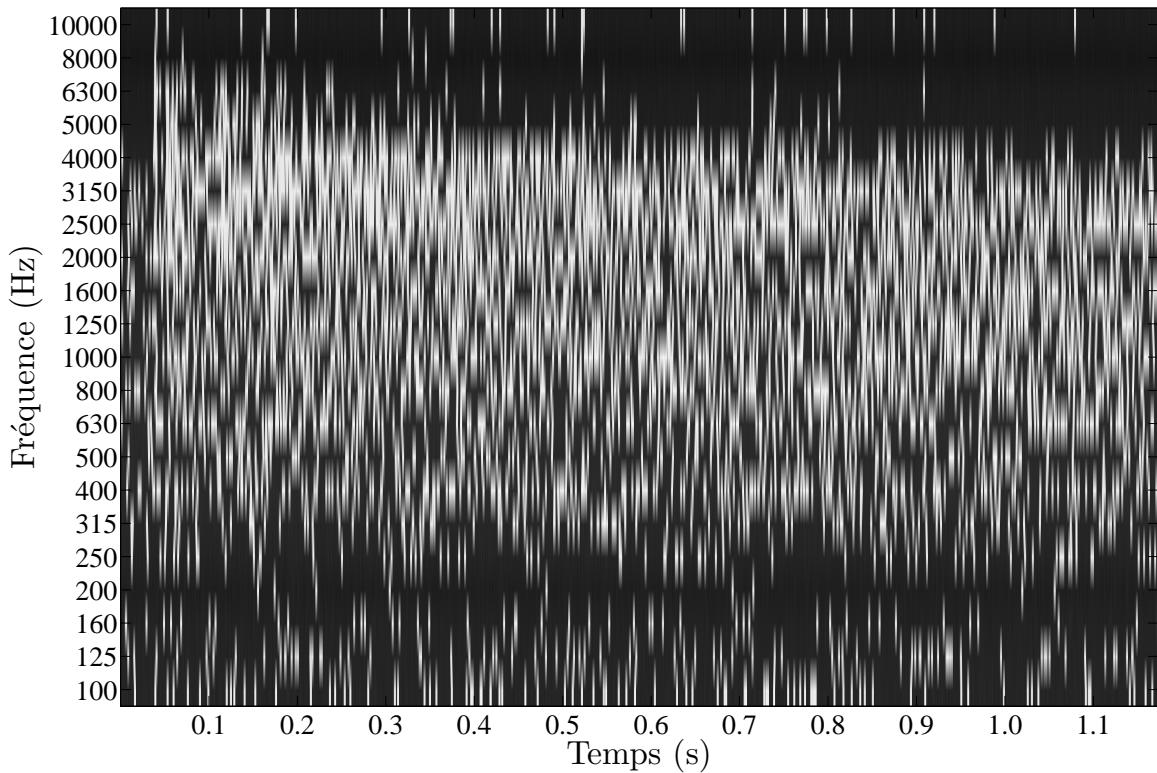


FIG. 5.39 – Décomposition d'une réponse impulsionnelle expérimentale en 1/3 de bandes d'octave ( $SRR = 5dB$ ). Chaque trait blanc marque l'occurrence d'un retour dans le plan temps/fréquence. On retrouve la coloration particulière de l'impulsion de revolver [47] ( voir aussi Chapitre 3).



# Chapitre 6

## Estimation du temps de mélange à l'aide de l'eXtensible Fourier Transfrom

### 6.1 Préambule

Ce Chapitre propose une méthode d'estimation de l'évolution temporelle des composantes fréquentielles d'un signal. On s'intéresse ici aux cas des réponses impulsionales de salles. La méthode proposée est une transformée de Fourier à fenêtre extensible, d'où son nom : *eXtensible Fourier Transform*. Dans le cadre des travaux présentés, la XFT sert à l'estimation du temps de mélange (Chapitres 4 et 5). La XFT permet d'estimer le temps à partir duquel un signal est statistiquement (dans les espaces temps et fréquence) distribué au regard de son état initial. Plus précisément, l'article que nous présentons par la suite, s'intéresse à la phase des spectres de la XFT et montre qu'il existe un point de rupture dans le comportement statistique du signal, qui a lieu au temps de mélange. Ce dernier est estimé sur une base de réponses impulsionales mesurées dans la Salle Pleyel [52]. Comme dans le cas de l'estimation des retours et du temps de mélange avec Matching Pursuit [33] (Chapitre 5), on établit qu'il existe une relation linéaire entre le temps de mélange et la distance source/récepteur. Ce résultat est confirmé par de nombreuses simulations issues du modèle de réponses impulsionales stochastiques [6] (Section 2.5).

Ce Chapitre fait l'objet d'un article soumis au journal *Applied Acoustics* sous le titre : G.Defrance and J-D. Polack, *Estimating the mixing time of concert halls using the eXtensible Fourier Transform*.

---

### 6.2 Estimating the mixing time of concert halls using the eXtensible Fourier Transform

#### 6.2.1 Abstract

In room acoustics, we measure room impulse responses (RIRs) in order to fully describe the hall. RIRs are composed of a succession of arrivals (i.e., some sound rays which have undergone one or more reflections on their way from the source to the receiver). We propose the eXtensible Fourier Transform (XFT) in order to investigate the time evolution of spectral components of RIRs. The phase evolution versus time allows to estimate the mixing time, which is defined as

the time it takes for initially adjacent sound rays to spread uniformly across the room. After presenting some properties of the XFT, we show why one must compensate the natural energy decay of the RIR in order to obtain stationary signals. We estimate mixing times for a set of experimental RIRs and several that are synthesized using a stochastic model. Then, we estimate the dependence of mixing time upon the source/receiver distance in all these RIRs. Results are consistent up to the lack of reproducibility of the sound sources, but are strongly dependent on some parameters used for computing the XFT. We finally discuss the relevance of the name *mixing time* with respect to the theory and in regard to the time we estimate, that we propose to call *cross-over time*.

### **Keywords**

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

## **6.3 Introduction**

In room acoustics, the room impulse response (RIR) is the quantity that fully describes a hall. A RIR is measured for a couple of source and receiver positions. The pulse emitted by the source in all spatial directions encounters boundaries of the room and is divided into multiple wavelets. At each reflection (specular or diffuse), the pulse is filtered by the intrinsic properties of the materials that compose the room. Each wavelet follows a different trajectory in the hall. The RIR is composed of all these trajectories, called *arrivals*, that reach the receiver position after travelling through the room. An arrival is a sound ray that has undergone one or more reflections on the boundaries of the hall. The role of these arrivals, and their time distribution, is well-documented [12, 42] and is known to be directly linked to the "quality" of the room. After a certain time, called the mixing time, initially adjacent sound rays are spread uniformly across the room. Mixing is one property of dynamical systems (such as billiards), which are also ergodic and chaotic.

Polack has shown [6], based on Joyce's work [3], that the relevance of billiards theory to room acoustics can be fully established in the limit of high frequencies. Billiards can also be characterized by their entropy, which is directly linked to ergodicity. The entropy can be related to the logarithmic rate of divergence of the trajectories of sound particles, which is the Lyapunov characteristic number. Mixing billiards have strictly positive entropy and Lyapunov characteristic numbers [94] and are always ergodic. After the mixing time, events can be considered as statistically independent. Let us recall that mixing is a progressive process. The theory of ergodic dynamical systems [3] is able to justify some spectral observations in the frequency domain [14], namely the existence of discrete modes at low frequencies and modal superposition at high frequencies.

This article aims to illustrate similar findings in the time and frequency domains, namely a domain where reflections are singled out, shortly after the direct sound, followed at later times by superposed reflections distributed differently in time (Fig.6.1). This paper focuses on spectral statistics of experimental and synthesized RIRs. We investigate the use of a newly introduced tool, the eXtensible Fourier Transform[67], to estimate mixing time from a given RIR. The knowledge of this time can be particularly interesting for synthesis of RIRs.

This article is organized as follows. First, we review statistical properties of room acoustics (Section 6.4), in order to link the time distribution of arrivals to the mixing time. Then, we investigate mixing time estimation by using the eXtensible Fourier Transform (XFT). We formally study the properties of the XFT (Section 6.5), and explain how mixing times are then estimated (Section 6.7). We then estimate mixing times on a set of RIRs measured in Salle Pleyel [52] (Paris) using different sound sources (Section 6.8). It is shown that the estimation

of mixing time depends upon the bandwidth of the sound source, as predicted by the theory [73]. We compare our results to those computed on RIRs synthesized by a stochastic model [6] (Section 6.9). In Section 6.10, we study a few issues on the use of the XFT as an estimator of mixing time. Finally, differences between results from experimental and synthesized RIRs are discussed (Section 6.11).

## 6.4 Statistical properties of room acoustics

Weyl was the first to show [13] (see also [17]) that the density of eigenfrequencies in bounded rooms increases with the square of the frequency. In 1954, Schroeder experimentally investigated [14] this relationship and found that, at high frequencies, this is not the case. Instead, Schroeder observed a superposition of many modes at many frequencies, and a constant density of peaks due to the finite width of the modes created by losses. This investigation led to the determination of a transition frequency between the two regimes, the well-known Schroeder frequency [14], which links the volume of the hall to the decay time, i.e., to mode width. This implies that at high frequencies the normal modes of a room overlap in the frequency domain, i.e., the average separation between normal frequencies is smaller than the bandwidth  $\Delta f$  of a mode. Thus, any source signal will excite several modes at high frequencies simultaneously.

In the time domain too, there is a transition time between two different behaviours of the RIR : an initial time domain where arrivals are discrete (early reflections), and a late time domain where arrivals are superposed and therefore statistically equidistributed (commonly called late reverberation). This is called the mixing time. Extending Weyl's theory to the time domain leads to investigate the temporal succession of arrivals from trajectories linking the source to the receiver and taking into account reflections at the boundaries.

As described by Joyce [3], room acoustics can be considered as some dynamical system, called billiards. The latter possess statistical properties such as ergodicity and mixing (and chaos, but the latter is not investigated in this paper). In ergodic billiards, a sound ray theoretically takes nearly all positions and all directions of the phase space. Ergodicity ensures that the sound particle spends equal time in the vicinity of each position and of each direction [3]. A counter example is the parallelopipedic billiard, which is not ergodic. In that case, the sound particle can only take eight different directions. Billiards can also be characterized by their entropy, which is directly linked to ergodicity. The entropy can be related to the logarithmic rate of divergence of the trajectories of sound particles, which is the Lyapunov characteristic number [94]. Mixing billiards have strictly positive entropy and Lyapunov characteristic numbers. Indeed, Krylov's theory of mixing [8] teaches us that most dynamical systems gradually lose 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. 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., its acoustical energy density and the direction of the intensity are uniformly distributed across the room.

However, arrival density is not constant at early times and increases with time. A simple theory derived for rectangular rooms [71], teaches us that the number of arrivals per second  $D_e(t)$  (Eq.(6.2)) is proportional to the square of the time elapsed since the sound was emitted

by the source.

$$N(t) = 4\pi c_0^3 \frac{t^3}{3V} \quad (6.1)$$

$$D_e(t) = \frac{dN}{dt} = 4\pi c_0^3 \frac{t^2}{V}, \quad (6.2)$$

where  $D_e(t)$  is in number of arrivals per second,  $N$  is the number of arrivals up to time  $t$ ,  $c_0$  is the speed of sound in  $m.s^{-1}$ , and  $V$  is the volume of the room in  $m^3$ . Note that, in contrast, we calculate room acoustical indices from the time of arrival of the direct sound [15].

A somewhat different definition of mixing time is proposed by Polack [9], which is linked to the resolution of the auditory system instead of the characteristic duration of arrivals. According to this definition, the mixing time is the earliest time when  $N = 10$  reflections overlap within the characteristic time resolution of the auditory system, taken equal to  $\Delta t = 24$  ms (as proposed by Cremer [12]). Then Eq.(6.1) leads to :

$$T_m \approx 0.16 \times 10^{-3} \sqrt{\frac{V}{\Delta t}} \quad (6.3)$$

$$\approx \sqrt{V}, \quad (6.4)$$

where  $T_m$  is the mixing time, expressed in  $ms$  in Eq.(6.4),  $\Delta t$  is the characteristic time resolution of the auditory system, and  $V$  is the volume of the room in  $m^3$ .

In his book, Cremer mentions the existence of a time over which arrivals, being statistically distributed, cannot discriminated by the human ear ( $N = 2000$  arrivals received within  $\Delta t = 1$  s) :

$$t_C \approx 2\sqrt{V}, \quad (6.5)$$

where  $t_C$  is expressed in  $ms$  and  $V$  in  $m^3$ .

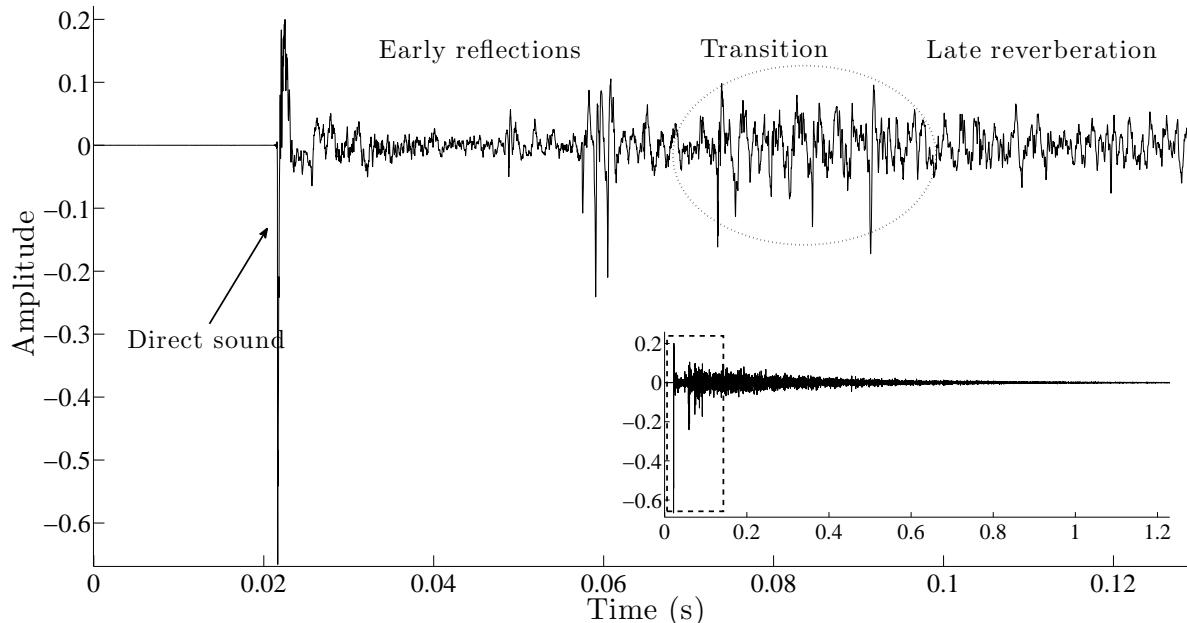


FIG. 6.1 – Detail of the early times of an experimental room impulse response. The encapsulated plot corresponds to the full record length RIR.

Eq.(6.4) is proposed as a reasonable approximation for the transition time between early reflections and late reverberation (Fig.6.1) [9]. One can define the late reverberation as an exponentially decaying noise, that is, as the superposition of pulses at a constant rate. It is shown [9, 27] that the exponentially decaying stochastic noise [72] is compatible with geometrical acoustics and billiard theory. As the mixing character of a room depends on its geometry and on the diffusing properties of its boundaries, we consider the value  $2\sqrt{V}$  to be an upper bound on the mixing time in mixing rooms [27, 73]. In the following, we concentrate on the experimental determination of the mixing time based on an analysis of the statistics in the time and frequency domains of the signal of the RIR.

## 6.5 The eXtensible Fourier Transform

We introduced the eXtensible Fourier Transform (XFT) [67] in order to estimate the mixing time of RIRs. We showed that using some Gaussianity estimators and sliding windows is not relevant and robust for estimating the transition time between early reflections and late reverberation, since results are strongly dependent on the size of the window of analysis [67, 66]. The latter report led us to use instead an extensible window, which allows accounting for progressive temporal/spectral changes in the signal given a fixed time origin. In the following, we present the principle, the definition and a few properties of the XFT.

### 6.5.1 Principle and definition

The XFT is comparable to a memory process. For instance, a human being is able to comprehend timbral changes thanks to his/her auditory memory, which is useful for detecting sound events in the time and frequency domains. The XFT is based on the same principle, that is, the time of occurrence of events (such as arrivals) can be detected compared to what happened in the past (the time of occurrence of the direct sound of a RIR, for instance). The XFT consists of computing the Discrete Fourier Transform of a signal using a window that grows from a given starting time (Fig.6.2). For a discrete signal  $x(n)$  of length  $N$  samples, its XFT is defined as :

$$X(p, f) = \sum_{n=0}^{N-1} w\left(\frac{n}{N} \frac{M}{p}\right) x(n) e^{-2j\pi n f/N}, \quad (6.6)$$

with  $w(t)$  the analysis window, which is a continuous function of time :

$$\text{supp}(w(t)) = [0, 1], \quad (6.7)$$

where  $X(p, f)$  is the complex matrix of the XFT,  $x(n)$  a signal of  $N$  samples length,  $w$  the analysis window that grows with a constant rate while  $p$  increases (the first window has a width  $L_0$ ),  $f$  is the discrete frequency in Hertz ( $f \in \mathbb{Z}$ ),  $M$  is the maximum number of windows  $w$  that the signal  $x$  can admit ( $M = N/L_0$ ),  $p$  is the number of windows ( $p = 1, 2, \dots, M$ ), and  $\text{supp}(w(t))$  is the support of the function  $w(t)$ .

### 6.5.2 Properties of the XFT

This subsection presents a few properties, related to the present study of the XFT from a signal processing point of view. Considering Eq.(6.6), it is obvious that at each step  $p$  of the calculation, the XFT is a summation of the information of the signal considered in the previous

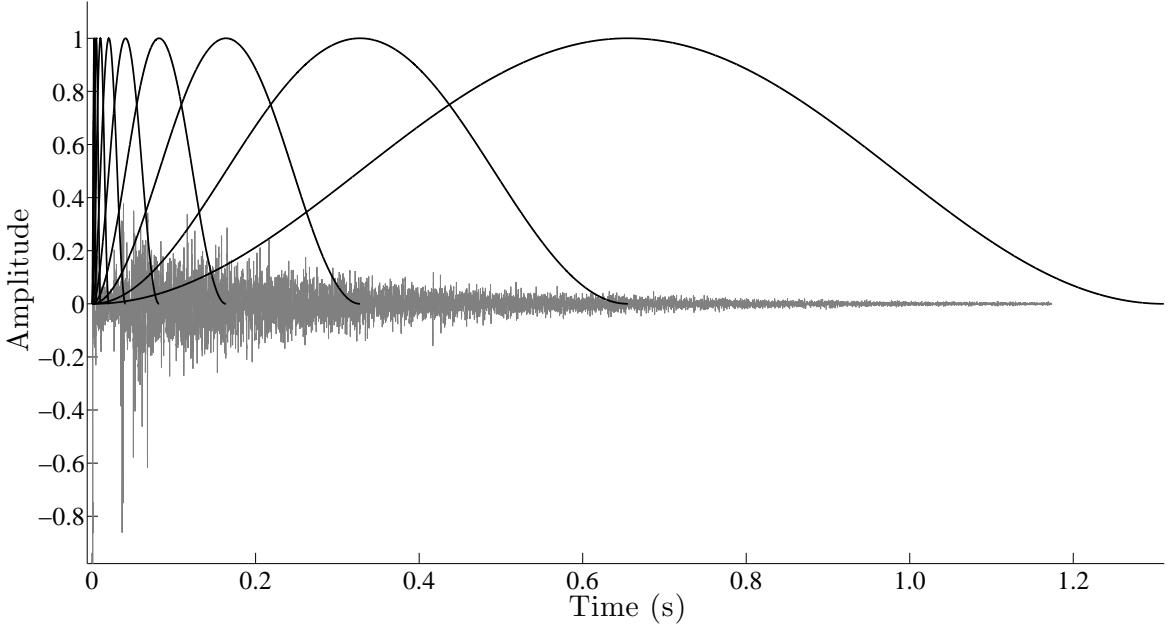


FIG. 6.2 – Scheme of the principle of the XFT. For each window, a Discrete Fourier Transform of the signal  $x$  is computed. In that case, the window  $w$  is a Hann window, which starts from the time of arrival of the direct sound. The latter is detected as described in [55].

steps. In other words, if  $p = 3$ , we can show that for a rectangular window  $w(n)$  :

$$X(3, f) = \sum_{n=0}^{3L_0-1} w\left(\frac{n}{3L_0}\right) x(n) e^{-2j\pi fn/N} \quad (6.8)$$

$$\begin{aligned} &= \sum_{n=0}^{L_0-1} w\left(\frac{n}{3L_0}\right) x(n) e^{-2j\pi fn/N} + \sum_{n=L_0}^{2L_0-1} w\left(\frac{n}{3L_0}\right) x(n) e^{-2j\pi fn/N} \\ &\quad + \sum_{n=2L_0}^{3L_0-1} w\left(\frac{n}{3L_0}\right) x(n) e^{-2j\pi fn/N}. \end{aligned} \quad (6.9)$$

Generalizing Eq.(6.9) :

$$X(p, f) = \sum_{m=1}^p \sum_{n=(m-1)L_0}^{mL_0-1} w\left(\frac{n}{pL_0}\right) x(n) e^{-2j\pi fn/N}. \quad (6.10)$$

Now, if we extend Eq.(6.10) to any type of window, we obtain :

$$X(p, f) = \sum_{m=1}^p \sum_{n=0}^{L_0-1} w\left(\frac{(m-1)L_0+n}{pL_0}\right) x(n) e^{-2j\pi fn/N}. \quad (6.11)$$

Eq.(6.10) and (6.11) emphasize the fact that the XFT is a sum in the time domain of spectral components of the signal. Furthermore, the XFT can be seen as a DFT computed at different times of the signal but keeping the phaseshift information.

In practice, the first window  $w$  should be chosen as small as possible so that the XFT has the best time resolution. However, this choice is also very dependent on the total number of

samples of the signal  $x$ , and therefore on the performances of the computer. It should be noted that, at each time step, the XFT is calculated for the same total number of samples (using zero padding), in order to keep the same nominal frequency resolution for each window. The number of samples ( $N_{DFT}$ ) for computing each Discrete Fourier Transform is calculated as the highest power of 2 :  $N_{DFT} = 2^{\lceil (\log(N)/\log(2)) \rceil}$ , where  $\lceil . \rceil$  is the ceiling symbol, and  $N$  is the number of samples of the signal  $x$ . For instance, for  $N = 91200$  samples,  $N_{DFT} = 2^{17} = 131072$  samples and using  $M = N/L_0 = 713$  windows  $w$  ( $L_0 = 128$  samples), the computational time is approximately of 20 seconds, on an Apple MacBook Pro (v.2.2) (1 single processor of 2.16 GHz, 3 GB of memory).

## 6.6 Application to an audio signal

Fig.6.3 shows the XFT of an experimental RIR, both its magnitude and its unwrapped phase. In this case, the XFT is computed using  $L_0 = 128$  samples and a rectangular window. We discuss later the choice of window. There are two distinct ways of studying the matrix  $X(p, f)$  : its magnitude and its phase. As predicted by Eq.(6.11), the energy of the signal (Fig.6.3 -middle) is an increasing function of time. Thus, the XFT does not allow to detect directly the offset of a signal. But on the contrary, it seems to be able to estimate the onset of an audio event. Interpretation of the magnitude of the XFT is not obvious though. However, visualizing its time evolution shows the progressive establishment of modes in the room. This is not relevant for mixing time, although it is related to Schroeder's work [14]. The phases of the XFT are easier to interpret than its magnitude.

Let us consider a simple model of a RIR defined as a sum of scaled Dirac impulses :

$$x_K(t) = \sum_{k=0}^{K-1} A_k \delta(t - t_k), \quad (6.12)$$

where  $t$  is discrete,  $A_k$  is an amplitude coefficient related to the absorption of the room (and to the air absorption),  $\delta(t - t_k)$  is a train of impulses, and  $t_k$  is the time of occurrence of arrivals regarded as independent random variables. Note that  $t_{k+1} - t_k$  is a decreasing function of time that tends towards 0 after the mixing time [12, 6], when arrivals overlap.

Let  $X(f)$  be the Discrete Fourier Transform of  $x_K(t)$  :

$$X(f) = \sum_{n=0}^{N-1} x_K(t)e^{-j2\pi ft/N} = \sum_{k=0}^{K-1} A_k e^{-j2\pi ft_k/N}, \quad (6.13)$$

where  $f$  is the discrete frequency.

Now the phase of  $X(f)$  is defined by :

$$\Phi(f) = \text{atan} \left( -\frac{\Im(X(f))}{\Re(X(f))} \right) \quad (6.14)$$

$$= \text{atan} \left( \frac{\sum_{k=0}^{K-1} A_k \sin(2\pi ft_k/N)}{\sum_{k=0}^{K-1} A_k \cos(2\pi ft_k/N)} \right). \quad (6.15)$$

Eq.(6.15) shows a well-known result of DFT, that is, the estimation of  $\Phi(f)$  depends on the amplitude coefficients  $A_k$ , which are decreasing with time. Therefore, in order to give the same weight to all arrivals [96] (especially for the ones at large times), the energy decay of the RIR needs to be compensated by applying the inverse of  $A_k$ , that is :  $1/e^{-\alpha k}$ , where  $\alpha$  is related to

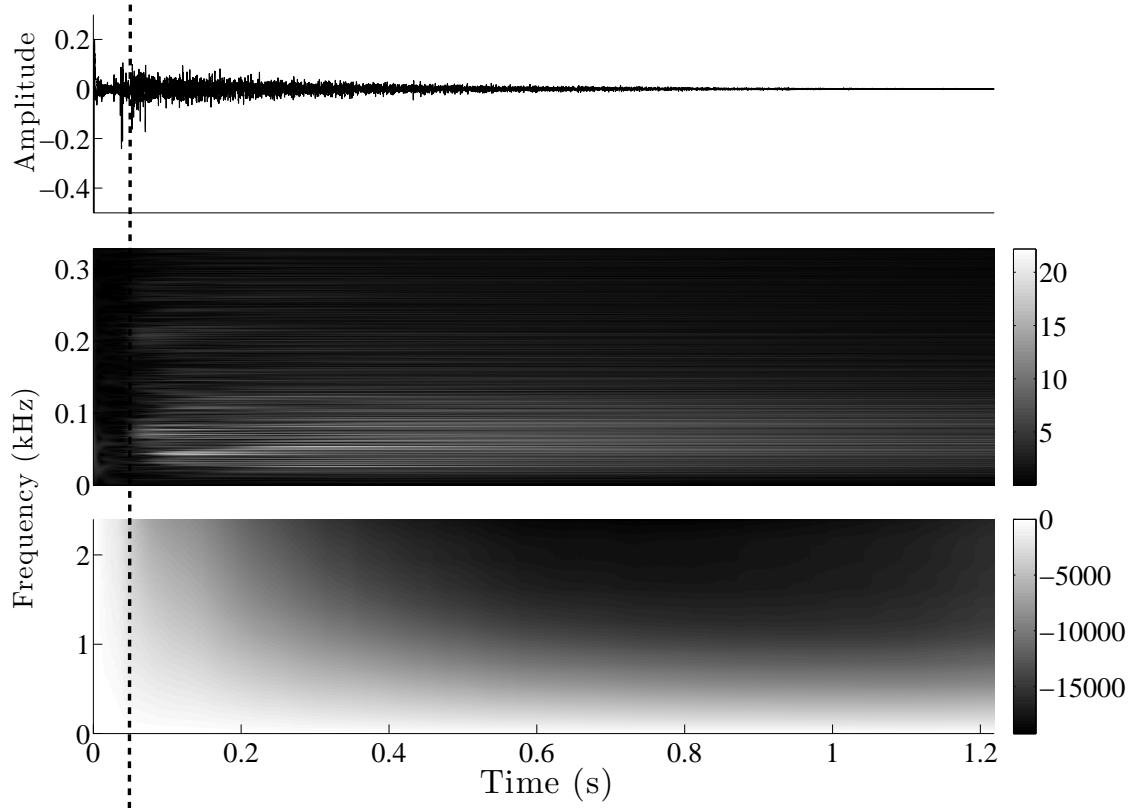


FIG. 6.3 – XFT applied to an experimental room impulse response signal for which the energy decay is compensated (but not shown here for an easier readability) (top). Middle : magnitude of spectra (linear scale). Bottom : unwrapped phases of spectra (linear scale in  $n\pi$ ). The dashed line marks a noticeable change in the statistics of the signal. Note the different frequency scales (middle and bottom graphs). The time of propagation between the source and the receiver is not shown here. The XFT is calculated from the time of occurrence of the direct sound. The RIR is the same as Fig.6.2.

the absorption of the hall and to the air attenuation [12]. Therefore, when compensating the energy decay, the phase  $\Phi(f)$  becomes :

$$\Phi(f) = \text{atan} \left( \frac{\sum_{k=0}^{K-1} \sin(2\pi ft_k/N)}{\sum_{k=0}^{K-1} \cos(2\pi ft_k/N)} \right). \quad (6.16)$$

The compensation of energy decay leads us to use a rectangular window  $w$  (see Eq.6.6) in order to avoid any other modifications of the amplitude of the signal  $x$ . However, using a rectangular window changes the spectral estimation of  $x_K(t)$  [97] since the window is not centered on  $t = 0$ , but on  $t = N/2$ , a phase shift of  $N/2$  is introduced [97]. Let us mention that cosine windows (such as Hann, Hamming, Bartlett, etc.) could also be used since they would compensate the exponential decay of the signal. In this case, the coefficients of the window should be chosen such that they compensate the energy decay based on the reverberation time.

## 6.7 Estimating mixing time of RIRs by calculating phase differences

### 6.7.1 Phase of the XFT

In this paper, the XFT is computed on RIRs for which the energy decay is compensated. Consider the simple model of RIRs given by Eq.(6.12) where the energy decay is compensated, and where  $K$  arrivals reach the receiver point during the time interval  $[0, T]$  :

$$x_K(t) = \sum_{k=0}^{K-1} \delta(t - t_k). \quad (6.17)$$

The coefficients in the Fourier series expansion of  $x_K(t)$  over the interval  $[0, T]$  are  $a_{nK}$  and  $b_{nK}$ , where :

$$a_{nK} - jb_{nK} = \frac{2}{T} \sum_{k=0}^{K-1} \int_0^T \delta(t - t_k) e^{-j\frac{2\pi nt}{T}} dt \quad (6.18)$$

$$= \frac{2}{T} \sum_{k=0}^{K-1} \int_0^T \delta(t) e^{-j\frac{2\pi n}{T}(t+t_k)} dt \quad (6.19)$$

$$= R_n e^{-j\phi_n} \sum_{k=0}^{K-1} e^{-jn\theta_k}. \quad (6.20)$$

In this expression :

$$\theta_k = \frac{2\pi t_k}{T} \quad (6.21)$$

$$R_n e^{-j\phi_n} = \frac{2}{T} \int_{-\infty}^{-\infty} \delta(t) e^{-j2\pi nt/T} dt = \frac{2}{T}. \quad (6.22)$$

As the times of arrivals  $(t_1, t_2, \dots, t_K)$  are regarded as  $K$  independent identically distributed random variable, each  $\theta_k$  may be regarded as a random variable distributed uniformly over  $[0, 2\pi]$ . From Eq.(6.20), one can see that there occurs the sum of  $K$  complex random variables, whose magnitude is unity. When  $K$  becomes very large, as it does when  $\nu \rightarrow \infty$  (where  $\nu$  is the number of arrivals per second), it is known [98] that the real and imaginary parts of Eq.(6.20) are also random variables, which tend to become independent and normally distributed about zero. This suggests the manner in which the normal distribution of coefficients arises [83]. Then, following from Eq.(6.20) (when averaging over the  $\theta_k$ ) :

$$\overline{a_{nK}} = \overline{b_{nK}} = 0 \quad (6.23)$$

$$\overline{a_{nK}^2} = \overline{b_{nK}^2} = R_n^2 \frac{K}{2} \quad (6.24)$$

$$\overline{a_{nK} b_{nK}} = \overline{a_{nK} a_{mK}} = \overline{b_{nK} b_{mK}} = 0, \quad (6.25)$$

for  $n \neq m$  and  $n, m > 0$ .

Now, instead of considering  $K$  arrivals in an interval of length  $T$ , we consider the probability of having any number of arrivals in the interval  $T$ , i.e.,  $T\nu$  arrivals. Then following from Eq.(6.20) :

$$\overline{a_n} = \overline{b_n} = 0 \quad (6.26)$$

$$\overline{a_n^2} = \overline{b_n^2} = R_n^2 \frac{\nu T}{2} = \sigma_n^2 \quad (6.27)$$

$$\overline{a_n b_n} = \overline{a_n a_m} = \overline{b_n b_m} = 0, \quad (6.28)$$

for  $n \neq m$  and  $n, m > 0$ .

In Eq.(6.27),  $\sigma_n$  denotes the standard deviation of the coefficients  $a_n$  and  $b_n$ . However,  $\sigma_n$  can be written differently. We introduce :

$$f_n = \frac{n}{T} = n\Delta f \quad (6.29)$$

$$\Delta f = \frac{1}{T}, \quad (6.30)$$

where  $f_n$  is the frequency of the  $n^{th}$  component.

Substituting Eq.(6.22) and (6.27) provides an expression for  $\sigma_n^2$  :

$$\sigma_n^2 = 2\nu\Delta f \left| \int_{-\infty}^{\infty} \delta(t) \exp[-j2\pi f_n t] dt \right|^2 = 2\nu\Delta f. \quad (6.31)$$

Thus, the square of the standard deviation is proportional to  $\nu/T$ .

Based on Eq.(6.6) and (6.16), the phase  $\Phi(p, f)$  of the XFT of  $x_K(t)$  can be written :

$$\Phi(p, f) = \text{atan} \left( \frac{\sum_{k=0}^{K(p)-1} \sin(2\pi f t_k / N)}{\sum_{k=0}^{K(p)-1} \cos(2\pi f t_k / N)} \right), \quad (6.32)$$

where  $K$ , the number of arrivals of the RIR occurring in each window  $p$  ( $p = 1, 2, \dots, M$ ), is an increasing function of  $p$ .

At first times (scale  $p$  is small), only a few arrivals are present in the signal. Thus, the slope of the phase of the XFT is expected to be proportional to the time delay of this group of arrivals. For instance, if  $p = 2$  and  $K = 2$ , then Eq.(6.32) becomes :

$$\Phi(2, f) = \text{atan} \left[ \frac{\sin(2\pi f t_1) + \sin(2\pi f t_2)}{\cos(2\pi f t_1) + \cos(2\pi f t_2)} \right]. \quad (6.33)$$

Fig.6.4 shows the arrival time distribution of a RIR synthesized by a stochastic model [6] (see Section 6.9.1 for details about the model). We plot the time elapsed between successive arrivals, as well as Eq.(6.32) unwrapped and computed at several  $p$ . For small  $p$  (Fig.6.4-c-d), the numerator and the denominator of Eq.(6.32) are sums of sines (and cosines) interfering constructively and/or destructively. As long as  $t_{k+1} - t_k \gg 0$ , the function  $\text{unwrap}(\Phi(p, f))$  is approximately a linear function of frequency  $f$ , which is more or less robust (depending on the number of sines summed and on the values of  $t_k$ ). While for larger  $p$  (Fig.6.4-e-h), if the number of arrivals  $K$  is large enough, then the coefficients of the Fourier series expansion become independent and normally distributed [98, 83]. We have not found in the literature any analytical estimation of the values of  $p$  and  $K$  such that the spectra of  $x_K(t)$  is normally distributed. Since this determination highly depends on the time distribution of arrivals that are stochastically distributed [99], this estimation might appear to be illusory. However, the XFT allows the experimental determination of the values of  $p$  and  $K$  based on the observation of the Fourier coefficients which are moving progressively from a deterministic to a random state. The XFT does not permit one to claim if a signal is normally distributed, but it does allow the visualization of the evolution of the system towards a random behaviour. The XFT can reveal the statistical evolution of a signal by looking at the time evolution of the phase of the XFT. We then estimate the mixing time as the time when we observe the coefficients (Eq.6.32) in the Fourier series expansion behave statistically.

### 6.7.2 Phase differences

The time evolution of the XFT phase embodies the changing statistics of a signal. We propose to calculate the sum of the differences between the unwrapped phases in the first window ( $p = 1$ ) and successive ones ( $p = 2, 3, 4, \dots, M$ ). The phase differences are calculated at each frequency  $f$  as follows :

$$\delta\Phi(p) = \sum_{f=0}^{N-1} |\text{unwrap}[\angle X(1, f)] - \text{unwrap}[\angle X(p, f)]|. \quad (6.34)$$

Eq.(6.34) can be seen as an estimate of the distance between the phases of  $X(1, f)$  and  $X(p, f)$ . Therefore, the smaller the quantity  $\delta\Phi(p)$ , the more similar are the phase spectra. Overall, Eq.(6.34) is an estimate of the time of occurrence of arrivals, regarding the time origin given by the first window ( $p = 1$ ). One may link this to the work of Defrance *et al.* [96, 33] in which the statistics of RIRs are investigated by looking at the cross-correlations in the time domain between the direct sound and the rest of the signal. The mixing time is found by detecting arrivals in the RIRs, and then by estimating the time at which two successive arrivals cannot be discriminated by the algorithm used. Note that in this previous work, the energy decay is also compensated for the reasons given in Section 6.6. Before the mixing time, a few arrivals were detected after the direct sound. It was shown that the number of arrivals is a cubic function of time, as predicted by the theory for small times (Eq.6.1). After the mixing time, the number of arrivals increases at an apparent constant rate with time. These previous studies showed that the statistics of a RIR are evolving from a deterministic to a stochastic state.

Fig.6.5 shows several graphs of the unwrapped phase of the XFT computed on an experimental RIR. As expected [12], the phase wraps faster between  $\pm\pi$  with time, denoting a gradual changing of behaviour in the signal, due for instance to an increasing number of events (such as arrivals and diffusion). This confirms findings from earlier investigations [96, 35, 33]. Estimating the mixing time is thus equivalent to estimating the time at which an important change of behaviour occurs in the phases.

As seen in Fig.6.6 and 6.7,  $\delta\Phi(p)$  is an increasing function of time, up to a certain time, and comprises one main inflection point. This is a breaking point in the statistics of the signal, and separates two states of behaviour of the impulse response. The first state is deterministic, that is, arrivals can be discriminated, and their time of occurrence can be predicted (Section 6.4). The function that best fits this part of the curve  $\delta\Phi(p)$  is a cubic function of time (Fig.6.6), which confirms Eq.(6.1). The second state of behaviour of the curve  $\delta\Phi(p)$ , which has a constant slope until a certain time, denotes a statistical behaviour : arrivals are statistically distributed in time.

### 6.7.3 Estimating of the mixing time in practice

For a proper estimate of the mixing time (Section 6.5.1), that is, in order to include the direct sound into the first window, the onset of the RIR has to be detected. We achieve this by using a method based on the spectral energy of the signal [55]. Obviously, an erroneous detection of the onset of the RIR will give an erroneous estimate of the mixing time. This issue will be investigated in a future work. The choice of the width  $L_0$  of the analysis window ( $w$ ) arises from our estimation of sound source durations. Choosing a first window width of the same size as the direct sound is in agreement with the method of mixing time estimation presented above (Sections 6.7.1 and 6.7.2). Note that the method used for estimating the direct sound duration has been published recently [47].

We propose to estimate the mixing time by searching the time of occurrence of the maximum of the curve  $\delta^2\Phi(p)$ , which is defined as :

$$\delta^2\Phi(p) = \delta\Phi(p+1) - \delta\Phi(p) \quad (6.35)$$

$$T_m = argmax_p(\delta^2\Phi(p)) * L_0/f_s \quad (6.36)$$

where  $p = 1, 2, \dots, M-1$ ,  $L_0$  is the width of the first window  $w$  of analysis in samples,  $f_s$  is the sampling frequency in  $Hz$ , and  $T_m$  is the mixing time in second.

We detect  $T_m$  using Cremer's criteria assuming that the mixing would not occur after that time. From Eq.(6.5), we find that  $t_C \approx 275 ms$  when  $V = 19000 m^3$ . This choice avoids to detect any breaking points in the statistics of the signal that can also occur later than  $T_m$  (Fig.6.7). Let us recall that mixing time takes into account the propagation time between the source and the receiver positions in the room.

## 6.8 Estimating the mixing time of a real concert hall

Mixing times have been estimated from 42 RIRs measured in Salle Pleyel [52], using pistol shots, balloon bursts and cracker impulses as sound sources. For each type of source, the first analysis window has a width  $L_0$  corresponding to the mean duration of the source impulse [47] (Table 6.1). Mixing times are presented versus the source/receiver distance. First, the mixing times of 21 impulse responses are estimated (Fig.6.8). For this set of measurements, pistol shots were used as sound sources.

Source	Mean duration $d$ (ms)	$L_0$ (samples)
Cracker impulse	2.0	128
Pistol shot	2.2	128
Balloon burst	3.2	256

TAB. 6.1 – Sound sources versus their mean duration  $\tilde{d}$  (ms), and the corresponding width of the first window ( $L_0$ ) in samples at a sampling frequency of  $48 kHz$  [47].

Large variations of mixing times (average=  $91 ms$ ,  $\sigma = 30\%$ ) are in contradiction with the ergodic theory of room acoustics, which predicts that the mixing occurs at the same time in the whole room. Surprisingly, mixing times can be well-described by an increasing linear function of the source/receiver distance. The relationship is given by :

$$T_m = 0.0027 \times d + 0.044 \quad (6.37)$$

$$r = 0.82 \quad (6.38)$$

where  $T_m$  is the mixing time in  $s$ ,  $d$  is the source/receiver distance in  $m$ ,  $r$  is the correlation coefficient, and  $1/0.0027 \approx 107\% c_0 \approx 370 m.s^{-1}$  (where  $c_0 = 345 m.s^{-1}$ ).

For some of the same source and receiver positions, two other sets of measurements have been carried out in Salle Pleyel, using balloon bursts and cracker impulses. Note that the width  $L_0$  of the first window used for calculating the XFT on RIRs measured with cracker impulses and with balloon bursts is not equal, but has a ratio of 2 (Table 6.1). Therefore, in order to keep the same time resolution, we calculate the XFT on RIRs measured with balloon bursts by using an increment window of  $1/2$  (while the increment window for RIRs measured by crackers or pistol shots is 1). Fig.6.8 shows mixing times estimated regarding the type of source used for the room acoustical measurements. The relationships that links mixing time to the source/receiver

distance are slightly different for each type sound sources, but remain consistent (Table 6.2). Moreover, Fig.6.8 shows that for some of the same source and receiver positions mixing times are consistent up to the type of source.

The spreading of the results is assumed to be linked to the reproducibility of each sound source. It is difficult to claim that the spreading of the results regarding the source is only linked to the temporal reproducibility of the impulse duration. The lack of reproducibility of the directivity of these sources can also be involved [100, 48, 50, 101]. Defrance *et. al* [47] show that the cracker impulse is the most reproducible, then comes the pistol shot and finally the balloon burst. Note that as the number of measurements is not equal for the three types of sources it is hard to compare the spreading of mixing times. Furthermore, the small number of measurements carried out using cracker impulses does not allow real comparison with the other sources, but only gives an overview of what can be obtained with a third type of sound source.

Source	Linear regression	r
Pistol shot	$T_m = 0.0027 \times d + 0.044$	0.82
Balloon burst	$T_m = 0.0023 \times d + 0.052$	0.81
Cracker impulse	$T_m = 0.0032 \times d + 0.027$	0.93

TAB. 6.2 – Sound sources versus the relationship between mixing times  $T_m$  (s) and the source/receiver distance  $d$  (m) in the room, where  $r$  is the correlation coefficient.

Finally, let us notice (Fig. 6.8) that a large number of mixing times estimated with the different sources for the same source and receiver positions are almost identical (mean variations are below 3%). As explained above, differences in results are due to the lack of reproducibility of the sources, and also to the estimation of a mean duration for each of them. Note in particular the value at 31 m equal to 157 ms (pistol shot), which is outside the cloud of data.

Consider the equation of the linear regression of Fig.6.8 :

$$T_m = a \times d + b \quad (6.39)$$

$$= 0.0026 \times d + 0.046. \quad (6.40)$$

As seen previously the slope coefficient of Eq.(6.40) is related to the speed of sound  $c_0$ , while the offset  $b$  is the time at which mixing occurs when the source and receiver positions coincide. The offset  $b$  corresponds to a distance of  $b \times c_0 \approx 15.5m$ , which is approximately the mean free path of the considered concert hall ( $\bar{l} = 14.6$  m). Recall that the mean free path is defined by Cremer [12] as : "[...] if some sound rays experience (on the average) the same number of reflections during the same time of observation, then they will have travelled the same mean free path."

## 6.9 Estimating the mixing time of synthesized concert halls

This section aims at verifying experimental observations from Section 6.8. This is achieved by using a stochastic model of RIRs [27, 7], described in Section 6.9.1, with which we synthesize a large number of RIRs for the same input parameters. This is inspired by the Monte Carlo methods, often used when simulating physical and mathematical systems [82]. Then, Section 6.9.2 presents relationships that exist between mixing times, source/receiver distances and the type of sources used for room acoustical measurements (Section 6.8).

### 6.9.1 Stochastic model of RIRs

The model used in this study has also been used in [27, 7, 96]. It assumes that the arrivals of the RIR are distributed in time according to a Poisson process, with a parameter (the mean number of arrivals Eq.6.1) that is time dependent. The set of arrivals is synthesized, as well as their respective amplitude, using the reverberation time ( $RT_{30}$  in second), the mean absorption ( $\tilde{\alpha}$  in percentage), the volume of the hall ( $V$  in  $m^3$ ), and the source/receiver distance (in meter) as input parameters. Input parameters of the model are those of Salle Pleyel [52] :  $RT_{30} = 1.9$  s,  $\tilde{\alpha} = 30\%$ ,  $V = 19000 m^3$ . This model has been validated [7] on a set of 8 American concert halls for which the room acoustical indices were recovered. The model allows the synthesis of diffusion, which is assumed to occur after the first arrival or reflection. Diffusion is synthesized as arrivals decaying exponentially, according to Morse and Feshbach [30]. Note that the model does not take into account the filtering of the room either. The impact of such simplifications is discussed in Section 6.11.

### 6.9.2 Estimation of the mixing time on synthesized RIRs

We synthesize thirty RIRs for several source/receiver distances varying from 5 to 44 m, by steps of 3 m. We consider four types of source : a Dirac, a balloon burst, a pistol shot, and finally a cracker impulse. We analyze with the XFT a total of  $4 \times 420 = 1680$  RIRs. We then estimate mixing times as presented in Section 6.7. We select the first window width  $L_0$  and the increment window by considering the source and the mean duration of the impulse (Table 6.1). We perform linear regressions between mixing times and distances and find mean mixing times calculated at each distance.

Mixing times are estimated on a set of RIRs synthesized (the source is a Dirac,  $L_0 = 4$  samples) with and without diffusion. Figure 6.9 shows mixing times versus the source/receiver distance. As seen in Fig.6.10, the presence of diffusion in the RIRs has a slight impact on the spreading of mixing times. This was expected since diffusion adds events of small amplitude in the early times of the signal. Therefore, mixing within RIRs synthesized with diffusion is expected to occur earlier than without diffusion. But on the other hand, the amplitude of diffusion is not important enough to have a strong impact on the phase. Fig.6.10 shows the mean values of mixing times (with and without diffusion) and their standard deviations. Note that mean variations in both cases are equal ( $\bar{\sigma} = 13.8\%$ ). Finally, the relationships linking mean mixing times to distance are very similar (Table 6.3). However, synthesized RIRs with diffusion are expected to be similar to measured RIRs. Consequently, in the following, RIRs are synthesized with diffusion.

RIRs synthesized	Linear regression	r
Without diffusion	$T_m = 0.0023 \times d + 0.039$	0.98
With diffusion	$T_m = 0.0024 \times d + 0.038$	0.98

TAB. 6.3 – Relationships between mixing times  $T_m$  (s) and the source/receiver distance  $d$  (m) in the room, where  $r$  is the correlation coefficient when diffusion is synthesized and when it is not.

Now, consider mixing times estimated with the different sound sources (Fig.6.9-6.13). Results show a strong dependency of mixing times on the source/receiver distance. The spread of mixing times diminishes when distance becomes large. This can be explained thanks to the time delay between the direct sound and the first arrival of the RIR, the so called ITDG (Initial Time Delay Gap) [42]. This time delay 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.

Differences of mixing times between sound sources can be explained as follows. Not only is the time of occurrence of arrivals important for determining the mixing time, but also the time delays between successive arrivals. We observe that when two successive arrivals of the model are delayed by a duration inferior or equal to the duration of the source impulse, the XFT considers not two but one arrival (which is thus filtered). Consequently, the more arrivals overlap, the more arrivals we may need to take into account in a window to determine the mixing time. Therefore, the longer the source impulse duration, the larger the estimated mixing time can be.

Finally, the offset of mixing times estimated for all sources is constant and equals  $0.038\text{ s}$ , which is approximately the mean free path ( $13\text{ m}$ ) of the considered hall estimated by the model based on the input parameters (Section 6.9.1) (the mean free path of Salle Pleyel equals  $14.6\text{ m}$ ). This means that mixing would occur (if the source and receiver positions are the same and not located on the boundaries of the hall) once sound rays have travelled the mean statistical path and undergone one reflection. The comparison of the hall radius of Salle Pleyel [12] ( $r_0 = 6.7\text{ m}$ ) and the offset of mixing time of Fig.6.9 gives a ratio of two. This might be a coincidence, since there is not any obvious relationship between the hall radius and the mean free path.<sup>1</sup>

## 6.10 Variability of the XFT

This section aims at estimating the variability of the mixing time when the background noise level of RIRs varies, as well as when the initial analysis window width changes. This is achieved by considering 21 experimental RIRs measured with pistol shots as sound sources, and 420 synthesized RIRs convolved with one pistol shot. Results are presented using interquartiles and boxplots. Differences between results from synthesized and experimental RIRs are discussed in Section 6.11.

### 6.10.1 Variability with the background noise level

As expected from Section 6.7, the estimation of mixing time depends on the background noise level, since two random processes are added : the RIR and the background noise. In experimental RIRs, background noise is observed and is commonly due to air conditionning, spot lights, etc. In this section, we estimate the mixing time of synthesized and experimental RIRs for which the signal to noise ratio ( $SNR$ ) is varying from  $80\text{ dB}$  to  $10\text{ dB}$ . The reference mixing time is taken as the one estimated using  $SNR = 80\text{ dB}$ , based on the  $SNR$  measured in Salle Pleyel [52]. The noise, added before the energy decay compensation, is white Gaussian distributed.

Considering synthesized RIRs (Fig.6.14), variations in mixing time estimates, as measured by interquartiles and the median values, are negligible for  $SNR \geq 40\text{ dB}$ , and greater than a variation of 60% points for  $SNR < 40\text{ dB}$ . Considering interquartiles of experimental RIRs (Fig.6.15), variations of mixing times are negligible (below 2%) if  $SNR \geq 30\text{ dB}$ , while a variation of almost 30% points is reached for  $SNR < 30\text{ dB}$  (up to extreme values). Note that the density of extreme values (values standing outside the whiskers of boxplots) is larger for  $SNR \leq 30\text{ dB}$  than for  $SNR > 30\text{ dB}$ . Further, negative variations correspond to some mixing times estimated earlier than the reference time. This simple investigation emphasizes the importance of measuring some RIRs with a good signal to noise ratio (i.e., better than  $30\text{ dB}$ ), as recommended in standards [102].

---

<sup>1</sup>Further investigations on two other large rooms showed different ratios between the hall radius and the mean free path. Therefore, as long as we cannot conclude on that particular point, we have chosen not to posit such a relationship.

### 6.10.2 Variability with the analysis window width

Through the paper, different window widths have been used, depending on the sound source, and paying attention to keep the same time resolution for every RIRs (experimental and synthesized). This section aims at studying the impact of the window width on the reproducibility of the XFT.

Sixteen sizes of the first window width are considered, while the window width increment is kept to one. The window width reference ( $L_0$ ) is set to 128 samples, and the mixing times are estimated with window width varying from  $-85\%$  up to  $1500\%$  ( $\Delta L_0$ ). Results (Fig.6.16-6.17) are the variations in percentage of mixing times ( $\Delta T_m$ ) versus the variations ( $\Delta L_0$ ) in percentage of the first window width ( $L_0$ ). In the case of synthesized RIRs (Fig.6.16), variations do not exceed  $10\%$ , if  $\Delta L_0 < 100\%$ . This shows that the XFT is robust to variations of the analysis window width when ideal RIRs are considered (without background noise and filtering of the hall). While for experimental RIRs (Fig.6.17), variations are larger, even when  $\Delta L_0$  is small (the average is  $50\%$  points). Therefore, one can see the importance of using an analysis window of the size of the impulse duration. Obviously, taking a window width larger than the expected mixing time would not make any sense. Fig.6.18 shows the median values of the variations of mixing time estimated for some length of windows between varying between  $-85\%$  and  $100\%$ . One can see that median values of mixing times estimated from the model are close to experimental ones. Finally, note that negative variations correspond to some mixing times estimated earlier than the reference time.

## 6.11 Discussion

In Salle Pleyel, the volume is  $V = 19000 m^3$ , therefore the theoretical upper bound value of mixing time is :  $T_{M_{th}} = 138 ms$ , according to Eq.(6.4). Mean mixing times estimated from experimental RIRs ( $\bar{T}_m = 99 ms$ ) (considering all sources) on the one hand, and from synthesized RIRs ( $\bar{T}_m = 101 ms$ ) on the other hand, recover the theoretical assumption that states that  $T_{M_{th}} \approx \sqrt{V}$  (ms) is an upper bound of mixing time (Section 6.4).

Further, even if mixing times estimated from synthesized RIRs are quite close to experimental ones (Fig.6.11-6.13), there exists some differences between the model and the experiments that can be explained as follows. 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. Recall that all measurements are taken near a boundary, a particular situation that may significantly differ from the statistical average of the model. Further, the reproducibility of the sound sources has a well-known incidence on room acoustical indices [76]. This also explains the spreading of the results obtained with experimental RIRs. While there are as many impulses as there are RIRs measured with each sound source (Section 6.8), only one impulse of each source was used in the synthesized RIRs. This explains the small spreading of mean mixing times in that case.

Recall that the model does not take into account any kind of filtering of the room, while this well-known phenomenon is obviously found in experimental RIRs. This is of great importance since Eq.(6.34-6.36) estimate the time at which the direct sound is not spectrally correlated to the rest of the signal (Section 6.7.2). However, let us mention that the convolution of a sound source by a train of impulses statistically distributed in time leads to a filtering of the sound source, especially when two arrivals are delayed in time by a duration inferior or equal to the one of the source impulse. This is nothing compared to the filtering of the room, which is however assumed to be more complex due to the many different materials. Consequently, a next step into this research is to add filtering of the hall and to compare experimental results and synthesized ones. This could be achieved using a ray-tracing model of room impulse responses

[5]. However, the theoretical problem is that ray tracing does not display mixing as long as the reflecting surfaces are plane [103]. Furthermore, the introduction of diffusion coefficients affects the comparison by artificially introducing mixing.

In this article, the XFT is applied to impulsive RIRs, but it can also be applied to RIRs measured by means of MLS, sine sweep, etc. Obviously, for each type of signal, one has to know (or estimate) the duration of the source impulse after deconvolution. A benefit to mathematical signals is that they provide reproducible sound sources, which is not the case with gunshots, balloon bursts, etc [47, 50, 100, 101]. Furthermore, the XFT is applicable to any kind of spaces or media in which waves propagate. Using the phase differences defined in Eq.(6.34), one is able to compute an estimate of arrivals statistics. As a remark, we wish to emphasize that if one does not take into account the time of propagation between the source and the receiver positions, the mixing time cannot be defined as such. However, one may see here the possibility of defining a new room acoustical index as the time at which the mixing is achieved starting from the time of arrival of the direct sound.

The match between experiments and the model (mixing times and the recovery of the mean free path) hints to choose the XFT as an estimator of mixing time, and this stochastic model of RIRs, which has already been through a first step of validation based on its reliability to synthesize RIRs with the same acoustical indices than experimental ones [7]. Another estimator of mixing time within RIRs has been proposed [33], using the algorithm of Matching Pursuit. Though the comparison between the two techniques is out of scope for this article, we wish to mention that the results we obtain with both estimators are in agreement. However, none of these estimators can tell us what the "true" mixing time is. Therefore, neither Matching Pursuit nor the XFT can serve as references, but only a model of RIRs as the one used in this article.

We wish to emphasize that the time we estimate using the XFT and Eqs.(6.34)-(6.36) is not, strictly speaking, the mixing time. First, since mixing is an asymptotic process that must be observed over an infinite time, mixing time does not exist as such according to the theory of dynamical systems. Second, in Sections 6.3 and 6.4, we define the mixing time as a function of divergence rate of sound rays. While we do not focus on sound rays, we consider the statistical distribution of arrivals at a given position in the room. Therefore, it may be more appropriate to use the term *cross-over time* rather than mixing time, since we are estimating the time at which the system behaves randomly. The name "cross-over time" is given in reference to the "cross-over frequency" after Schroeder [14], which defines a transition frequency between eigenmodes which can be easily discriminated at low frequencies, and eigenmodes at high frequencies which have, in appearance, a constant density. The cross-over time defines a time transition within RIRs between early reflections and late reverberation. Since the relationship between that time and the mixing time is not experimentally shown yet, one cannot claim that the two times are identical.

Finally, the results we obtain are also a function of the detection method used. The cross-over time is estimated using Eq.(6.36), that is, by detecting the time at which the maximum of Eq.(6.35) is reached. In [67] and [66], it is shown that the signal of a RIR recorded in a large room can have a Gaussian distribution at several different times. That means that there are some moments even after the beginning of the late reverberation (that one can estimate by hand) where the signal is not randomly distributed. In that case, we observe what we call some *bursts of coherence*. Therefore, using Eq.(6.36), one can estimate several cross-over times. By restricting our estimation up to Cremer's criteria (Eq.6.5), we avoid to detect too many of these bursts of coherence. However, such an assumption is arbitrary, and future work would consist in finding a more robust detector of the breaking point in the statistics of the XFT. Moreover, it would be especially interesting to study the probability of appearance of these bursts of coherence and to understand the physics that is related.

## 6.12 Conclusion

This article investigates the use of the eXtensible Fourier Transform as an estimator of the cross-over time. After reviewing a few properties of the XFT, we explain, using a stochastic approach, how the estimation of this time is possible. This estimation is achieved by calculating phase differences of the XFT, with the phase of the direct sound as a reference. Results obtained with experimental RIRs are confirmed by those of synthesized RIRs. This shows the relevance of the XFT, and also of the stochastic model used in this article. Furthermore, cross-over times are estimated using different sound sources during measurements are consistent, up to the lack of reproducibility of the sources. Results show that cross-over time is a function of source/receiver distance. Differences between cross-over times of experimental and synthesized RIRs are assumed to be due to the filtering of the hall, which is not taken into account by the model. A next step into this research would consist in supplementing the model by integrating a random filtering of the room. We also note that results show that for identical source and receiver positions, the late reverberation starts once the sound has travelled the mean free path of the hall. However, this assumption would necessitate some experimental validations.

In this article, we also study the robustness of the XFT as an estimator of the cross-over time. We identify that size of the first analysis window is one of the major issue of this estimator. We propose using a window width equal to the length of the direct sound of the RIR in order to be in agreement with the principle of the XFT first, and second to avoid erroneous estimations. On the other hand, we also study the impact of the background noise level on the precision of estimation, and emphasizes the need to carry out measurements with respect to the standard (ISO 3382 [102]). The relevance of the estimation of the cross-over time upon the signal to noise ratio should also be studied using listening tests. Moreover, from a signal processing point of view, there remains to investigate the relevance of using the XFT compared to a simple spectrogram, for estimating spectral components of a signal, their time of occurrence, but also the impact of using different types of windows. The impact of using an extensible window that grows with a certain rate inferior to its initial size will be the topic of a future work.

Finally, we emphasize that the term “mixing time” might no be appropriate. Instead, we propose to use “cross-over time”, since the relationship between the transition time from early reflections to late reverberation and mixing has not been experimentally shown yet.

## 6.13 Acknowledgments

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

---

## 6.14 Développements ultérieurs

### 6.14.1 Comparaison avec Matching Pursuit

Nous avons présenté précédemment l'estimation du temps de mélange à l'aide deux méthodes : Matching Pursuit (MP) [33] et la XFT [106]. Nous nous proposons de comparer les résultats obtenus à l'aide de ces deux méthodes.

L'estimation du temps de mélange à l'aide de MP et de la XFT aboutissent à la même observation : le temps de mélange est une fonction linéaire croissante de la distance source/récepteur

(Table 6.4). Dans le cas des réponses impulsionales mesurées dans la Salle Pleyel, nous obtenons les relations suivantes :

Estimateur	Relation temps de mélange/distance
Matching Pursuit ( $SRR = 5dB$ )	$Tm = 0.0026 \times d + 0.026$ ( $r = 0.82$ )
XFT	$Tm = 0.0027 \times d + 0.044$ ( $r = 0.82$ )

TAB. 6.4 – Relation temps de mélange et distance source/récepteur pour chaque estimateur du temps de mélange sur les réponses impulsionales mesurées dans la Salle Pleyel [52].

Les deux estimateurs aboutissent sensiblement à la même relation (Table 6.4). Cependant, lorsque la position de la source et du récepteur sont confondus, MP estime qu'il faut  $26ms$  (soit environ  $8m$ ) avant que le mélange ne soit atteint, contre  $44ms^2$  pour la XFT (soit  $13m$ ). A la connaissance de l'auteur, aucune étude n'a encore été menée pour estimer le temps de mélange à la position d'émission de la source dans la salle. Il est donc impossible de juger du meilleur résultat. Remarquons que la XFT prévoit que le son parcourt une distance une demie fois plus longue que celle prédicta par MP avant que le mélange ne soit atteint dans cette configuration particulière.

La Fig. 6.19 présente les temps de mélange estimés pour la Salle Pleyel à l'aide deux estimateurs. Remarquons d'abord que certaines valeurs coïncident. Cependant, la différence moyenne entre les deux séries est de 20% et l'écart-type est de 80%. Les différences entre les résultats des deux estimateurs s'expliquent par les différents paramètres que l'on doit estimer et/ou choisir avant l'estimation. Ainsi, Matching Pursuit estime les retours de la réponse impulsionnelle à partir d'un dictionnaire d'atomes composé du son direct translaté dans le temps. Il est donc nécessaire au préalable de définir l'atome-mère du dictionnaire, autrement dit de connaître les limites temporelles du son direct [47], qui doivent être estimées dans le cas du revolver. D'autre part, le Chapitre 5 insiste sur le fait que le nombre de retours estimés -et donc le temps de mélange- dépend du critère d'arrêt de Matching Pursuit.

Estimer le temps de mélange à l'aide de la XFT présente également quelques étapes nécessaires dont dépend la valeur du temps de mélange : estimation de la durée du son direct et incrément temporel des fenêtres.

### 6.14.2 Ouvertures

Nous avons montré comment la XFT appliquée à une réponse impulsionnelle permet d'estimer le temps auquel son spectre devient statistiquement distribué. Dans ce travail de thèse, nous n'avons pas considéré d'autres domaines d'application de cet outil temps/fréquence. Cependant, il devrait pouvoir trouver son utilité dans toute application où l'on souhaite visualiser l'évolution temporelle des propriétés spectrales d'un système relativement à l'état initial de ce système.

---

<sup>2</sup>Soit presque deux fois plus longtemps.

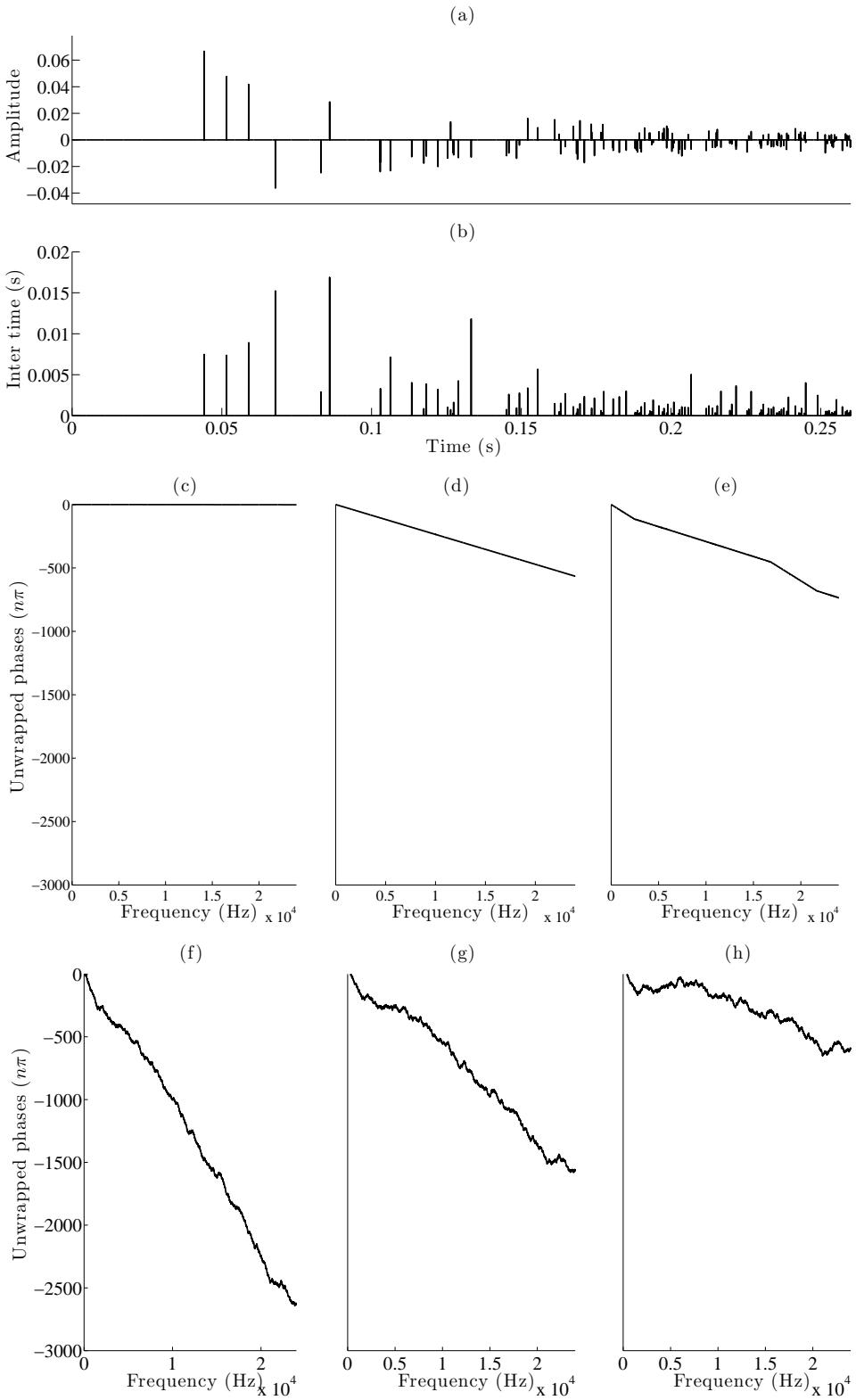


FIG. 6.4 – (a) : Zoom on the early times of a synthesized RIR (the energy compensation is applied, but not shown for an easier readability). (b) : Time elapsed between successive arrivals. (c)-(h) : Unwrapped phase at different times starting from the time of arrival of the direct sound. (c) :  $0 \leq t \leq 60$  ms. (d) :  $0 \leq t \leq 80$  ms. (e) :  $0 \leq t \leq 90$  ms. (f) :  $0 \leq t \leq 110$  ms. (g) :  $0 \leq t \leq 120$  ms. (h) :  $0 \leq t \leq 130$  ms. Note that phases scales are the same for each graph from (c) to (h).

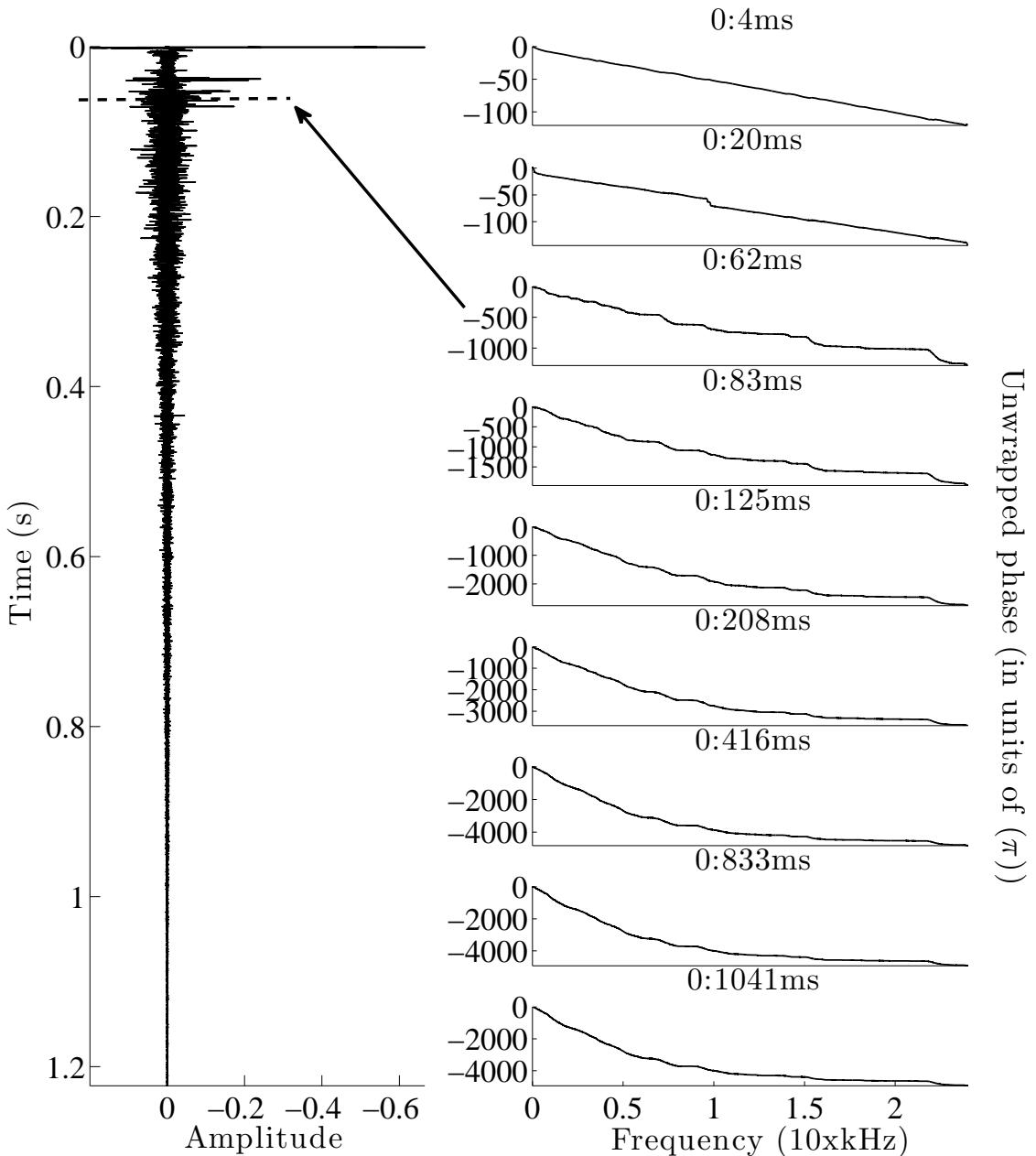


FIG. 6.5 – Visualization of the unwrapped phases of the XFT computed on a RIR measured in Salle Pleyel (source : pistol shot). Note that the compensation of the energy decay of the RIR is applied, but not shown for a better readability. Furthermore, the time of propagation between the source and the receiver positions is not taken into account. First window width :  $L_0 = 128$  samples. The estimated mixing time is materialized by a dashed line superposed to the RIR.

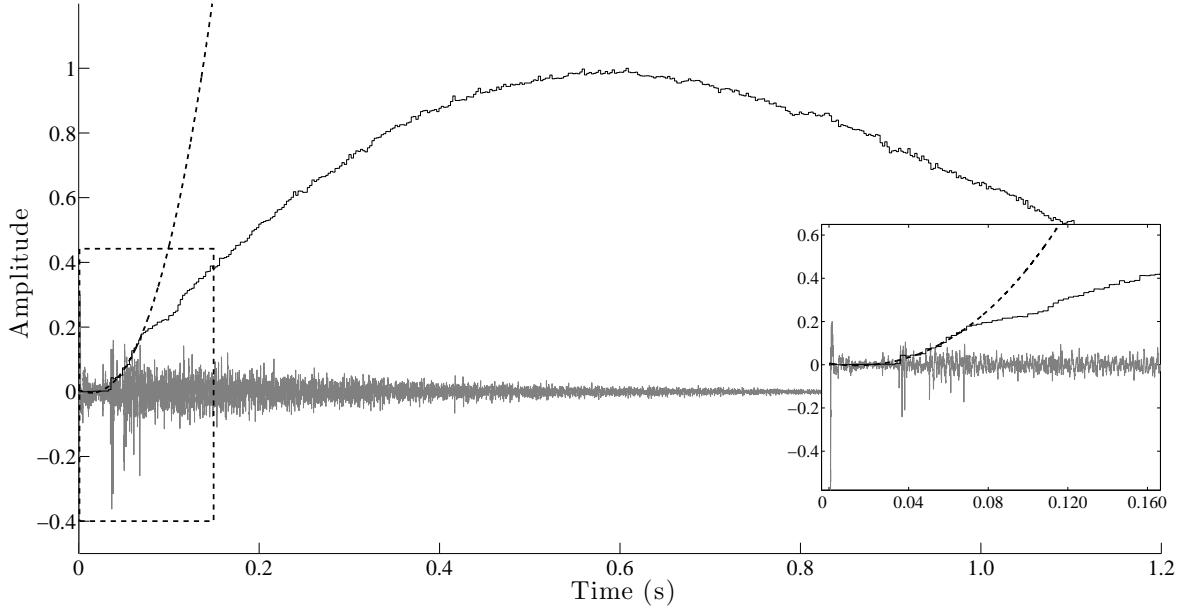


FIG. 6.6 – Grey : experimental RIR (the energy decay compensation is applied, but not shown for a better readability). Plain black line : curve  $\delta\Phi(p)$  (Eq.(6.34)) (normalized). Black bold dashed line : cubic fit. Note that  $\delta\Phi$  is over sampled for a better readability. The graphic included shows the details of the selected region within the dashed rectangle. The RIR is the same as Fig.6.2.

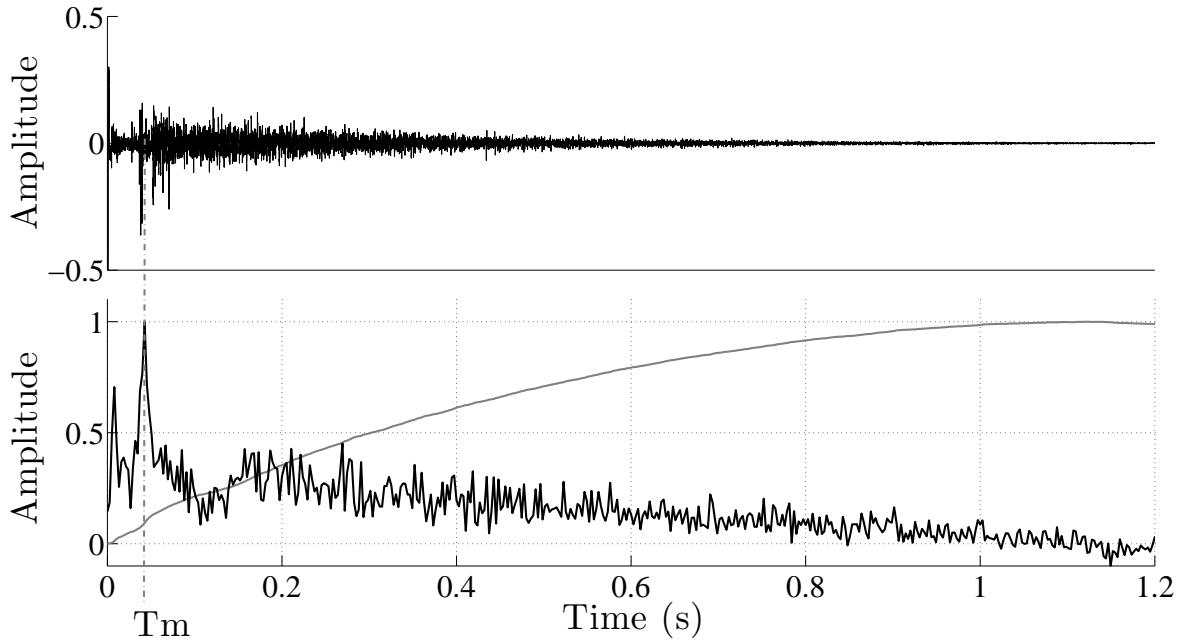


FIG. 6.7 – Top : an experimental RIR (the energy decay compensation is applied, but not shown for a better readability). Bottom : Eq.(6.34) in grey; Eq.(6.35) in black. The breaking point (i.e., the mixing time) in the statistics of the RIR is materialized by a dashed line. Note that bottom graphic is over sampled for a better readability. Notice the other beaking points after the mixing time (other peaks). The time of propagation between the source and the receiver positions is not shown here. On bottom graphic, curves are normalized for an easier readability. The RIR is the same as Fig.6.2.

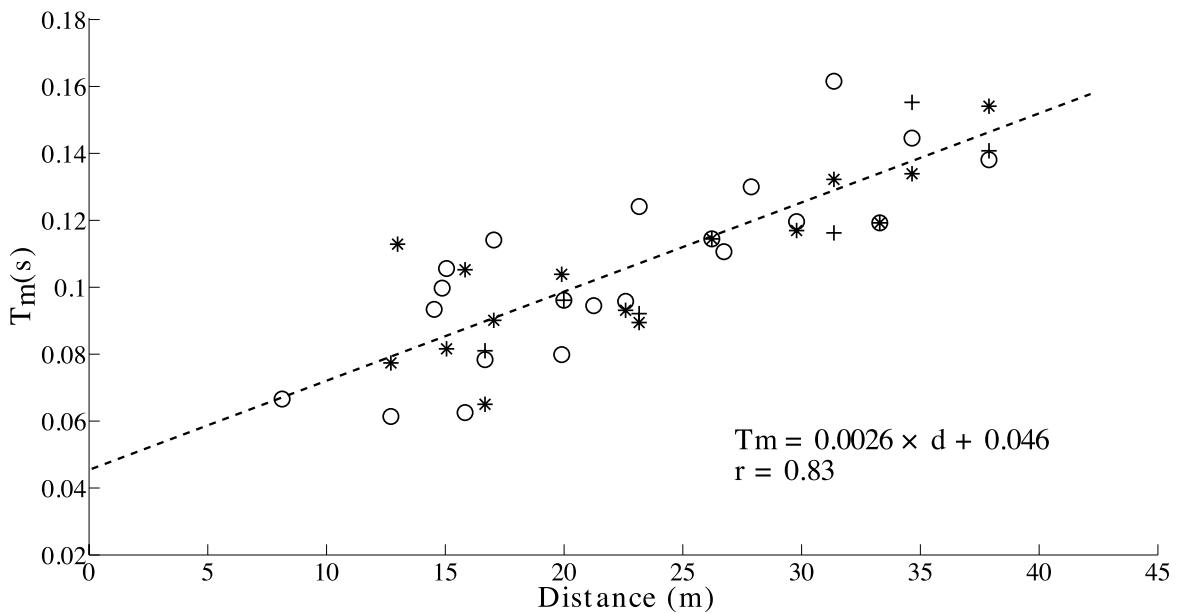


FIG. 6.8 – Estimated mixing times on 42 RIRs measured in Salle Pleyel (sources : pistol shots = circles, balloon bursts = stars, cracker impulses = crosses). Note the superposition of some data. Dashed line : linear regression calculated on mixing times for the three types of sources.

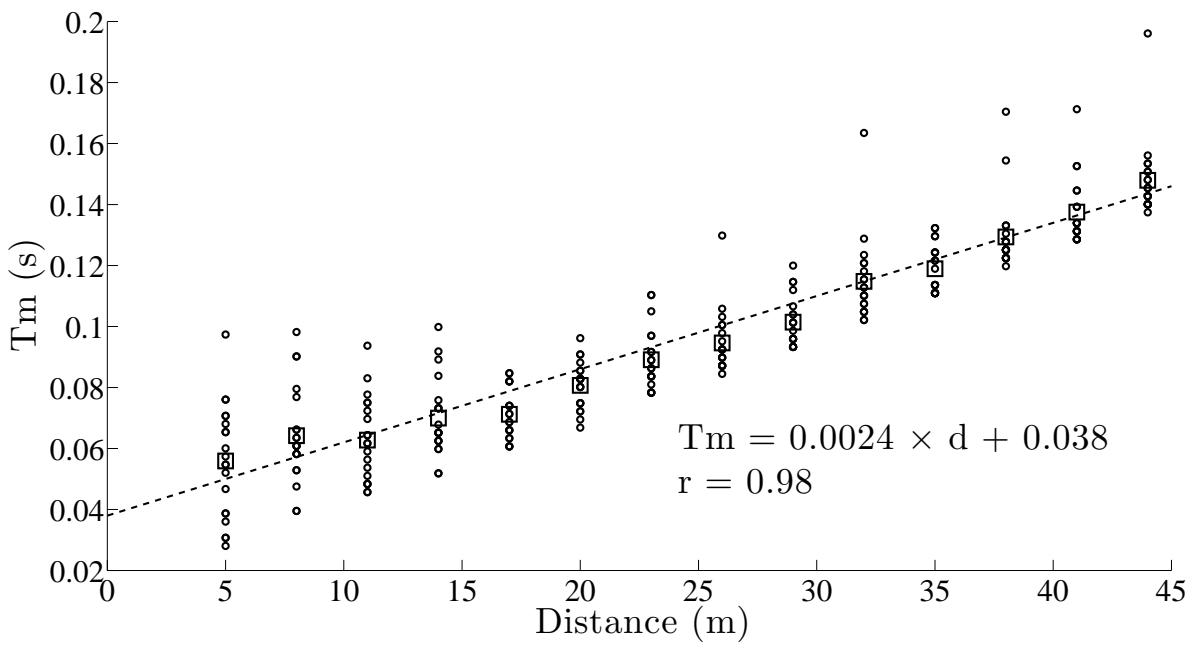


FIG. 6.9 – Dots : estimated mixing times on 420 synthesized RIRs for the Salle Pleyel (source : Dirac). First analysis window width :  $L_0 = 4$  samples. In that case, the model synthesizes diffusion. Squares : mean mixing times. Dashed line : linear regression calculated on mean mixing times.

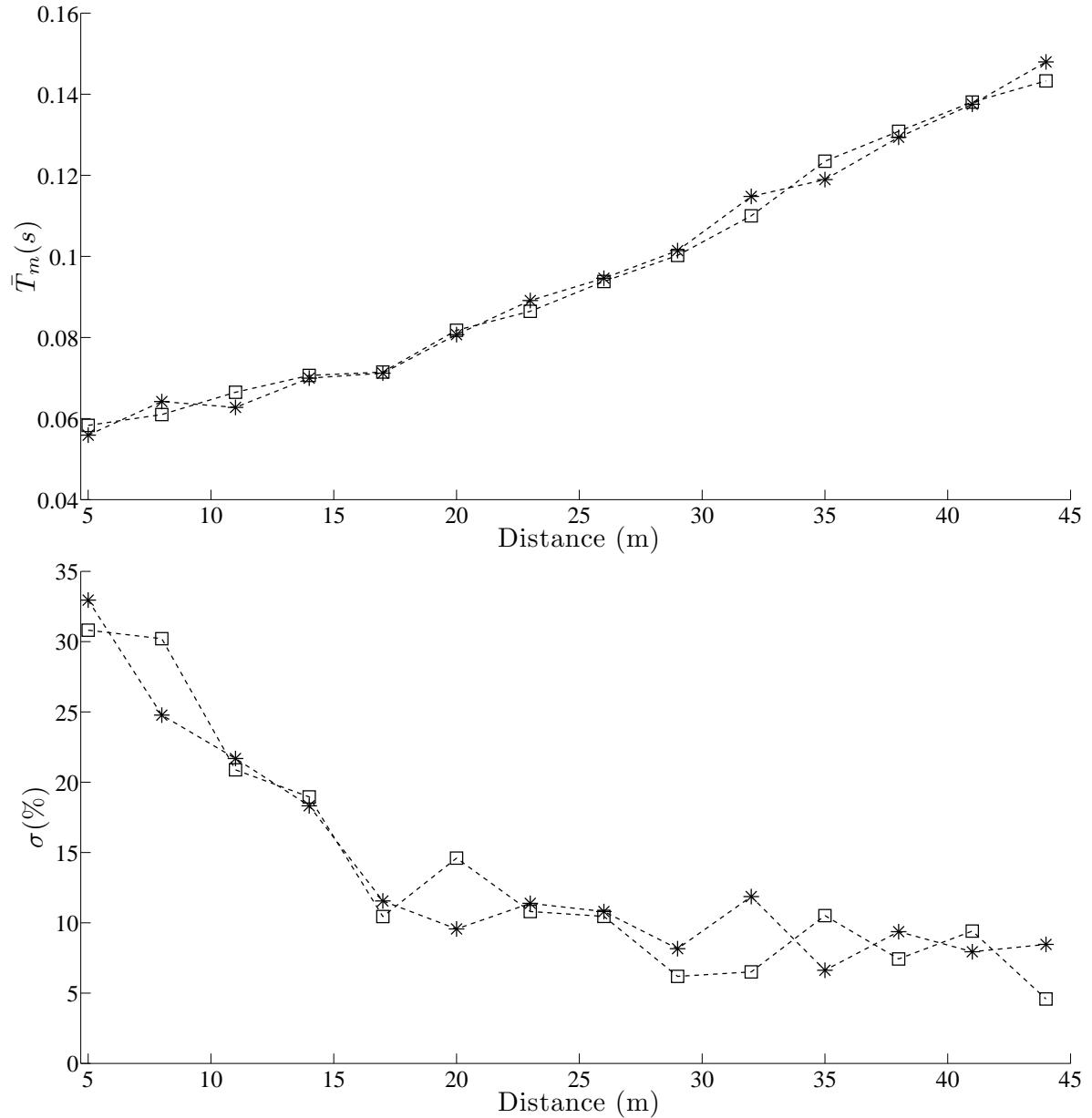


FIG. 6.10 – Top : mean mixing times  $\bar{T}_m$  (s) estimated on RIRs synthesized with diffusion (stars) and without diffusion (squares). Bottom : standard deviations  $\sigma$  (%) of mixing times estimated on synthesized RIRs with diffusion (stars) and without diffusion (squares). The source is a Dirac. First analysis window width :  $L_0 = 4$  samples.

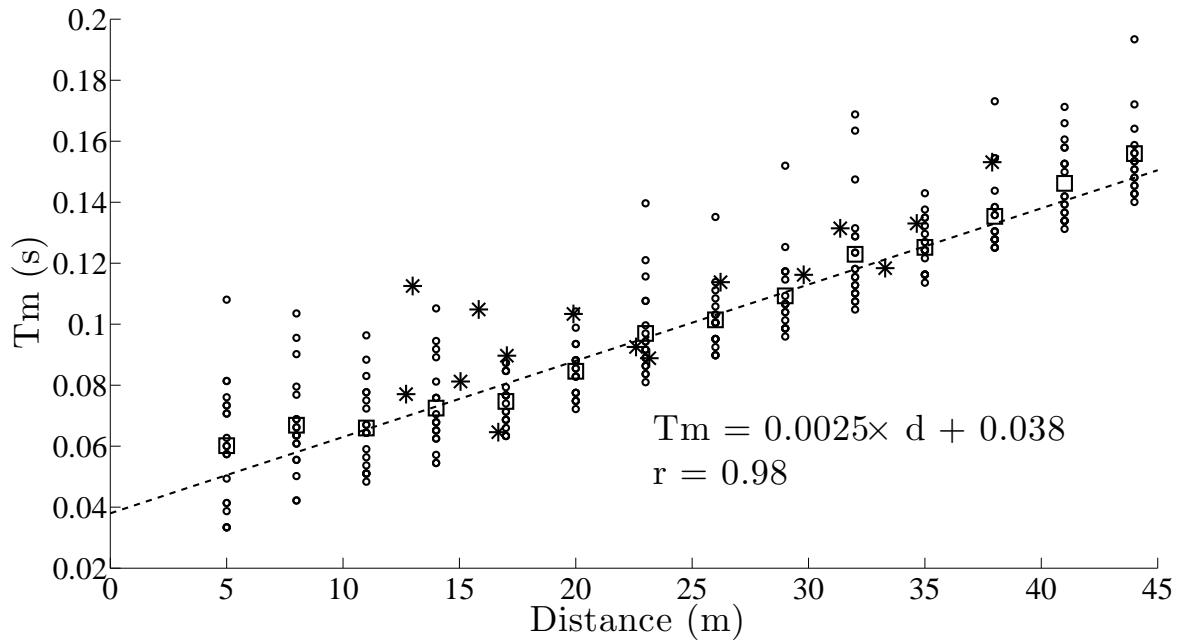


FIG. 6.11 – Dots : estimated mixing times on 420 synthesized RIRs for the Salle Pleyel (source : balloon burst). Stars : experimental mixing times (using balloon bursts). First window width :  $L_0 = 256$  samples. Squares : mean mixing times. Dashed line : linear regression calculated on mean mixing times.

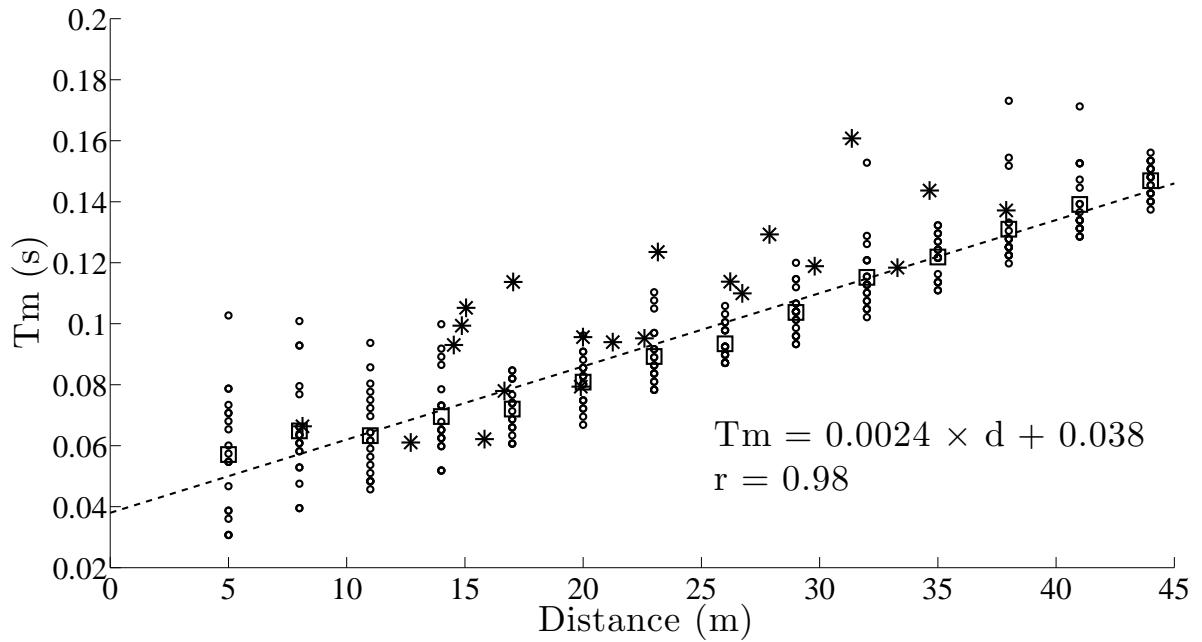


FIG. 6.12 – Dots : estimated mixing times on 420 synthesized impulse responses for the Salle Pleyel (source : pistol shot). Stars : experimental mixing times (using pistol shots). First window width :  $L_0 = 128$  samples. Squares : mean mixing times. Dashed line : linear regression calculated on mean mixing times.

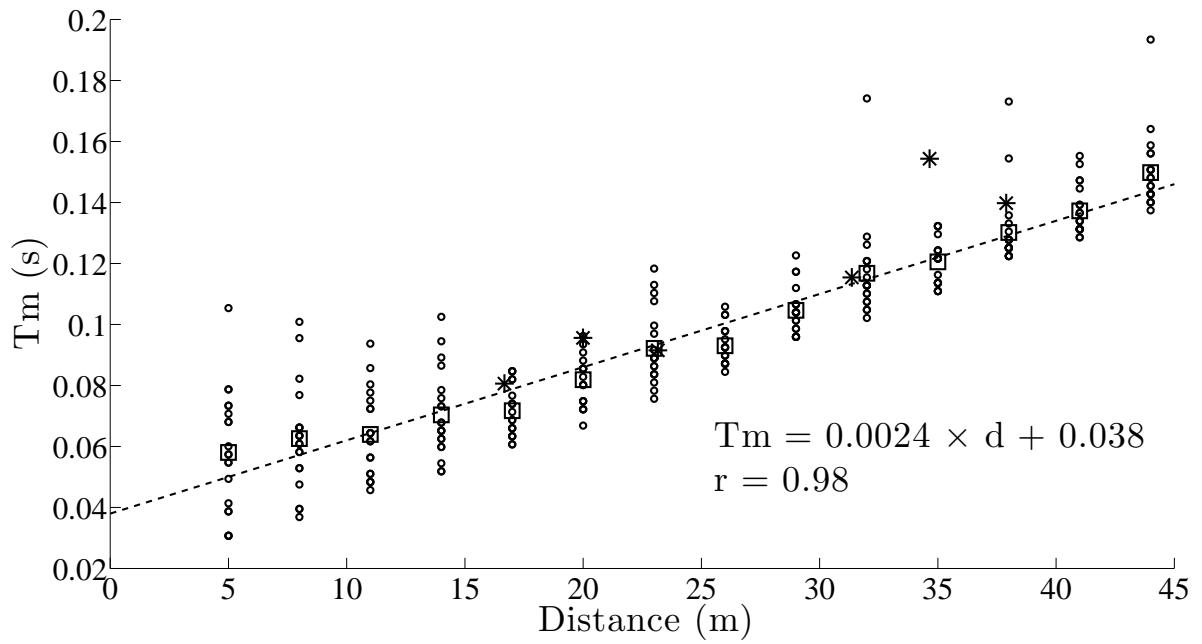


FIG. 6.13 – Dots : estimated mixing times on 420 synthesized impulse responses for the Salle Pleyel (source : cracker impulse). Stars : experimental mixing times (using cracker impulses). First window width :  $L_0 = 128$  samples. Squares : mean mixing times. Dashed line : linear regression calculated on mean mixing times.

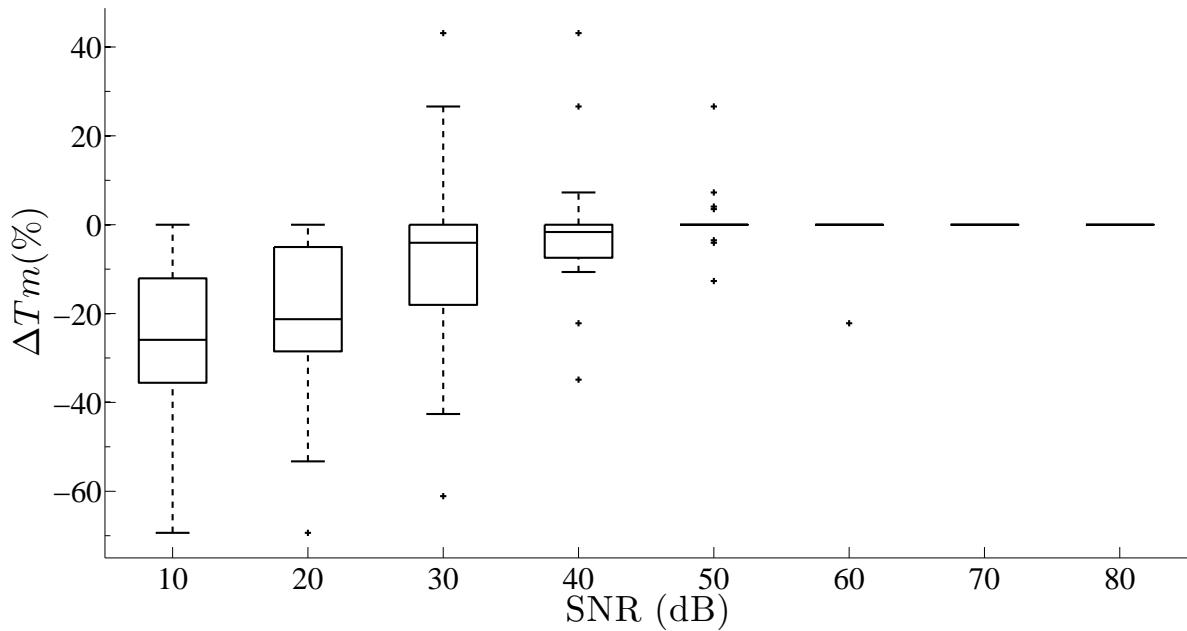


FIG. 6.14 – Variations in percentage (interquartiles) of mixing time versus the signal to noise ratio  $SNR$  (dB). 420 synthesized RIRs are considered (the source is a pistol shot). First analysis window width :  $L_0 = 128$  samples. The reference mixing time is estimated for  $SNR = 80$  dB. Dots stand for extreme values.

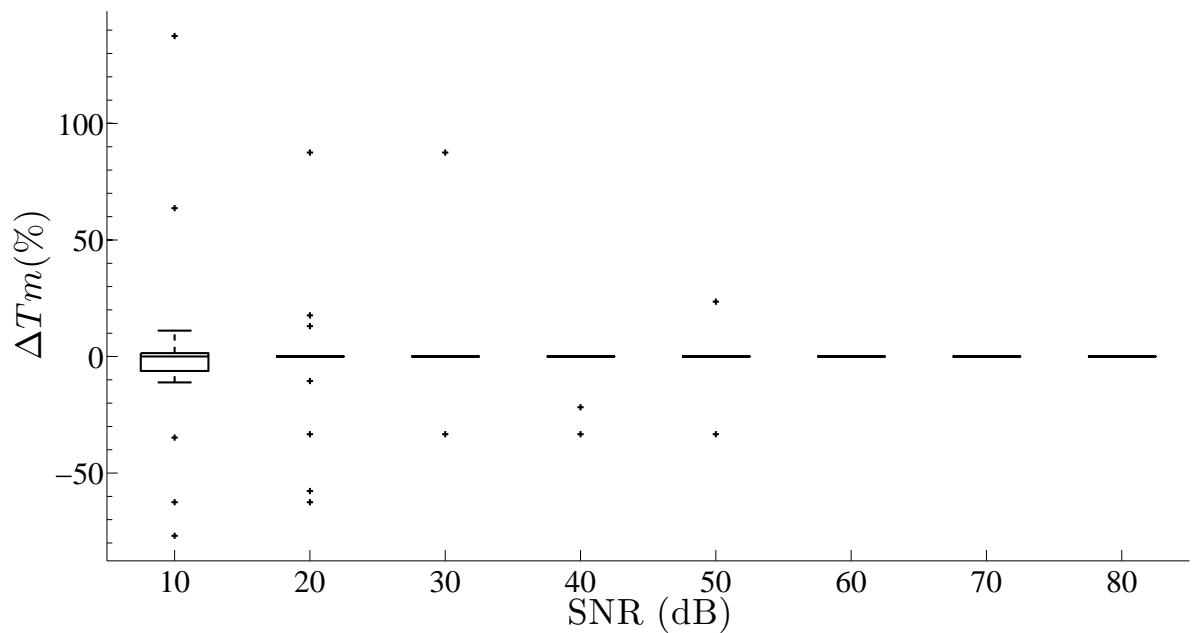


FIG. 6.15 – Variations in percentage (interquartiles) of mixing time versus the signal to noise ratio ( $SNR\ (dB)$ ). 21 RIRs measured in Salle Pleyel are considered (source : pistol shots). First analysis window width :  $L_0 = 128$  samples. The reference mixing time is the one estimated at  $SNR = 80\ dB$  (experimental conditions [52]). Dots stand for extreme values.

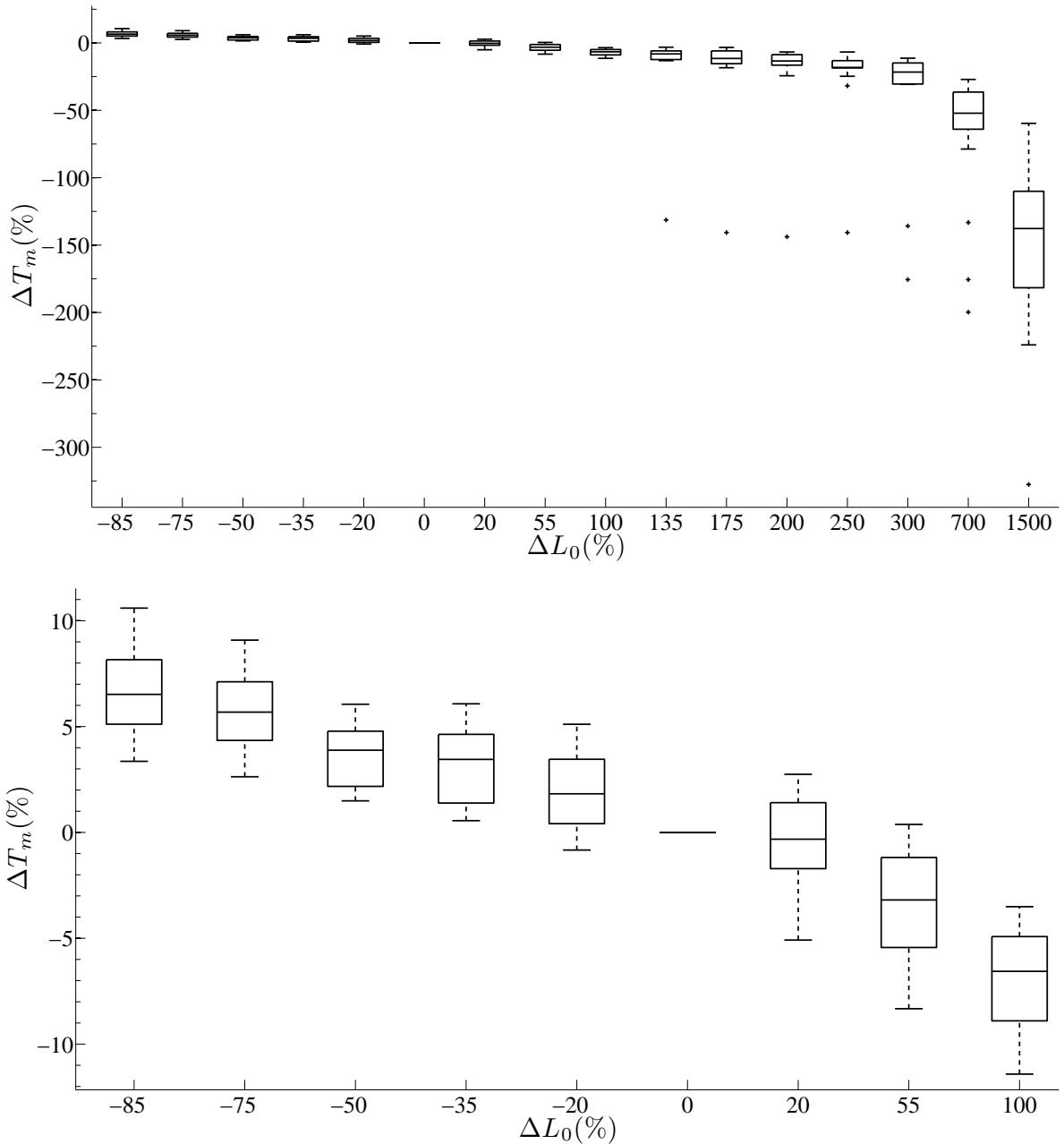


FIG. 6.16 – Signed variations in percentage (interquartiles) of mixing time  $\Delta T_M$  (%) versus variations  $\Delta L_0$  (%) of the analysis window width ( $L_0$ ). 420 synthesized RIRs are considered (the source is a pistol shot). The reference mixing times are the ones estimated for  $\Delta L_0 = 0\%$ , that is,  $L_0 = 128$  samples. Bottom : zoom on the top graphic. Dots stand for extreme values.

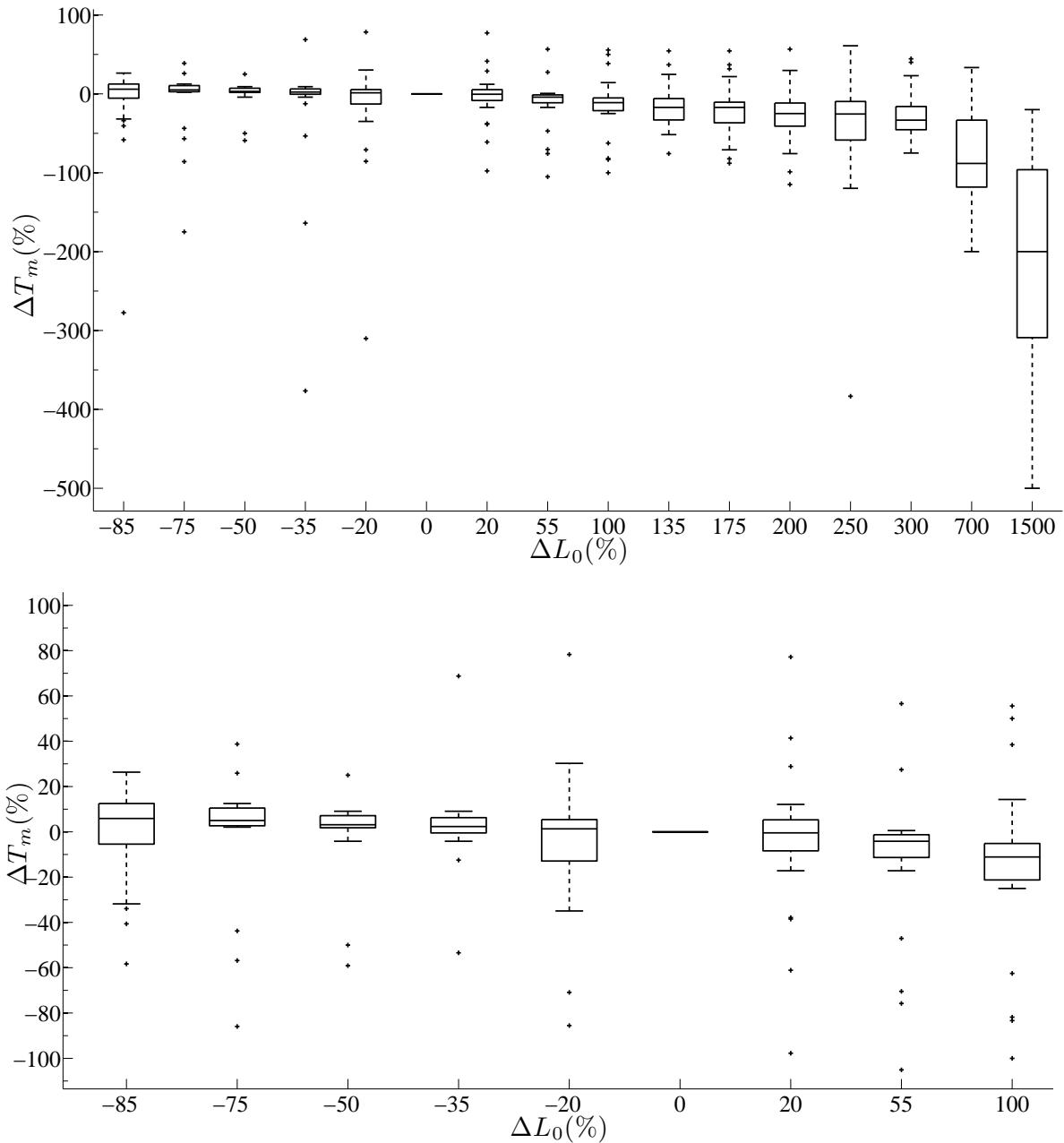


FIG. 6.17 – Top : signed variations in percentage (interquartiles) of mixing time  $\Delta T_M$  (%) versus variations  $\Delta L_0$  (%) of the analysis window width ( $L_0$ ). 21 experimental RIRs (source : pistol shots) are considered (with compensation of the energy decay). The reference mixing times are the ones estimated for  $\Delta L_0 = 0\%$ , that is,  $L_0 = 128$  samples. Bottom : zoom on the top graphic. Dots stand for extreme values.

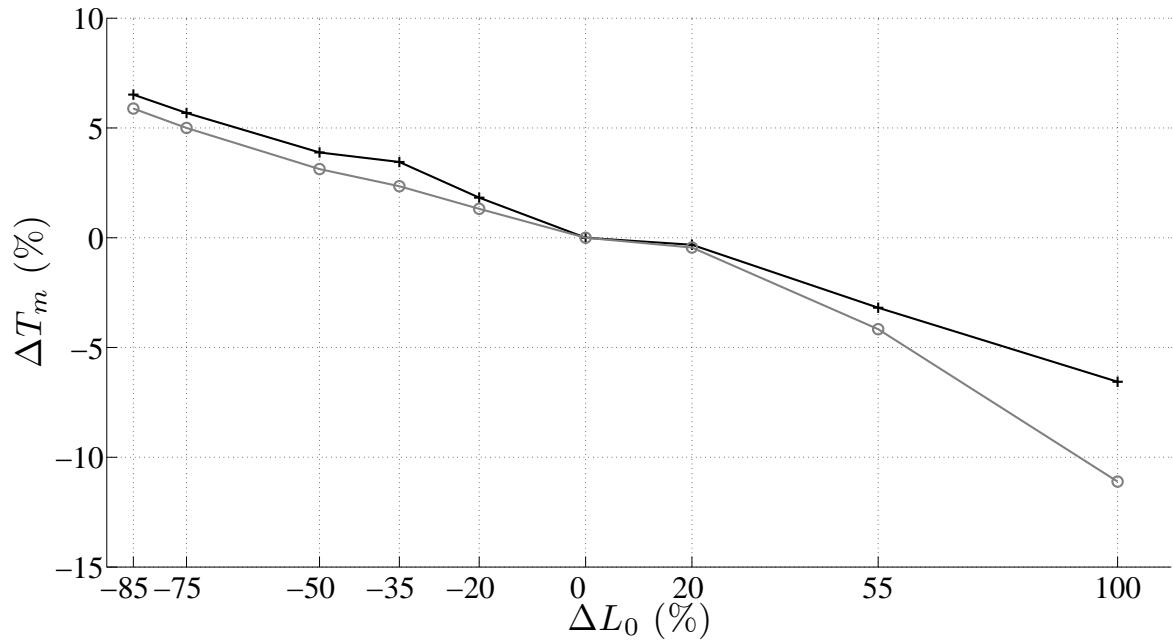


FIG. 6.18 – Median variations in percentage of mixing time  $\Delta T_M$  (%) versus variations  $\Delta L_0$  (%) of the analysis window width ( $L_0$ ). Black line : 420 synthesized RIRs are considered. Grey : 21 experimental RIRs. The source is a pistol shot. First analysis window width :  $L_0 = 128$  samples.

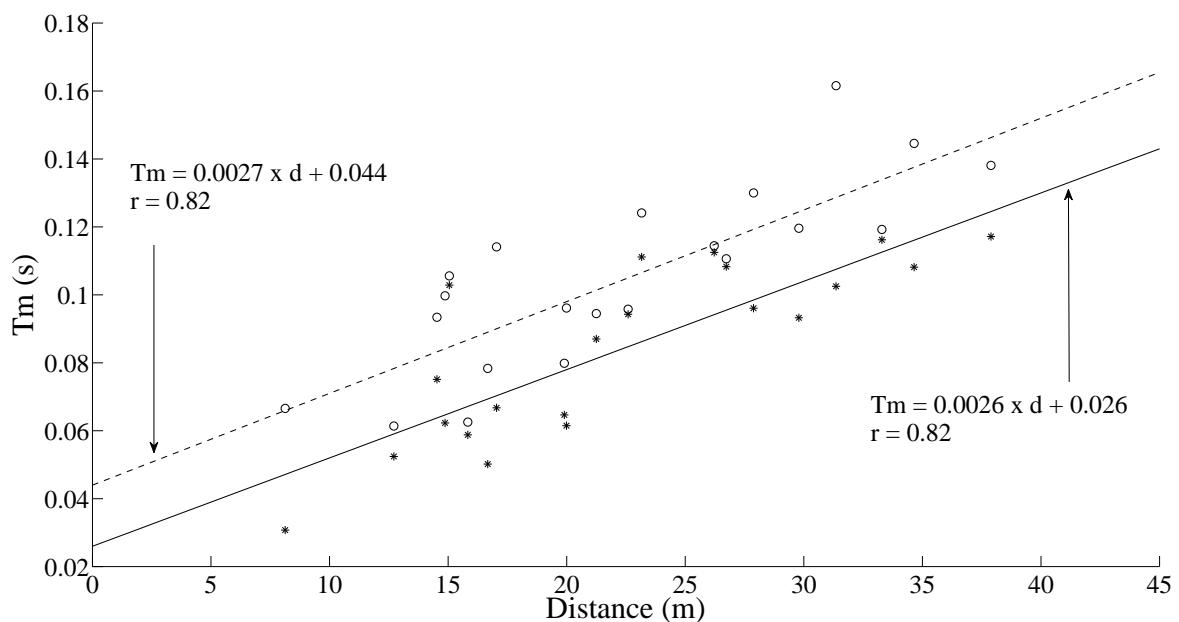


FIG. 6.19 – Temps de mélange estimé dans la Salle Pleyel à l'aide Matching Pursuit [ $SRR = 5dB$ ] (étoiles) et la XFT (cercles). Source : pistolet.

# Chapitre 7

## Discussion des résultats

Les Chapitres 4, 5 et 6 présentent plusieurs méthodes d'estimation de la statistique des réponses impulsionales de salles de spectacles : tests de Gaussianité, Matching Pursuit et XFT. Si les tests de Gaussianité ne constituent pas des outils concluants pour estimer le début du champ diffus, les deux premiers sont en accord, et montrent qu'il existe un temps auquel la dynamique du système change brutalement. Nous avons défini ce temps comme le temps de mélange (Chapitre 2). Il est important de retenir que les résultats présentés dans ce document ne constituent pas une démonstration de l'ergodicité des salles, mais une interprétation de cette théorie.

Dans ce Chapitre, nous tentons d'évaluer la pertinence des résultats obtenus et dégageons de nouveaux axes de recherche.

### 7.1 Principaux résultats

L'analogie entre la propagation du son dans une salle et un système dynamique introduite par Joyce [3] a ouvert la voie à de nombreuses simulations numériques vérifiant la pertinence d'une telle approche. Ainsi, Mortessagne, Kawabe et Koyanagi [107, 108, 109] retrouvent la loi de décroissance de l'énergie de salles de spectacles prédicta par Sabine en considérant des billards dont l'absorption est uniformément répartie. A ces travaux s'ajoute le modèle stochastique de Polack qui est au centre de ce mémoire [27] et qui est validé tant du point de vue des indices acoustiques de salles que de la statistique des retours. Néanmoins, l'hypothèse d'ergodicité des salles de spectacles n'a jamais été vérifiée expérimentalement, ce qui justifie que certains puissent douter de la validité de cette hypothèse [110]. Rappelons qu'il existe en réalité peu de systèmes pour lesquels cette hypothèse a été démontrée.

Dans le présent manuscrit, nous tentons une validation expérimentale de cette hypothèse à l'aide de différents outils : tests du  $\chi^2$ , de Lilliefors, de Kolmogorov-Smirnov, moments d'ordre 2 et 4 d'une distribution Gaussienne, Matching Pursuit, eXtensible Fourier Transform. Nous avons mis en évidence plusieurs phénomènes comme l'apparition de la diffusion, l'augmentation de l'entropie au cours du temps et le temps auquel la distribution des retours devient statistique et que nous avons défini comme le temps de mélange. Ces résultats ne valident pas pour autant les hypothèses d'ergodicité et de mélange des salles énoncées par Joyce.

De plus, nous avons montré au Chapitre 5 qu'après le temps de mélange, le nombre total de rayons sonores captés en différents points d'une salle variait, mais que l'évolution temporelle de ce nombre était en apparence constante dans toute la salle. Cette observation tendrait à montrer que l'énergie n'est pas equipartie, et donc que la salle ne serait pas ergodique. Cependant, ce constat est tributaire de la méthode d'estimation des retours que nous avons employée (Matching Pursuit) et largement discutée (Chapitre 5). En effet, rappelons que le nombre total de retours

captés aux différents points d'observation varie peu d'un emplacement à un autre, et que ce nombre est avant tout une fonction du critère d'arrêt de la poursuite (dans le cas de Matching Pursuit).

## 7.2 Les définitions du temps de mélange

Il semble important de préciser que le temps de mélange est un concept naïf. En effet, un système mélangeant tend vers l'équilibre progressivement (asymptotiquement) et n'atteint pas cet état d'une manière soudaine. Enfin, rappelons que l'hypothèse des salles mélangeantes est utile aux acousticiens, car elle permet de postuler que l'énergie est équitable dans la salle après un nombre suffisant de réflexions [2, 3, 110]. Si l'expérience montre que le taux de décroissance de l'énergie est statistiquement le même dans toute la salle [15], nous n'avons pas connaissance de validations expérimentales prouvant que l'énergie est uniformément répartie dans l'espace des phases. De plus, cette équpartition de l'énergie supposée en acoustique des salles n'est valable qu'à un certaine distance des parois, là où la migration d'énergie ne peut être observée.

Il nous paraît essentiel d'insister sur les différentes définitions, que nous avons données au cours de ce mémoire, du temps de mélange. Dans un premier temps, nous avons défini le mélange comme une propriété de certains systèmes dynamiques. Mais nous avons également défini le temps de mélange comme le temps de transition entre deux régimes (des premiers retours au champ diffus) de la réponse impulsionale d'une salle, et aussi comme le temps au bout duquel dix retours se superposent dans une fenêtre d'observation de 24 ms. Si nous voulons résumer, nous avons trois définitions du temps de mélange : l'une est propre aux systèmes dynamiques, l'autre est relative à l'acoustique d'une salle et à la distribution temporelle des retours en point de la salle, et enfin la dernière fait appel à la perception sonore. Ces trois définitions rendent parfois difficiles la compréhension de ce qu'est le mélange et par conséquent de ce qu'est le temps de mélange.

Si nous souhaitons estimer le temps de mélange au sens des systèmes dynamiques, alors il nous faut observer l'évolution temporel du système dans son espace des phases (ce dont nous discutons dans la Section 7.3). Cette approche requiert de suivre dans l'espace les trajectoires des particules sonores<sup>1</sup>, alors que les deux autres définitions s'appliquent à une analyse physique ou perceptive du système en un point donné de l'espace. L'approche dynamique n'est en rien reliée, nous le concevons bien, à l'approche perceptive proposée par Polack [9], puisque cette dernière s'intéresse au nombre de retours perçus dans une fenêtre de temps donné et qu'elle ne s'intéresse pas aux trajectoires divergentes des rayons dans la salle. Le mélange au sens perceptif du terme est à comprendre comme une impossibilité de distinguer, de discriminer plus de dix événements dans un temps donné. Dans le Chapitre 5, nous avons comparé le temps de mélange estimé par Matching Pursuit et le critère de Polack (nous nous intéressons donc ici à la statistique des retours). Nous montrons que le lien entre les deux approches n'est pas évident, mais que ces seules mesures ne suffisent pas à conclure. En revanche, le Chapitre 6 se concentre sur la différence de phase de la XFT pour estimer le temps de mélange, c'est-à-dire que nous considérons la statistique des retours ainsi que l'évolution de la corrélation du système à un temps donné relativement à son état initial. Cette méthode se rapproche de la définition du mélange au sens des systèmes dynamiques, en ce qu'un système mélangé est défini entre autre par une perte de mémoire progressive, un baisse de corrélation au regard de son état initial. Néanmoins, comme la XFT et la différence de phases ne considèrent pas l'évolution spatiale des trajectoires, nous ne pouvons conclure quant à l'équivalence stricte entre les approches dynamique et statistique.

---

<sup>1</sup>Ce que nous n'avons pas fait dans le travail présenté dans les précédents Chapitres.

Nous sommes tentés de faire l'hypothèse que le temps que nous estimons grâce à la XFT et Matching Pursuit est une limite inférieure du temps de mélange de la salle. En effet, localement des retours peuvent être statistiquement distribués sans que ce soit le cas dans toute la salle, c'est-à-dire à un niveau macroscopique (Fig. 7.1). Le temps nécessaire pour qu'en tout point de la salle nous puissions mesurer une distribution statistique des retours (une équpartition de l'énergie, une densité homogène et isotrope des rayons) est au moins égal au temps de propagation entre le point d'émission et le point d'observation. A ce temps de propagation, il faut ajouter le temps d'établissement d'un régime diffus local (Fig. 7.1), plus le temps pour que ce régime diffus envahisse tout l'espace.

Il paraît donc plus sage de définir le temps que nous avons appelé "temps de mélange" comme "*temps de transition*" (ou "*cross-over time*" en Anglais, en référence à la fréquence de coupure de Schroeder, ou "*cross-over frequency*"). Le temps de transition définit le temps à partir duquel les retours d'une réponse impulsionale sont statistiquement distribués en un point de l'espace. Le mélange, qui suppose une distribution homogène et isotrope des rayons dans tout l'espace, sera atteint à un temps supérieur au temps de transition.

Dans la suite, nous proposons deux approches expérimentales afin de vérifier le caractère mélangeant ou tout au moins ergodique des grandes salles.

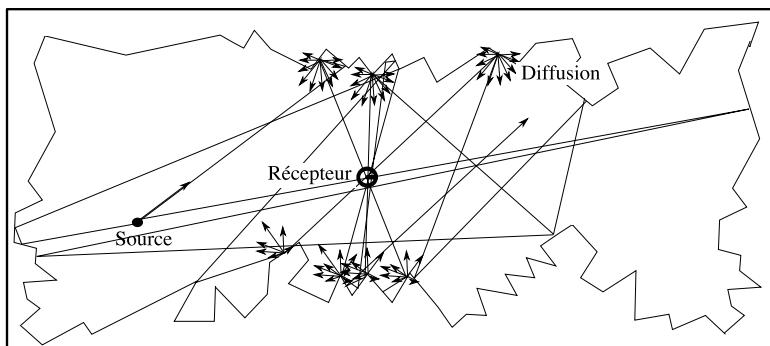


FIG. 7.1 – Schéma d'une salle aux parois diffusantes. La diffusion s'établit progressivement dans l'espace.

### 7.3 Alternatives

La principale raison qui fait que le travail que nous avons présenté jusqu'ici ne prouve pas -au sens des systèmes dynamiques- le caractère mélangeant ou ergodique des salles de spectacles est la technique d'observation. En effet, en toute rigueur, il faut observer sur des temps très longs les états du système dans l'espace des phases qui est défini, pour une salle, par la position et la direction des particules sonores. Les résultats présentés précédemment sont obtenus à partir de mesures de réponses impulsionales. Autrement dit, nous avons analysé la distribution temporelle de formes d'onde en plusieurs points de l'espace sur des temps relativement courts<sup>2</sup>. Des

<sup>2</sup>Le temps d'observation est de l'ordre de 1.5 secondes pour la Salle Pleyel. Pour des temps supérieurs au temps de réverbération, le niveau sonore de la réponse devient trop faible, et seul le bruit de fond est capté. C'est une contrainte forte pour l'étude de ce type de système dynamique. De plus, les mesures de réponses impulsionales que nous avons exploitées dans ce travail auraient sans doute dû être réalisées loin des parois (ou d'obstacles tels

réponses mesurées, nous avons extrait les temps d'arrivée des particules sonores en ces points, et l'évolution de ces temps d'occurrence au cours du temps (Chapitre 5) selon plusieurs directions de l'espace ou à l'aide d'une captation omnidirectionnelle.

Afin d'observer l'espace des phases, il nous faut suivre le parcours d'une ou plusieurs particules dans l'espace et dans le temps. Naturellement, nous concevons la difficulté de telles mesures dans une salle : cela nécessite de discréteriser en trois dimensions l'espace plus ou moins finement et de manière régulière. La discréterisation de l'espace introduit évidemment une incertitude de détection des particules. Il faut donc imaginer un système de captation à l'allure d'un grille de microphones omnidirectionnels ou B-format (Fig. 7.2). A chaque point du maillage, une réponse impulsionnelle est mesurée et les retours (temps d'occurrence et limites temporelles) sont estimés. Le trajet d'une particule est obtenu en estimant les retours les plus corrélés entre deux points du maillage. Entre autres difficultés, nous avons identifié au Chapitre 5 celle qui consiste à estimer les limites temporelles d'un retour. De plus, nous supposons que même pour des temps faibles (i.e., après seulement quelques réflexions), il existe des particules possédant (presque) la même information. En d'autres termes, plusieurs particules peuvent avoir subi le même filtrage global et parcouru la même distance. Dans ce cas, la question est d'identifier notre particule parmi un ensemble de possibilités. Ce choix à faire introduit une incertitude qu'il faut quantifier. Ce problème est d'autant plus difficile à surmonter qu'il y a au cours du temps de plus en plus de particules créées (Chapitre 2).

Une telle approche, qui doit s'appliquer à un grand nombre de particules, doit permettre d'une part de connaître la position et la direction d'une particule au cours du temps, et d'autre part, d'estimer la perte d'information de chaque particule - ou leur entropie- au cours du temps. Nous avons ainsi la possibilité d'évaluer le caractère ergodique et mélangeant des grandes salles. En effet, un critère possible pour définir le temps de mélange consiste à déterminer le temps auquel l'algorithme de suivi d'une particule ne sait plus identifier une particule parmi d'autres dans une intervalle de confiance que nous fixons. Autrement dit, le mélange est atteint quand statistiquement toutes les particules ont la même information et donc la même probabilité d'être reconnues. Le mélange est aussi atteint quand nous mesurons une densité de particules homogène et isotrope en tout point de la matrice de microphones. Enfin, cette approche permet une estimation d'une part, du filtrage de la salle (Chapitre 5), et d'autre part, de la répartition de l'énergie dans l'espace. Nous sommes alors capables de vérifier l'équipartition de l'énergie du système.

Une alternative au problème que nous avons soulevé précédemment est de ne considérer qu'un volume restreint de la salle, et d'estimer au cours du temps le nombre de particules traversant ce volume. Il faut pour cela disposer d'une antenne de microphones plus petite que celle précédemment décrite (Fig. 7.2). Une antenne mobile de forme cubique ou sphérique semble facilement réalisable<sup>3</sup>. De plus, une condition nécessaire à la réussite d'une telle approche est de disposer d'une source sonore parfaitement reproductible. En effet, le moindre changement de directivité ou de la forme d'onde émise modifie les données observées. C'est-à-dire que les particules émises par la source ne suivent plus les même chemins ; d'une mesure à l'autre, nous n'observons pas le même nombre de particules dans notre volume. Nous retrouvons ici la sensibilité aux conditions initiales des systèmes chaotiques, si tant est que les salles sont des systèmes chaotiques !

A chaque pas de temps, le nombre et la position des particules sont rapportés, comme cela est

---

que les sièges ou les balcons) afin de s'affranchir de la migration de l'énergie et des phénomènes de diffraction par exemple. Autrement dit, des mesures judicieuses de réponses impulsionales de salles consisteraient en des mesures loin de toute paroi.

<sup>3</sup>Le nombre de microphones à utiliser, leur directivité et leur emplacement sur l'antenne est à étudier.

représenté dans la Fig. 7.3. Comme nous l'avons montré précédemment (Chapitre 5), le nombre de particules au cours du temps doit être une fonction croissante du temps. Cette approche permet également valider les résultats obtenus dans les Chapitres 5 et 6 où le nombre cumulé de retours obtenu est une fonction cubique du temps jusqu'à un certain temps à partir duquel il devient –en apparence– une fonction linéaire du temps.

La détection de particules traversant le volume d'observation peut être réalisée à l'aide de Matching Pursuit. Nous obtenons une série d'images comme celles de la Fig. 7.3. Des outils d'analyse d'images permettent de compter le nombre de particules et d'estimer leur position au cours du temps. Nous supposons qu'il existe un temps à partir duquel le nombre de particules ne varie plus significativement et la densité spatiale de particules oscille autour d'une constante. Statistiquement, toutes les positions de l'espace doivent être occupées par une ou plusieurs particules à partir de ce temps. Ceci permet alors de vérifier l'équipartition de l'énergie dans un échantillon de volume de notre espace et par conséquent de conclure –en étendant cette observation à tout le volume de la salle– quant à la validité de l'ergodicité de la salle.

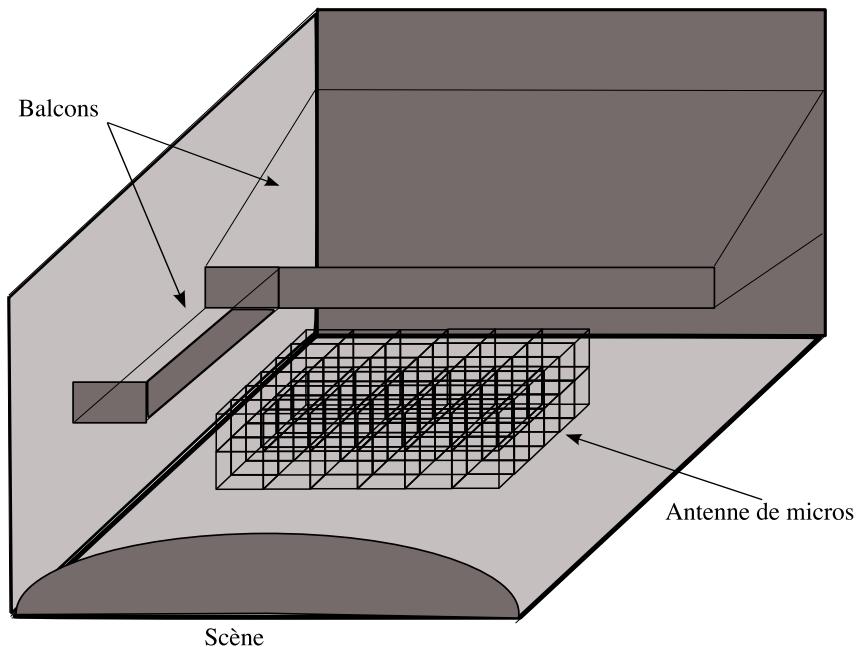


FIG. 7.2 – Schéma d'une salle et d'une antenne de microphones.

Comme nous l'avons déjà mentionné plus haut, la mise en œuvre des techniques de mesure décrites précédemment est sans aucun doute assez lourde. Une approche peut-être plus simple, dans un premier temps, est de réaliser ces mesures dans une maquette de salle (Section 2.6). Mais, nous pouvons d'ores et déjà identifier la limite de cette technique. Une antenne de microphones si petits soient-ils perturbe le champ sonore de manière conséquente (selon l'échelle de la maquette). A moins d'imaginer une réseaux de microphones transparents acoustiquement, une méthode de visualisation de propagation des fronts d'onde paraît plus appropriée. Ainsi, la technique de “Schlieren” (Section 2.6, [53]) semble toute indiquée, même si la détection des front d'onde ne peut se faire qu'en deux dimensions de l'espace (Fig. 7.4). Dans la pratique, il s'agit de filmer à l'aide d'une caméra rapide les fronts d'onde se propageant dans une zone de l'espace qui nous intéresse. Puis à l'aide d'un algorithme d'analyse d'images, mailler la zone d'observation et compter le nombre de fronts d'onde interceptés par chaque point du maillage. Si cette technique

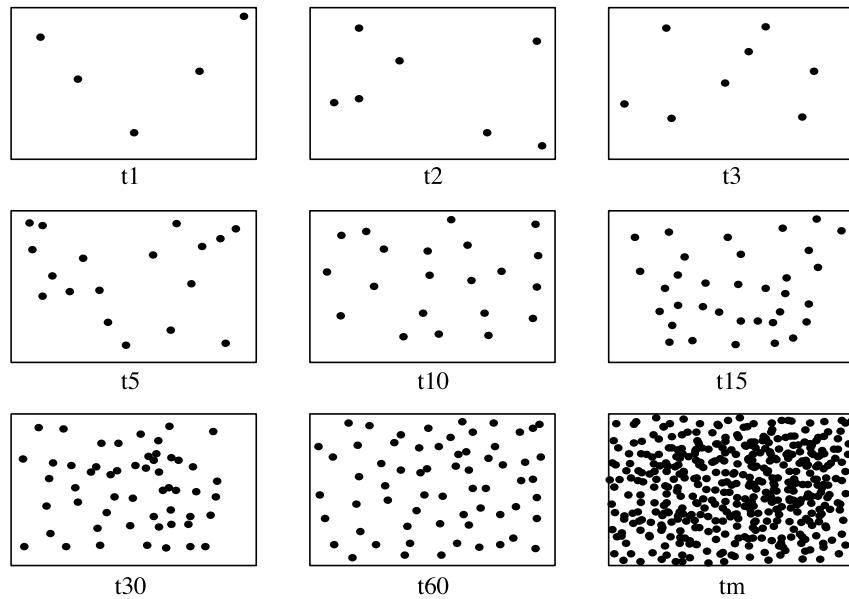


FIG. 7.3 – Observation schématique en deux dimensions du nombre de particules captées dans un volume de l'espace au cours du temps ( $t_1 \dots t_m$ ). Ici,  $t_m$  correspond au temps à partir duquel la densité de particules est uniforme sur tout l'espace observé.

ne permet pas d'estimer l'entropie du système, elle offre néanmoins la possibilité d'estimer la propriété d'ergodicité du système comme nous l'avons décrit précédemment.

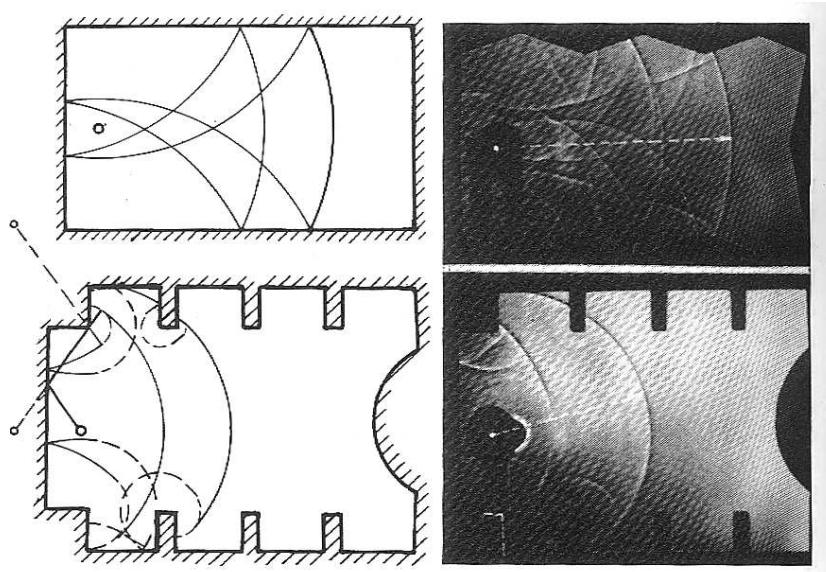


FIG. 7.4 – Photographies de la propagation d'ondes dans des maquettes à l'aide la technique de Schlieren [12].

# Chapitre 8

## Conclusion

Malgré son titre, ce manuscrit ne s'est pas seulement attaché à étudier le mélange et estimer son temps d'occurrence dans les salles de spectacles. Différents aspects de l'acoustique des salles ont été traités : la théorie statistique, la perception, la mesure, la synthèse, les indices acoustiques, la caractérisation de sources sonores, la modélisation stochastique de réponses de salles, l'estimation des retours et du temps de mélange. Nous nous sommes particulièrement concentrés sur l'estimation des propriétés de la théorie statistique de l'acoustique des salles et ce, d'une part, afin de vérifier certains points de cette théorie, et d'autre part, dans le but d'améliorer le modèle stochastique de réponses de salles. Nous avons finalement été amenés à redéfinir ce qu'au début de ce mémoire nous appelions temps de mélange en *temps de transition*<sup>1</sup>.

En nous basant sur la théorie statistique de l'acoustique des salles, nous avons tenté de détecter le début du champ diffus à l'aide de estimateurs de Gaussianité. Nous avons donc implicitement admis que ce régime du système est Gaussien. Les quelques pages qui présentent cette approche mettent cependant en évidence la limite de tels tests : la taille de fenêtre d'analyse, et le choix de ces tests. Malgré ces limitations qui excluent l'emploi de ces outils comme estimateurs du temps de transition, nous montrons qu'il existe des *bouffées de cohérence*, c'est-à-dire des retours encore fortement corrélés au son direct alors que le champ diffus est déjà établi. Si l'étude poussée de ces bouffées dépasse le cadre de cette thèse, nous soulignons l'intérêt que représentent celles-ci. En effet, appréhender ce phénomène permettrait de mieux comprendre le système physique que nous étudions : les salles de spectacles.

L'estimation des retours à l'aide Matching Pursuit permet de vérifier que leur distribution temporelle vérifie la théorie statistique de l'acoustique des salles dans les premiers temps de la réponse impulsionale. Passé le temps de transition, la densité de retours augmente d'un rapport constant (*en apparence*) avec le temps, alors que la théorie prédit une évolution temporelle cubique. L'estimation des retours à partir du modèle stochastique permet de vérifier ces résultats. L'estimation du temps d'arrivée des retours ouvre la voie à l'estimation du filtrage de la salle comme une fonction du temps. Nous distinguons ici la possibilité d'estimer la perte d'information du système, ou entropie, et la coloration de la salle. Soulignons qu'une condition indispensable dans ce cas est la constitution d'un dictionnaire d'atomes plus raffiné (choix des filtres, durées des atomes, etc.) que celui que nous avons employé. L'estimation des retours permet également d'étudier la statistique des retours en trois dimensions et ouvre certainement la voie au développement de nouveaux indices acoustiques.

Parallèlement, nous avons présenté l'*eXtensible Fourier Transform* qui permet une visualisation pas-à-pas des composantes fréquentielles d'un signal en gardant la même origine du

---

<sup>1</sup>Ou *cross-over time* en Anglais.

temps. Appliquée aux réponses de salles, la phase de la XFT nous permet de rendre compte de l'évolution de la statistique des composantes spectrales des réponses. Il apparaît ainsi un point de rupture que nous définissons comme le temps de transition. Nous montrons que pour la Salle Pleyel, les résultats de Matching Pursuit sont semblables à ceux de la XFT.

D'autre part, Matching Pursuit et l'eXtensible Fourier Transform s'accordent tant sur des mesures de réponses de salles que sur le modèle stochastique pour établir que le temps de transition est une fonction de la distance source/récepteur et est borné (limite supérieure) par sa formule heuristique proposée par Polack [9].

Nous avons soulevé une question dont la réponse sera donnée par l'expérimentation : quel est le temps de transition lorsque le point réception est identique au point source ? Nous avons vu que le résultat différait selon les paramètres du modèle (synthèse ou non de la diffusion et du filtrage de la salle) et aussi selon l'estimateur utilisé. Il semblerait que dans cette configuration, le temps de transition soit atteint après un libre parcours moyen. Il est pour l'instant difficile de conclure quant à la pertinence de ce résultat. En effet, nous avons vu et discuté l'importance de raffiner le dictionnaire d'atomes et de choisir judicieusement les paramètres de la XFT.

Le modèle stochastique dont il est sujet tout au long de ce mémoire a été validé au regard de la statistique des retours. Il semble important de rappeler que, lors de cet exposé, nous avons introduit quelques apports au modèle : la synthèse de la diffusion et du filtrage aléatoire de la salle. Si la diffusion est synthétisée à partir de l'équation de diffusion, les méthodes de filtrage sont, elles, intuitives et peuvent sûrement être soumises à controverse. Nous retiendrons néanmoins qu'une partie de ces ajouts constituent des améliorations : les temps de transition estimés se trouvent alors similaires à ceux mesurés expérimentalement. De plus, des premiers tests perceptifs de validation du modèle montrent clairement que certains des filtrages proposés se rapprochent de l'expérience des sujets (Annexe E). Nous sommes alors en mesure de conclure que les salles sont sans doute des générateurs de nombres aléatoires, comme l'annonçait Polack [7]. C'est plus particulièrement le cas du filtrage par Variations Spectrales Aléatoires. Dans le futur, il serait intéressant de vérifier la pertinence de ce genre de filtrage et de le confronter au filtrage mesurable dans des réponses impulsionales de salles.

De plus, dans un Chapitre à part, nous insistons sur le regard critique qui doit être porté aux résultats présentés dans ce manuscrit. Ainsi, ces résultats relèvent de l'interprétation de la théorie des systèmes dynamiques et ne constituent pas une démonstration. En confrontant les différentes définitions du temps de mélange (théorie des systèmes dynamiques, Polack, acoustique des salles), nous concluons sur le fait que le temps que nous estimons n'est pas le temps de mélange. A la place, nous proposons l'emploi du terme de *temps de transition*.

Nous identifions quelques approches expérimentales utiles à la vérification de l'hypothèse de Joyce selon laquelle les salles sont des systèmes dynamiques mélangeants.

Les conclusions et les questions soulevées par les travaux que nous avons présentés nécessitent de plus amples recherches naturellement. Toute conclusion hâtive serait grave, et il serait judicieux de confronter nos résultats aux données obtenues à partir de campagnes de mesures dans de nombreuses salles à la géométrie différente (complexe ou simple) et en utilisant des sources sonores reproductibles. Néanmoins, loin d'être négatifs, nous voulons insister sur le fait que si nos données concernent peu de salles, nos résultats et conclusions sont également issus de l'analyse de plusieurs milliers de réponses impulsionales synthétisées (soit autant de salles) par le modèle stochastique.

# Bibliographie

- [1] F. V. Hunt, *Origins in acoustics. The Science of Sound from Antiquity to the Age of Newton*, Acoustical Society of America, 1992.
- [2] W. C. Sabine, *Collected Papers on Acoustics*, (Harvard Univ. Press, Cambridge) (reprinted by Dover, New York), 1964.
- [3] W. B. Joyce, “Sabine’s reverberation time and ergodic auditoriums,” *J. Acoust. Soc. Am.*, vol. 58, no. 3, pp. 643–655, 1975.
- [4] J. B. Allen and D. A. Berkely, “Image method for efficiently simulating small-room acoustics,” *J. Acoust. Soc. Am.*, vol. 65, no. 4, pp. 943–950, 1979.
- [5] A. Krokstad, S. Strom, and S. Sorsdal, “Calculating the acoustical room response by the use of a ray tracing technique,” *J. Sound and Vib.*, vol. 8, no. 1, pp. 118–125, 1968.
- [6] J-D Polack, “Modifying chambers to play billiards : the foundations of reverberation theory,” *Acta Acustica*, vol. 76, pp. 257–272, 1992.
- [7] J-D. Polack, “Reverberation time and mean absorption in concert halls,” in *Proceedings of the Institute of Acoustics*, 2006, vol. 28, p. 2.
- [8] N. S. Krylov, *Works on Foundations of Statistical Physics*, Princeton University Press (translated by A.B. Migdal and Ya. G. Sinai and Yu. L. Zeeman), 1979.
- [9] 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.
- [10] M. Bruneau, *Introduction aux théories de l’acoustique*, Ed. Université du Maine, Le Mans, 1983.
- [11] J. Jouhaneau, *Acoustique des salles et sonorisation*, Acoustique Appliquée. Cnam - TEC DOC, 2nd edition, 2003.
- [12] L. Cremer, H.A. Muller, and T.J. Schultz, *Principles and Applications of Room Acoustics*, vol. 1, Applied Science Publishers Ltd, 1982.
- [13] H. Weyl, “Das asymptotische Verteilungsgesetz der Eigenwerte linearer partieller Differentialgleichungen,” *Math. Ann.*, vol. 71, pp. 441–479, 1912.
- [14] 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.
- [15] H. Kuttruff, *Room Acoustics*, Elsevier Applied Science, 3 edition, 1991.
- [16] 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.
- [17] 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.

- [18] M. R. Schroeder and H. Kuttruff, "On Frequency response Curves in Rooms. Comparison of Experimental, Theoretical and Monte Carlo results for the Average Frequency Spacing Between Maxima," *J. Acoust. Soc. Am.*, vol. 34, no. 1, pp. 76–80, 1962.
- [19] James A. Moorer, *About This Reverberation Business*, Ircam, 1978.
- [20] X. Meynial, G. Dodd, J-D. Polack, and A-H. Marshall, "All-scale model measurements :the midas system," in *121st ASA Meeting, special session on auditorium measurements*, Baltimore, April 1991.
- [21] X. Meynial, J-D. Polack, and G. Dodd, "Comparison between full-scale and 1 :50 scale model measurements in theatre municipal, le mans," *Acta Acustica*, vol. 1, pp. 199–212, October/December 1993.
- [22] N. Korany, J. Blauert, and O. Alim Abdel, "Acoustic simulation of room with boundaries of partially spacial reflectivity," *Applied Acoustics*, vol. 62, pp. 875–887, 2001.
- [23] N. Campo, P. Rissone, and M. Toderi, "Adaptive pyramid tracing : a new technique for room acoustics," *Applied Acoustics*, vol. 61, pp. 199–221, 2000.
- [24] X. Y. Zeng, "An improved broad-spectrum room acoustics model including diffuse reflections," *Applied Acoustics*, vol. 66, pp. 1309–1319, 2005.
- [25] F. P. Mechel, "Improved mirror source method in room acoustics," *J. Sound and Vibration*, vol. 256, no. 5, pp. 873–940, 2002.
- [26] F. Cotana, "An improved room acoustic model," *Applied Acoustics*, vol. 61, pp. 1–25, 2000.
- [27] J-D Polack, "Playing Billiards in the Concert Hall : The Mathematical Foundations of Geometrical Room Acoustics," *Applied Acoustics*, vol. 38, pp. 235–244, 1993.
- [28] O. Legrand and D. Sornette, "Role of the absorption distribution and generalization of exponential reverberation law in chaotic rooms," *J. Acoust. Soc. Am.*, vol. 88, pp. 865, 1993.
- [29] A. Ruegg, *Processus stochastiques avec applications aux phénomènes d'attente et de fiabilité*, Presses polytechniques romandes, 1st edition, 1989.
- [30] P. Morse and H. Feshbach, *Methods of Theoretical Physics*, vol. II, McGraw-Hill Book Company, 1953.
- [31] J. Picaut, *Propagation du son en milieu urbain*, Ph.D. thesis, Université du Maine, 1996.
- [32] M. Kleiner, B. Dalenback, and P. Svensson, "Auralization - An Overview," *J. Audio Eng. Soc.*, vol. 41, no. 11, pp. 861–875, 1993.
- [33] G. Defrance, L. Daudet, and J-D. Polack, "Using Matching Pursuit for estimating mixing time within Room Impulse Responses," *Acta Acustica united with Acustica*, vol. 95, no. 6, pp. 1082–1092, 2009.
- [34] A. A. A. Bassuet, "Acoustics of a selection of famous 18th century opera houses : Versailles, Markgräflisches, Drottningholm, Schweitzingen," in *Acoustics'08*, 2008, pp. 1645–1650.
- [35] G.R. Schodder, "Über Die Verteilung Der Energiereicheren Schallruckwurfe In Salen," *Acustica*, vol. 6, pp. 445–465, 1956.
- [36] H. Kuttruff, "Auralization of Impulse Responses Modeled on the Basis of Ray-Tracing Results," *J. Audio Eng. Soc.*, vol. 41, no. 11, pp. 876–880, 1993.
- [37] M. R. Schroeder, "Integrated-impulse method measuring sound decay without using impulses," *J. Acoust. Soc. Am.*, vol. 66, no. 2, pp. 497–500, 1979.
- [38] M. Emerit, *Simulation binaurale de l'acoustique des salles de concert*, Ph.D. thesis, CSTB - Institut national polytechnique de Grenoble, 1986.

- [39] A. Hurtado-Huyssen, *Intensité acoustique appliquée aux métiers du son*, CNAM, Mémoire pour le diplôme d'ingénieur, 2006.
- [40] X. Pelorson, J-P. Vian, and J-D. Polack, “Stabilité et variabilité des critères objectifs utilisés en acoustique des salles,” *Journal d'Acoustique*, vol. 4, pp. 599–619, 1991.
- [41] ISO 3382, *Acoustics-measurements of the reverberation time of rooms with reference to other acoustical parameters*, 1997.
- [42] Leo Beranek, *Concert and Opera Halls : How They Sound*, Acoustical Society of America, 1996.
- [43] M. R. Schroeder, “New method of measuring reverberation time,” *J. Acoust. Soc. Am.*, vol. 37, pp. 409–412, 1965.
- [44] P. Penna, “Some statistical properties of reverberation process in diffuse sound fields,” M.S. thesis, The acoustics laboratory, Danmark tekniske højskole, 1993.
- [45] V. L. Jordan, “A group of objective acoustical criteria for concert halls,” *Applied Acoustics*, vol. 14, pp. 253–266, 1981.
- [46] Vilhelm. L. Jordan, “A comprehensive musical criterion : the inversion index,” *47<sup>th</sup> AES Convention*, March 1974.
- [47] G. Defrance, L. Daudet, and J-D. Polack, “Characterizing sound sources for room acoustical measurements,” in *Proceedings of the Institute of Acoustics*, 2008, vol. 30.
- [48] 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.
- [49] Juichi Igarashi, “Sound produced by pistol shot,” *The reports of the Institute of Science and Technology, University of Tokyo*, vol. 11, no. 13, pp. 155–175, 19571205.
- [50] A. Nash, “On the acoustical characteristics of a balloon,” in *Proc. Int. Symp. Room Acoustics*, Sevilla, September 2007.
- [51] A. C. Gade and J. H. Rindel, “Akustic i danske koncertsale,” *The acoustics laboratory, Techn. Univ. of Denmark*, vol. 22, 1984.
- [52] G. Defrance, J-D. Polack, and B-FG. Katz, “Measurements in the new Salle Pleyel,” in *Proc. Int. Symp. Room Ac.*, Sevilla, September 2007.
- [53] W. Merzkirch, *Flow visualization*, Academic Press INC, London, 2nd edition, 1987.
- [54] L. D. C. Gonzalez-Rodriguez and F. E. Rodriguez-Manzo, “Simulation of the reverberation time of an existing architectural space using a 1 :10 scale model,” in *Acoustics'08*, 2008, pp. 4551–4556.
- [55] 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.
- [56] X. Pelorson, J-P. Vian, and J-D. Polack, “On the variability of room acoustical parameters : reproducibility and statistical validity,” *Applied Acoustics*, vol. 37, pp. 175–198, 1992.
- [57] J. Bello, L. Daudet, S Abdallah, C. Duxbury, M. Davies, and M. Sandler, “A tutorial on onset detection in musical signals,” *IEEE Trans. Speech and Audio Proc.*, vol. 13, no. 5, pp. 1035–1047, 2005.
- [58] Dik J. Hermes, “Vowel-onset detection,” *The Journal of the Acoustical Society of America*, vol. 87, no. 2, pp. 866–873, 1990.
- [59] S. Mallat, *A wavelet tour of signal processing*, New York : Academic Press, 1999.
- [60] M. Basseville and I. V. Nikiforov, *Detection of Abrupt Changes-Theory and Application*, Englewood Cliffs, NJ : Prentice-Hall, 1993.

- [61] D. Griesinger, “Beyond MLS- occupied hall measurement with fft techniques,” 2004.
- [62] S. Mallat and Z. Zhang, “Matching pursuit with time-frequency dictionnaires,” *IEEE Trans. Signal Process.*, vol. 40, no. 12, pp. 3397–3415, December 1993.
- [63] P. Flandrin, “A time-frequency formulation of optimum detection,” in *IEEE Trans. Acoust. Speech, Sig. Proc.*, 1988, vol. 36, pp. 1377–1384.
- [64] G. Defrance, L. Daudet, and J-D. Polack, “Detecting arrivals within Room Impulse Response using Matching Pursuit,” in *Proc. of the 11th Int. Conference on Digital Audio Effects (DAFx-08)*, Espoo, Finland, September 1-4 2008.
- [65] S. Krstulovic and R. Gribonval, “Mptk : Matching pursuit made tractable,” Toulouse, France, May 2006, ICASSP’06.
- [66] 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.
- [67] G. Defrance and J-D. Polack, “Measuring mixing time in auditoria,” *J. Acoust. Soc. Am.*, vol. 123, no. 5, pp. 3499, 2008.
- [68] C. E. Shannon, “The Mathematical Theory of Communication,” *Bell Sys. Tech. J*, vol. 27, pp. 379–423 and 623–656, 1948.
- [69] M. Kuster, “Reliability of estimating the room volume from a single room impulse response,” *J. Acoust. Soc. Am.*, vol. 124, no. 2, pp. 982–993, 2008.
- [70] T. Hidaka, Y. Yamada, and Nakagawa T., “A new definition of boundary point between early reflections and late reverberation in room impulse responses,” *J. Acoust. Soc. Am.*, vol. 122, no. 1, pp. 326–332, 2007.
- [71] 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.
- [72] A. Baskind and J-D. Polack, “Sound Power Radiated by Sources in Diffuse Sound Field,” in *Proceedings of the 108th AES Convention*, 2000, Paris.
- [73] 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.
- [74] V. L. Jordan, “Room acoustics and architectural development in recent years,” *App. Acoustics*, vol. 2, pp. 59–81, 1969.
- [75] R. Kürer, “Zur Gewinnung von Einzahlkriterien bei Impulsmessung in der Raumakustik,” *Acustica*, vol. 21, pp. 370–372, 1969.
- [76] X. Meynil, 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.
- [77] D. Griesinger, “Beyond MLS- Occupied hall measurement with FFT techniques,” 2008 March, <http://world.std.com/griesngr/sweep.pdf>.
- [78] G. Leveau, L. Daudet, and G. Richard, “Methodology and Tools for the evaluation of automatic onset detection algorithms in music,” in *Proceedings of the International Symposium on Music Information Retrieval*. ISMIR’04, 2004.
- [79] J. S. Bendat and A. G. Piersol, *Random Data : Analysis and Measurements Procedures*, New York : Wiley, 1971.
- [80] W. D. Stanley and Peterson S. J., “Equivalent Statistical Bandwidths of Conventional Low-Pass Filters,” *IEEE Transactions on Communications*, vol. 27, no. 10, pp. 1633–1634, October 1979.

- [81] 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.
- [82] W. H. Press, S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery, *Numerical Recipes in C*, Cambridge University Press, second edition, 1992.
- [83] S.O. Rice, “Mathematical Analysis of Random Noise,” *Bell Sys. Tech. J.*, vol. 23, pp. 13–29, 1944.
- [84] S. Moreau, J. Daniel, and S. Bertet, “3D Sound Field Recording with Higher Order Amisonics – Objective Measurements and Validation of a 4th Order Spherical Microphone,” in *120th AES Convention, Paris, France*, 2006.
- [85] C. Gustavino, *Etude sémantique et acoustique de la perception des basses fréquences dans l'environnement sonore urbain*, Ph.D. thesis, Université Pierre et Marie Curie (Paris VI), 2003.
- [86] E. Vincent, “MUSHRAM (MULTi Stimulus test with Hidden Reference and Anchor),” .
- [87] H. Levitt, “Transformed up-down methods in psychoacoustics,” *J. Acoust. Soc. Am.*, vol. 49, no. 2, pp. 467–477, 1971.
- [88] B. Kollmeier, R. H. Gilkey, and U. K. Sieben, “Adaptive staircase techniques in psychoacoustics : A comparison of human data and a mathematical model,” *J. Acoust. Soc. Am.*, vol. 83, no. 5, pp. 1852–1862, 1988.
- [89] J. Daniel, *Représentation de champs acoustiques, application à la transmission et à la reproduction de scènes sonores complexes dans un contexte multimedia*, Ph.D. thesis, Université Pierre et Marie Curie (Paris VI), 2000.
- [90] B. Duval, “Etudes de techniques d’extraction de l’information dans une scène sonore multicanal,” M.S. thesis, Université Pierre et Marie Curie, 2006.
- [91] N. and Gover, “Measurements of directional properties of reverberant sound fields in rooms using a spherical microphone array,” *J. Acoust. Soc. Am.*, vol. 116, no. 4, pp. 2138–2148, 2004.
- [92] T. J. Hargreaves, T. J. Cox, and Y. W. Lam, “Diffusion coefficients,” *J. Acoust. Soc. Am.*, vol. 104, no. 3, 1998.
- [93] R. H. Bolt and P. E. Doak, “A Tentative Criterion for the Short-Term Transient Response of Auditoriums,” *J. Acoust. Soc. Am.*, vol. 22, no. 4, pp. 507–509, 1950.
- [94] G. Benettin and J.M. Strelcyn, “Numerical experiments on the free motion of a point mass moving in a plane convex region : stochastic transition and entropy,” *Phys. Rev. A*, vol. 17, pp. 773–785, 1978.
- [95] G. Defrance, L. Daudet, and J-D. Polack, “Detecting arrivals within room impulse responses using matching pursuit,” in *Proc. of the 11th Int. Conference on Digital Audio Effects (DAFx-08), Espoo, Finland*, 2008.
- [96] F. J. Harris, “On the Use of Windows for Harmonic Analysis with the Discrete Fourier Transform,” *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 66, no. 1, pp. 51–83, 1978.
- [97] Jr H. Hurwitz and M. Kac, “Statistical Analysis of Certain Types of Random Functions,” *The Annals of Mathematical Statistics*, vol. 15, no. 2, pp. 173–181, 1944.
- [98] S. M. Ross, *Initiation aux probabilités (A first course in probability)*, Presses polytechniques et universitaires romandes, 1990.

- [99] M. Arana, A. Vela, and L. San Martin, “Calculating the Impulse Response in Rooms Using Pseudo-Impulsive Acoustic Sources,” *Acta Acustica*, vol. 89, pp. 377–380, 2008.
- [100] D. Sumarac-Pavlovic, M. Mijic, and H. Kurtovic, “A simple impulse sound source for measurements in room acoustics,” *Applied Acoustics*, vol. 69, pp. 378–383, 2008.
- [101] ISO 3382-1, *Acoustics – Measurement of room acoustic parameters – Part 1 : Performance spaces*, 2009.
- [102] Ya G. Sinai, “Dynamical Systems with Elastic Reflections,” *Russ. Math. Surv.*, vol. 25, pp. 141, 1972.
- [103] G. Defrance and J-D. Polack, “Estimating the mixing time of concert halls using the eXtensible Fourier Transform,” *Applied Acoustics*, under review.
- [104] T. Kawabe, K. Aono, and M. Shin-ya, “Acoustic ray chaos and billiard system in Hamiltonian formalism,” *J. Acoust. Soc. Am.*, vol. 113, no. 2, pp. 701–704, 2003.
- [105] S. Koyanagi, T. Nakano, and T. Kawabe, “Application of Hamiltonian of ray motion to room acoustics,” *J. Acoust. Soc. Am.*, vol. 124, no. 2, pp. 719–722, 2008.
- [106] F. Mortessagne, *Dynamique et interférences géométriques dans les billards chaotiques*, Ph.D. thesis, Univ. Paris 7, Denis Diderot, 1994.
- [107] J. B. Lee, “Why concert halls are not ergodic ?,” *J. Acoust. Soc. Am.*, vol. 85, no. 6, pp. 2680–2681, 1989.
- [108] J. Borish, “Extension of the image model to arbitrary polyhedra,” *J. Acoust. Soc. Am.*, vol. 75, no. 6, pp. 1827–1836, 1984.
- [109] Polack. J-D and S. Nezan, “La relaxation dans les gaz visco-thermiques : application à l'étude de l'atténuation et au déphasage d'une onde sonore,” *J. Acoustique*, vol. 5, pp. 299–319, 1992.
- [110] T. J. Schultz and B. G. Watters, “Propagation of Sound accross Audience Seating,” *J. Acoust. Soc. Am.*, vol. 36, no. 5, pp. 885–896, 1964.
- [111] G. M. Sessler and J. E. West, “Sound Transmission over Theatre Seats,” *J. Acoust. Soc. Am.*, vol. 36, no. 9, pp. 1725–1732, 1964.
- [112] P. D'Antonio, P.nio and J. Konnert, “The Directionnal Scattering Coefficient : Experimental Determination,” *J. Audio Eng. Soc.*, vol. 40, no. 12, pp. 997–1017, 1992.
- [113] H. Juricic and F. Santon, “Images et Rayons sonores dans le calcul numérique des échogrammes,” *Acustica*, vol. 28, no. 2, pp. 77–89, 1973.
- [114] U. Stephenson, “Comparison of the Mirror Image Source Method and The Sound Particle Simulation Mehod,” *Applied Acoustics*, vol. 29, pp. 35–72, 1990.
- [115] A. Kulowski, “Error investigation for the ray tracing technique,” *Applied Acoustics*, vol. 15, pp. 263–274, 1982.
- [116] D. V. Maercke and J. Martin, “The Prediction of Echograms and Impulse Responses within the Epidaur Software,” *Applied Acoustics*, vol. 38, pp. 93–114, 1993.
- [117] N. Faillet, “Validation of the pyramid tracing algorithm for sound propagation outdoors : comparison with experimental measurements and with ISO-DIS 9613 standards,” *Advances in Engineering Softwares*, vol. 31, pp. 241–250, 2000.
- [118] R. Stewart and M. Sandler, “Towards an efficient interactive auralization system : evaluation of directional impulse responses,” in *IEEE Workshop on Applications of Signal Processing to Audio and Acoustics*, New Paltz, NY, 2009.
- [119] M. Barron, “Subjective Study of British Symphony Concert Halls,” *Acustica*, vol. 66, no. 1, pp. 1–14, 1988.

## Annexe A

# Annexe : Méthodes de synthèse de réponses de salles

### A.1 Méthode des sources-images

A partir d'une source ponctuelle émettant un signal quelconque dans la salle, la méthode des sources-images permet de calculer la pression en chaque point de l'espace par superposition des contributions venant de la source et de l'ensemble des sources-images dont la position est obtenue géométriquement. La réponse impulsionale entre la source et le récepteur est ainsi estimée.

#### A.1.1 Principe

La méthode des sources-images est un modèle de réflexions spéculaires qui consiste à calculer le champ sonore en un point de l'espace par superposition d'ondes sonores issues de la source et d'une série de sources fictives. Ces dernières, appelées sources-images, sont obtenues en considérant les images réfléchies de la source réelle sur les parois du volume et les images de ces images (Fig.A.1). Le calcul analytique des sources-images est ais   lorsque la g  om  trie est simple. De plus, lorsque les parois sont parfaitement r  f  l  chissantes et que la source est omnidirectionnelle, le temps de calcul n  cessaire  l'estimation de la r  ponse impulsionale est court. Le r  sultat obtenu est alors celui donn   par la th  orie des modes [4].

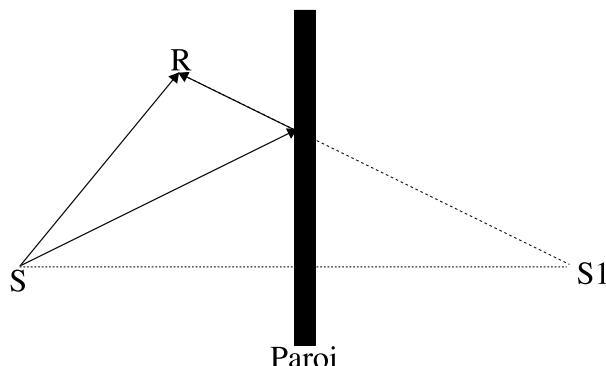


FIG. A.1 – Source-image du premier ordre ( $S_1$ ).  $S$  : point source.  $R$  point de r  ception.

Le champ sonore en un point de l'espace, r  sultant de l'  mission d'une source harmonique

de pulsation  $\omega$  s'écrit :

$$P(\vec{r}, \omega) = P_0(\omega) \times e^{-j\omega t} \sum_{i=0}^{+\infty} \frac{e^{-jkr_i}}{r_i}, \quad (\text{A.1})$$

où  $P_0(\omega)$  est la pression sonore émise par la source à la pulsation harmonique  $\omega$ , et  $r_i$  est la distance entre la source-image  $i$  et le récepteur.

L'Eq. A.2 formalise la réponse impulsionnelle entre une source et un récepteur comme une somme de potentiels retardés ou d'impulsions. Ces dernières correspondent aux contributions de chaque source-image, dont la position est égale au temps de parcours entre la source-image correspondante et le récepteur. Le niveau de chaque impulsion est inversement proportionnel à la distance parcourue, soit :

$$\vec{p}(r), t = p_0 \times \sum_{i=0}^{+\infty} \frac{\delta(t - r_i/c_0)}{r_i}, \quad (\text{A.2})$$

où  $c_0$  est la célérité du son.

Dans un volume de géométrie complexe, la position des sources-images doit être calculée numériquement. Certaines sources-images ne doivent pas être prises en compte, car elles ne contribuent pas à la réponse de la salle (Fig.A.2) [111]. De plus, le temps de calcul est fonction du nombre de sources-images, c'est-à-dire qu'il est fonction du nombre de parois de la salle et de réflexions subies par les ondes, soit :  $T_{\text{calc}} \propto M^n$ , où  $M$  est le nombre de parois et  $n$  l'ordre de sources-images.

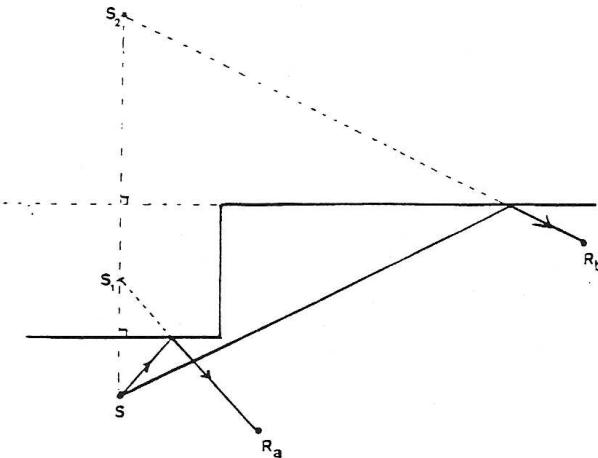


FIG. A.2 – Exemple de source-image ne contribuant pas à la réponse impulsionnelle de la salle. La source  $S_1$  contribue à la réponse au récepteur  $R_a$  mais pas à celle du récepteur  $R_b$  [38].

### A.1.2 Limites de la méthode

Ce modèle suppose que les parois sont planes, ce qui est rarement rencontré dans la réalité. Cette approche est valable pour le cas des grandes salles de spectacles si [10, 15] :

- la source est loin des parois relativement aux longueurs d'ondes considérées ;

- les parois sont planes relativement aux longueurs d'ondes considérées ;
- les longueurs d'ondes sont petites par rapport aux dimensions des parois ;
- l'atténuation aux hautes fréquences (effets viscothermiques, relaxation [112]) est négligeable.

De plus, cette approche géométrique néglige les phénomènes de diffraction que l'on peut rencontrer en incidence rasante sur le dossier des sièges, sur la fosse d'orchestre, ou encore sur le rebord d'un balcon [113, 114]. Il en va de même avec le phénomène de diffusion. Notons toutefois, que l'ajout des directivités mesurées d'une paroi engendrerait de longs temps de calculs [115]. Cependant, certains modèles tentent de synthétiser ces phénomènes [22, 23, 24, 25, 26]. Ces modèles améliorés prennent en compte la directivité de la source et du récepteur, des facteurs de réflexion des parois en introduisant un facteur de pondération des contributions des sources-images ainsi que l'angle d'incidence des ondes se réfléchissant sur la paroi.

Une approximation majeure concerne les filtres des parois qui ne tiennent pas compte de la direction d'incidence de l'onde. En effet, les matériaux sont généralement caractérisés par leur contribution moyenne suivant toutes les directions d'incidence. L'erreur introduite ici devient importante dans le cas de matériaux à faible impédance, ou multi-couche, ou non plans. Emerit [38] rappelle que cette approximation n'a que peu d'effet sur la modélisation des salles de spectacles, dès lors que celles-ci sont composées majoritairement de matériaux à forte impédance.

Enfin, comme cela a été vu au précédent paragraphe, la géométrie complexe d'une salle complique l'estimation de la réponse impulsionale, et par conséquent augmente le temps de calcul, rendant impossible le calcul de la réponse complète. La méthode des cônes tente de palier à ce problème. Cette méthode est basée sur celle du lancer de rayons.

## A.2 Méthode du lancer de rayons

### A.2.1 Principe

La méthode des rayons sonores permet également d'estimer la réponse impulsionale d'une salle en un point de l'espace, mais plus rapidement (et plus grossièrement, car elle nécessite d'introduire un volume de réception) qu'avec la méthode des source-images décrite précédemment (Section A.1) et avec des formes géométriques plus complexes.

Cette méthode est basée sur le principe de l'optique géométrique. De ce fait, c'est une approximation valable à hautes fréquences. Dans le cas d'obstacles plans, elle est équivalente à la méthode des sources-images.

La réponse synthétisée est un signal en énergie recueilli par un récepteur non-ponctuel. Il s'agit non plus de calculer des sources-images entre deux points d'une salle, mais de tirer - aléatoirement ou uniformément- un grand nombre de rayons dans l'espace et de suivre leur parcours dans l'espace en appliquant les principes de l'acoustique géométrique. Chaque fois qu'un rayon atteint le récepteur, sa contribution est accumulée. Le rayon est ensuite prolongé jusqu'à ce que son énergie soit devenue négligeable. La discrétisation de l'espace impose que le récepteur soit représenté par une sphère et non par un point. Dès lors, plus le récepteur sera près de la source, plus celui-ci aura de chance d'être atteint par un rayon. En effet, la probabilité d'atteindre le récepteur est inversement proportionnel au carré de la distance source/récepteur. La divergence géométrique se traduit donc par une décroissance de la densité de rayons qui traversent le récepteur lorsque la distance parcourue par le rayon augmente.

L'intensité reçue par le récepteur vaut :

$$I(t, \omega) = \sum_i \sum_{j=1}^{n_j} \delta(t - t_i) \times \frac{E_j(\omega)}{A}, \quad (\text{A.3})$$

où  $n_j$  est le nombre de rayons qui traversent la surface du récepteur à l'instant  $t_i$ ,  $E_j(\omega)$  est

l'énergie élémentaire porté par le rayon  $j$ , et  $A$  est la surface apparente du récepteur.

L'énergie élémentaire  $E_j(\omega)$  portée par le rayon  $j$  résulte de l'énergie totale  $W$  émise par la source, selon :

$$E_j(\omega) = \frac{W}{N} \Pi_j (1 - \alpha_j(\omega)), \quad (\text{A.4})$$

où  $N$  est le nombre de rayons,  $\alpha_j$  est un coefficient relatif à l'absorption des surfaces rencontrées par les rayons et à l'atténuation de l'air.

### A.2.2 Limites de la méthode

Le signal obtenu avec le modèle des rayons est un signal en énergie. Afin d'obtenir en signal en pression, il est possible de calculer la racine carrée du signal en énergie et de définir une phase aléatoire [36]. Cependant, ce faisant cette approche ne tient pas compte des phénomènes d'interférences des ondes sonores [38]. Enfin, au-delà d'un certain temps, la densité de rayons est tellement faible que le récepteur peut fréquemment se trouver entre deux rayons. Un effet de quantification apparaît, et l'amplitude des contributions ne tient plus compte de la divergence géométrique [116, 117, 118]. Une fois de plus, une approche statistique devient nécessaire afin de prolonger l'estimation de la réponse aux temps plus élevés.

## A.3 Méthode hybride

La méthode des cônes, développée par le C.S.T.B. [119]<sup>1</sup>, regroupe les avantages des deux méthodes précédentes sans leurs défauts. Elle consiste à calculer des sources-images ou des échogrammes grâce à une méthode du type "tir de rayons".

### A.3.1 Principe

**Calcul des sources-images** Dans ce modèle, la source émet de l'énergie à l'intérieur d'un grand nombre de cônes (Fig. A.3 et A.4), dont l'axe part du centre de la source. En assimilant chaque cône à son axe de symétrie, la propagation et la détermination des réflexions se fait de la même manière que pour la méthode des rayons (Section A.2). Lorsqu'un cône rencontre le récepteur, la position de la source-image correspondante est déterminée. Cependant, afin d'éviter toute erreur due à la largeur du cône, le tracé réel est retracé fictivement de manière rétro-propagative du récepteur à la source, dans le but de vérifier l'existence de la source-image estimée. Les sources-images sont estimées plus rapidement que par la méthode décrite dans la Section A.1. Le temps de calcul est équivalent à celui de la méthode des rayons et est proportionnel au cube du nombre maximum de réflexions ( $T_{\text{calc}} \propto n^3$ ).

Notons que la méthode des cônes est employée dans certains modèles hybrides [38] pour estimer plus précisément le nombre de sources-images qu'en utilisant simplement la méthode des sources-images (Section A.1).

**Calcul de l'échogramme** La réponse en énergie de la salle peut être obtenue de la manière suivante. La surface de la sphère qui entoure la source est discrétisée en autant d'éléments que l'on considère de cônes. Chaque axe de cône passe par le centre de la sphère. L'énergie des cônes est pondérée par une fonction en  $\cos^2$  ou une fonction de Gauss bornée [117]. La surface de la sphère est décrite par un chevauchement des cônes. L'échogramme de la réponse de la salle

---

<sup>1</sup>Ou sa variante, la méthode des pyramides, développée par Farina [120].

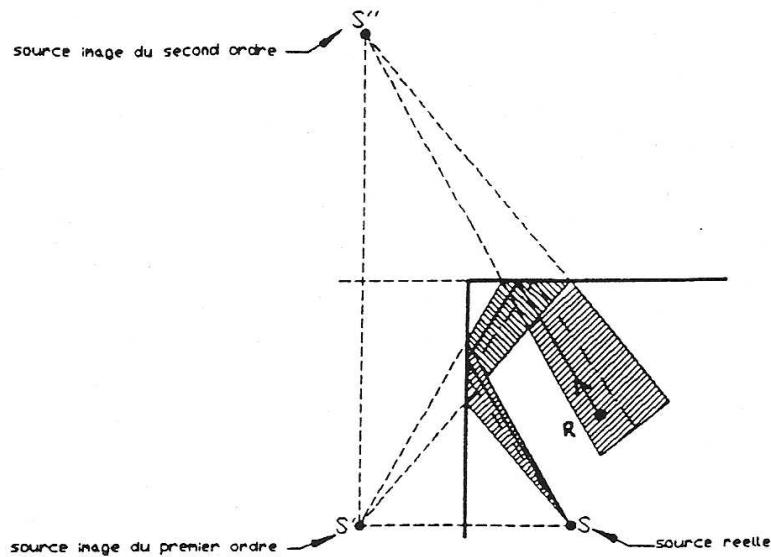


FIG. A.3 – Exemple d'estimation d'une source-image par rétro-propagation [38].

entre deux points est obtenu en sommant la contribution en énergie de chaque cône qui atteint le récepteur. Comparativement à la Section A.3.1, le nombre de sources-images obtenu par ordre de réflexion est plus précis. En effet, le nombre de sources-images manquées par ordre de réflexion est plus faible que précédemment grâce à la pondération des cônes en énergie [38].

### A.3.2 Limites de la méthode

Comme dans le cas de la méthode des rayons, la méthode des cônes discrétise l'espace, et ceci constitue un véritable inconvénient. En effet, lors de leur propagation, les cônes sont assimilés à leur axe de symétrie. De plus, l'espace entre deux axes augmente avec la distance. Par conséquent, la probabilité que le cône omette certaines images devient non-négligeable lorsqu'une paroi est suffisamment petite pour se situer entre les axes du cône. Autrement dit, après un certaine distance parcourue, l'ouverture entre les deux axes du cône peut devenir important relativement aux dimensions de la salle. Dans ce cas, le récepteur sera continuellement atteint par le cône. Ainsi, plus l'ordre de réflexions désiré est élevé, plus l'ouverture du cône doit être faible. Il en découle naturellement une augmentation du nombre de cônes et du temps de calcul. De plus, le temps de calcul est une fonction cubique de l'ordre de réflexion, ce qui entraîne des temps de calcul élevés pour des ordres élevés.

Dans la pratique, afin de calculer la réponse de la salle en un temps raisonnable, l'ordre de réflexion considéré est faible ( $\leq 10$ ). Le reste du signal est approximé par des méthodes statistiques, telles qu'un bruit blanc filtré et fenêtré en temps.

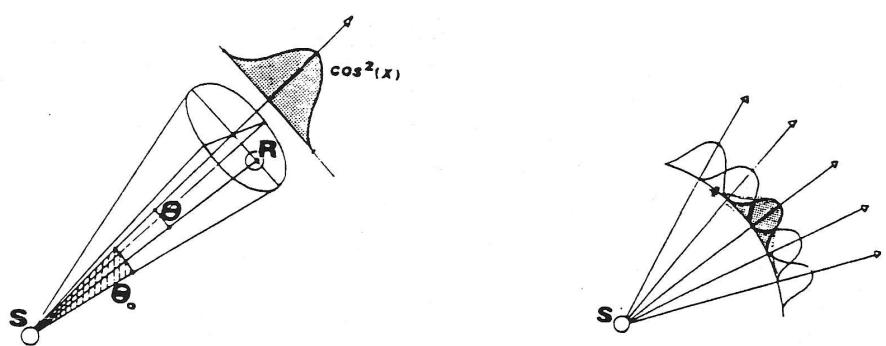


FIG. A.4 – A gauche : Cône, dont l'origine est la source  $S$ , et sa fonction de pondération en  $\cos^2$ .  
A droite : Recouvrement des cônes sur un élément de surface de la sphère entourant le point source ( $S$ ) [38].

## Annexe B

# Annexe : OpenMIDAS

### B.1 OpenMIDAS - Quick start of Processing Module

#### B.1.1 Installing OpenMIDAS package

The OpenMIDAS package is received by email. It is embedded in a ZIP archive. Unzip this archive in a folder wherever you want on your hard disk.

In the same folder, be sure to have :

- MatLab functions of OpenMIDAS (see the list below)
- a MIDAS acquisition file (for instance : 'Le Mans E1')
- a MIDAS parameter file (for instance : 'MIDAS-PARAM')

Note 1 : If you miss the acquisition file and/or the parameter file, OpenMIDAS will not work. The acquisition file contains on the one hand the impulse response measured and on the other hand some other parameters set during the acquisition. The parameter file contains some parameters set by the user during the use of MIDAS.

Note 2 : MIDAS parameter files all have the same name, that is, 'MIDAS-PARAM'.

Note 3 : In the list below appears some calibration files. These are optional : measurements can be carried out without calibration file. The existence (mentionned in the acquisition file) of the calibration file depends on the user.

Below is a list of all the MatLab functions and textfiles used by OpenMIDAS (alphabetical order) :

- banfilter.m
- compair.m
- ComputeDecayCurve.m
- ComputeDirectEnergy.m
- ComputeIIA.m
- ComputeLateralEnergy.m
- ComputeTransIndex.m
- FFA-M.m
- Full+Oct+BMT.txt
- Full+Oct+Half.txt
- Full+Oct+ Half+0-2k.txt
- Full+Oct+Third-2.txt
- Full+Oct+Third+BMT.txt
- IIA.m

- IndexComputation.m
- interpolation.m
- leastsquares.m
- Mesures MIDAS
  - Full scale (*folder*)
    - Le Mans E1 (acquisition file)
    - Le Mans Reference L1 (calibration file)
    - Le Mans U24 (acquisition file)
    - MIDAS-PARAM
  - 1/50 scale (*folder*)
    - Maqu Le Mans E1 (acquisition file)
    - Reference 4138+grid 20cm (calibration file)
    - Maqu Le Mans U24 (acquisition file)
    - MIDAS-PARAM
- MIDAS-W&R (*folder*)
  - MIDASread.m
  - MIDASwrite.m
  - MIDASwriteDefault.m
  - MIDASwritehelper.m
  - Music-Indices.txt
  - NoiseDetectionCompensat.m
  - ReadCritDescriptor.m
  - ReadCritFile.m
  - ReadParam.m
  - reducefft.m
  - SetAirComp.m
  - SetBandfilter.m
  - SetCalibration.m
  - SetConstants.m
  - setweight.m
  - SignalDecay.m
  - Speech Indices.txt
  - User-Indices.txt
  - WriteFiles.m

### B.1.2 Running OpenMIDAS

#### B.1.3 Set MatLab

Make sure you have MatLab installed on your computer (it can be a PC, Linux or a Mac (PPC or Intel)).

##### B.1.3.1 Run OpenMIDAS

If you want to calculate the acoustical indices of a file called 'Le Mans E1'. Proceede as follows :

- Launch MatLab.
- Add the OpenMIDAS files and folders path to the path of your MatLab.
- In your MatLab Command Window, type : "[OMDS] = IndexComputation('Le Mans E1');".

- Then read on the screen the different information useful to know (is there a calibration file ? is air compensation activated ? etc.) and the different steps of processing corresponding to each function of OpenMIDAS.
- When the work is done, the sentence "Write Files" appears on your screen, meaning that you can find 2 textfiles in your folder : "Le Mans E1-20-Nov-2008-17/42/50-Processing Report.txt" (*format : filename/date/hour/Processing Report*) and "Le Mans E1-20-Nov-2008-17/42/50-Processing Result.txt" (*filename/date/hour/Processing Result*). The Processing Report file exists mainly for debug, so it does not contain any crucial information for the user. While the last text file (Processing Result) is the more important, that is, it contains the acoustical indices calculated.
- The structure OMDS is created once the process is done. It contains all the parameters acquisition, calibration filters, noise detection and compensation, air compensation, etc. You will find in this super-structure as many structures as you have functions called by OpenMIDAS. Obviously, you also have access to the acoustical indices. This way you can easily access each step of the processing work.

Note that the acquisition, parameters and calibration files are called into different functions. You do not have to indicate MatLab other filename than the acquisition file.

#### B.1.4 Possible errors

Errors of computation may be due to :

- a parameter or acquisition file wrongly filled in (e.g : it is written that there exists a calibration, while there is not)
- some missing data in the parameter or acquisition file
- any mistake in the spelling of filenames
- to a wrong path of MatLab
- a missing 'MIDAS-PARAM' file
- ...

## B.2 Detailed description of the Processing module of OpenMIDAS

Two different uses can be made of the module P :

- (1)- Calculating acoustic criteria from an old MIDAS Acquisition file (typically for checking precision of the OpenMIDAS) ;
- (2)- Calculating acoustic criteria from a new acquisition file.

<b>Input</b> <i>SubStructure</i>	FUNCTION NAME	<b>Output</b> <i>SubStructure</i>
{Acquisition file; "MIDAS-PARAM" file}	$\Rightarrow \text{IndexComputation.m} \Rightarrow$ ↓	<i>Parameters</i>
{Parameters}	$\Rightarrow \text{SetConstants.m} \Rightarrow$ ↓	<i>Constants</i>
{Parameters; Constants}	$\Rightarrow \text{SetBandfilter.m} \Rightarrow$ ↓	<i>Bandfilter</i>
{Parameters}	$\Rightarrow \text{ReadCritDescriptor.m} \Rightarrow$ ↓	<i>CritDescriptor</i>
{Parameters; Constants; Bandfilter; CritDescriptor}	$\Rightarrow \text{SetCalibration.m} \Rightarrow$ ↓	<i>Calibration</i>
{Parameters}	$\Rightarrow \text{LateralEnergy.m} \Rightarrow$ ↓	<i>LateralEnergy</i>
{Parameters; Constants; Bandfilter}	$\Rightarrow \text{SetAirComp.m} \Rightarrow$ ↓	<i>AirCompensation</i>
{Parameters; Constants; Bandfilter; CritDescriptor; Calibration}	$\Rightarrow \text{ComputeDirectEnergy.m} \Rightarrow$ ↓	<i>DirectSoundEnergy</i>
{Parameters; Constants; Bandfilter; Calibration; AirCompensation}	$\Rightarrow \text{SignalDecay.m} \Rightarrow$ ↓	<i>Decays</i>
{Parameters; Constants; Bandfilter; Decays; AirCompensation}	$\Rightarrow \text{NoiseDetectionCompensation.m} \Rightarrow$ ↓	<i>NoiseComp</i>
{Constants; Bandfilter; CritDescriptor; LateralEnergy; NoiseComp}	$\Rightarrow \text{ComputeTransIndex.m} \Rightarrow$ ↓	<i>TransIndex</i>
{ Constants; Bandfilter; NoiseComp}	$\Rightarrow \text{ComputeDecayCurve.m} \Rightarrow$ ↓	<i>DecayCurve</i>
{Parameters; Constants; Bandfilter; CritDescriptor; Calibration; LateralEnergy; NoiseComp; TransIndex; DecayCurve}	$\Rightarrow \text{ComputeIIA.m} \Rightarrow$ ↓	<i>Index</i>
	Processed data are saved into a .mat file and written into a textfile named <i>ProcessingResults.txt</i>	

## B.3 Action of each brick

As seen in Section B.2, the module P is constituted by 13 bricks. The first brick, called '*IndexComputation.m*', needs two input arguments :

- (1) the acquisition file to analyse ;
- (2) the parameters file ("MIDAS-PARAM").

If this protocole is not respected, an ERROR message will pop up.  
In the following, we describe each brick of module P.

### B.3.1 IndexComputation.m

The file "MIDAS-PARAM" file and the Acquisition file are read, and data are stored into a sub-structure called *Parameters*. The parameters read are :

1. *SampFreq*. Sampling Frequency in Hz. Type : double.
2. *timebase*. Type : double.
3. *sampintval*. Interval between two samples. Type : double.
4. *nsamples*. Number of samples of the IR. Type : double.
5. *length*. Length of the IR in seconds. Type : double.
6. *filimit*. Filter low limit set during acquisition. A band is specified by two pairs of number -(low cut frequency ; low cut slope), (high cut frequency ; high cut slope) in the band descriptor file. The slopes of the filter response are set to infinity (in fact there is infinite attenuation of the signal) when the response drops below the value specified in 'low limit' (*i.e. filimit*). Type : double.
7. *dt*. Equivalent full scale time step of the acquisition, in second. Type : double.
8. *nfirst*. First sample of the IR. Type : double.
9. *nlast*. Last sample of the IR. Type : double.
10. *scale*. Scale. At full-scale, scale=1, in model scale (for instance 1/50), scale = 50.
11. *nmicro*. Number of microphones used during measurements. Type : double.
12. *press*. Pressure in Pascal during easurement. Type : double.
13. *extra\_ms*. Low time-boundary, in millisecond, for decay extrapolation, used for noise compensation. It is equivalent to the mixing time. Type : double.
14. *incr*. Increment between to adjacent windows expressed as a fraction of the window size. Type : double.
15. *winocc*. Proportion of the signal occupancy in each "slice" of the signal. Type : double.
16. *Lfms*. Lateral Energy time interval in milliseconds. This is the time interval used for computing the lateral energy fraction. Type : double.
17. *fmax*. Center frequency of the upper octave of analysis in Hz. Type : double.
18. *fmin*. Minimum frequency of analysis in Hz. Type : double.
19. *DryTemp*. Dry temperature in °C. Type : double.
20. *humid*. Relative humidity in %. Type : double.
21. *dist\_sm*. Distance, in meter, between the source and the microphone(s). Type : vector of double.
22. *filtcoef*. Coefficient for adjusting the cutoff frequency of the filter, in order to avoid aliasing. Type : double.

23. *margin*. Signal to noise margin in dB. This defines the range at the end of the decay curve on which measurements are not made because of the contamination by background noise. Type : double.
24. *mindyn*. Minimum dynamic in dB. This is the minimum dynamic range of decay required for the acoustic criteria to be computed. Type : double.
25. *caltion*. Calibration state. 1 for 'ON' ; 0 for 'OFF'. Type : logical.
26. *aircomp*. Air compensation state. 1 for 'ON' ; 0 for 'OFF'. Type : logical.
27. *doublemic*. Use of two microphones positionned as at the ears of a listener (for computing binaural indices). 1 for 'ON' ; 0 for 'OFF' (1 microphone used instead). Type : logical.
28. *hall*. Name of the hall. Type : character.
29. *sourcepos*. Source position in the hall. Type : character.
30. *micropos1*. Position of the first microphone in the hall. Type : character.
31. *micropos2*. Position of the second microphone in the hall. Type : character.
32. *mictype1*. Type of the first microphone. Type : character.
33. *mictype2*. Type of the second microphone. Type : character.
34. *acqtime*. Time of acquisition. Type : character.
35. *acqdate*. Date of acquisition. Type : character.
36. *calfile*. Calibration filename. Type : character.
37. *windowtype*. Type of windows used for windowing the signal ('Rectangular', 'Hamming', 'Hanning'). Type : character.
38. *bandfile*. This is the name of the file where the details of the frequency band used in the filtering will be read. Type : character.
39. *weitype*. Type of weighting curve ('OFF', 'Pink Curve', 'A Curve', 'Speech Curve'). Type : character.
40. *critype*. Indicates the filename of acoustic criteria to compute. Type : character.
41. *source*. Source type. Type : character.
  - Impulse (*source* = 'p'). The source is an impulse (pistol shot, balloon burst, spark,etc.) ;
  - Interrupted noise (*source* = 'i'). Sound decay is measured when interrupting a steady noise ;
  - Pseudo-Random noise (*source* = 'h'). Impulse response is computed using Maximum Length Pseudo-Random Sequence.
42. *cal\_ms*. Calibration window width, in ms. Duration of the reference signal from the calibration file. This duration is also used for testing the direct sound energy repeatability for acquisitions using an impulse response. Type : double.
43. *IR*. Measured impulse response in Pascal. Type : vector or matrix (according to *OpenMIDAS.Parameters* of double).

### **Ouput arguments : *OpenMIDAS.Parameters***

1. *OpenMIDAS.Parameters.SampFreq*. Sampling Frequency in Hz. Type : double.
2. *OpenMIDAS.Parameters.timebase*. Type : double.

3. *OpenMIDAS.Parameters.sampintval*. Interval between two samples. Type : double.
4. *OpenMIDAS.Parameters.nsamples*. Number of samples of the IR. Type : double.
5. *OpenMIDAS.Parameters.length*. Length of the IR in seconds. Type : double.
6. *OpenMIDAS.Parameters.filimit*. Filter low limit set during acquisition. Type : double.
7. *OpenMIDAS.Parameters.dt*. Equivalent full scale time step of the acquisition, in second. Type : double.
8. *OpenMIDAS.Parameters.nfirst*. First sample of the IR. Type : double.
9. *OpenMIDAS.Parameters.nlast*. Last sample of the IR. Type : double.
10. *OpenMIDAS.Parameters.scale*. Scale. At full-scale, scale=1, in model scale (for instance 1/50), scale = 50.
11. *OpenMIDAS.Parameters.nmicro*. Number of microphones used during measurements. Type : double.
12. *OpenMIDAS.Parameters.press*. Pressure in Pascal during easurement. Type : double.
13. *OpenMIDAS.Parameters.extra\_ms*. Low time-boundary, in millisecond, for decay extrapolation, used for noise compensation. It is equivalent to the mixing time. Type : double.
14. *OpenMIDAS.Parameters.incr*. This is the increment between to adjacent windows expressed as a fraction of the window size. Type : double.
15. *OpenMIDAS.Parameters.winocc*. Proportion of the signal occupancy in each "slice" of the signal. Type : double.
16. *OpenMIDAS.Parameters.Lfms*. Lateral Energy time interval in milliseconds. Type : double.
17. *OpenMIDAS.Parameters.fmax*. Center frequency of the upper octave of analysis in Hz. Type : double.
18. *OpenMIDAS.Parameters.fmin*. Minimum frequency of analysis in Hz. Type : double.
19. *OpenMIDAS.Parameters.DryTemp*. Dry temperature in °C. Type : double.
20. *OpenMIDAS.Parameters.humid*. Relative humidity in %. Type : double.
21. *OpenMIDAS.Parameters.dist\_sm*. Distance, in meter, between the source and the microphone(s). Type : vector of double.
22. *OpenMIDAS.Parametersfiltcoef*. Coefficient for adjusting the cutoff frequency of the filter, in order to avoid aliasing. Type : double.
23. *OpenMIDAS.Parameters.margin*. Signal to noise margin. Type : double.
24. *OpenMIDAS.Parameters.mindyn*. Minimum dynamic in dB. Type : double.
25. *OpenMIDAS.Parameters.caltion*. Calibration state. 1 for 'ON'; 0 for 'OFF'. Type : logical.
26. *OpenMIDAS.Parameters.aircomp*. Air compensation state. 1 for 'ON'; 0 for 'OFF'. Type : logical.
27. *OpenMIDAS.Parameters.doublemic*. Use of two microphones. 1 for 'ON'; 0 for 'OFF' (1 microphone used instead). Type : logical.
28. *OpenMIDAS.Parameters.hall*. Name of the hall. Type : character.
29. *OpenMIDAS.Parameters.sourcepos*. Source position in the hall. Type : character.
30. *OpenMIDAS.Parameters.micropos1*. Position of the first microphone in the hall. Type : character.

31. *OpenMIDAS.Parameters.micropos2*. Position of the second microphone in the hall.  
Type : character.
32. *OpenMIDAS.Parameters.mictype1*. Type of the first microphone. Type : character.
33. *OpenMIDAS.Parameters.mictype2*. Type of the second microphone. Type : character.
34. *OpenMIDAS.Parameters.acqtime*. Time of acquisition. Type : character.
35. *OpenMIDAS.Parameters.acqdate*. Date of acquisition. Type : character.
36. *OpenMIDAS.Parameters.calfile*. Calibration filename. Type : character.
37. *OpenMIDAS.Parameters.windowtype*. Type of windows used for windowing the signal ('Rectangular', 'Hamming', 'Hanning'). Type : character.
38. *OpenMIDAS.Parameters.bandfile*. Indicates the filename of the bandfilter to compute.  
Type : character.
39. *OpenMIDAS.Parameters.weitype*. Type of weighting curve ('OFF', 'Pink Curve', 'A Curve', 'Speech Curve'). Type : character.
40. *OpenMIDAS.Parameters.critype*. Indicates the filename of acoustic criteria to compute.  
Type : character.
41. *OpenMIDAS.Parameters.source*. Source type ('i', 'p', 'h'). Type : character.
42. *OpenMIDAS.Parameters.cal\_ms*. Calibration window width, in ms. Type : double.
43. *OpenMIDAS.Parameters.IR*. Measured impulse response in Pa. Type : vector or matrix  
(according to *OpenMIDAS.Parameters.nmicro*) of double.

### B.3.2 SetConstants.m

This routine calculates all constants needed for the processing, such as the number of octaves, the number of time points per FFT, the maximum and minimum frequency, etc.

**Input arguments :** *OpenMIDAS.Parameters*

1. *OpenMIDAS.Parameters.fmax*. Center frequency of the upper octave of analysis in Hz.  
Type : double.
2. *OpenMIDAS.Parameters.fmin*. Minimum frequency of analysis in Hz. Type : double.
3. *OpenMIDAS.Parameters.winocc*. Proportion of the signal occupancy in each "slice" of the signal. Type : double.
4. *OpenMIDAS.Parameters.nmicro*. Number of microphones used during measurement.  
Type : vector of double.
5. *OpenMIDAS.Parameters.nsamples*. Number of samples of the acquisition. Type : double.
6. *OpenMIDAS.Parameters.incr*. This is the increment between two adjacent windows expressed as a fraction of the window size. Type : double.
7. *OpenMIDAS.Parameters.dt*. Equivalent full scale time step of the acquisition, in second.  
Type : double.
8. *OpenMIDAS.Parameters.source*. Source type. Type : character.

- Impulse (*source* = 'p'). The source is an impulse (pistol shot, balloon burst, spark,etc.) ;
  - Interrupted noise (*source* = 'i'). Sound decay is measured when interrupting a steady noise ;
  - Pseudo-Random noise (*source* = 'h'). Impulse response is computed using Maximum Length Pseudo-Random Sequence.
9. *OpenMIDAS.Parameters.weitype*. Type of weighting curve ('OFF', 'Pink Curve', 'A Curve', 'Speech Curve'). Type : character.
  10. *OpenMIDAS.Parameters.fs*. Sampling frequency in Hz of the acquisition. Type : double.
  11. *OpenMIDAS.Parameters.aircomp*. Air compensation state. 1 for 'ON' ; 0 for 'OFF'. Type : logical.
  12. *OpenMIDAS.Parameters.caltion*. Calibration state. 1 for 'ON' ; 0 for 'OFF'. Type : logical.

#### **Ouput arguments : *OpenMIDAS.Constants***

1. *OpenMIDAS.Constants.noctave*. Number of octaves of the filter. Type : double.
2. *OpenMIDAS.Constants.lfft*. Number of time points per Fast Fourier Transform. Type : double.
3. *OpenMIDAS.Constants.nfft*. Half number of time points per Fast Fourier Transform (= *lfft*/2). Type : double
4. *OpenMIDAS.Constants.df*. Frequency resolution in Hz. Type : double.
5. *OpenMIDAS.Constants.nppw*. Number of points per window. Type : double
6. *OpenMIDAS.Constants.npts*. Number of points per channel. Type : double.
7. *OpenMIDAS.Constants.incpos*. Window increment samples. Type : double.
8. *OpenMIDAS.Constants.dwindow*. Window increment in millisecond. Type : double.
9. *OpenMIDAS.Constants.weigth*. Weighting curve. Type : vector of double.
10. *OpenMIDAS.Constants.normal*. Normalization coefficient. Type : double.
11. *OpenMIDAS.Constants.nwindow*. Number of windows used for processing the signal. Type : double.

### **B.3.3 SetBandfilter.m**

The frequency bands for which the computation is to be performed are read from the band descriptor file<sup>1</sup> (selected by the user) and the corresponding filters are computed (number of bands, central frequency, boundaries, energy and cut frequency of each band).

**Input arguments :** *OpenMIDAS.Parameters*, *OpenMIDAS.Constants*.

---

<sup>1</sup>*OpenMIDAS.Parameters.bandfile*

1. *OpenMIDAS.Parameters.filimit*. Filter low limit set during acquisition. A band is specified by two pairs of number -(low cut frequency ; low cut slope), (high cut frequency ; high cut slope) in the band descriptor file. The slopes of the filter response are set to infinity (in effect, there is infinite attenuation of the signal) when the response drops below the value specified in 'low limit' (*i.e.* *filimit*). Type : double.
2. *OpenMIDAS.Parameters.fmax*. Center frequency of the upper octave of analysis in Hz. Type : double.
3. *OpenMIDAS.Parameters.fmin*. Minimum frequency of analysis in Hz. Type : double.
4. *OpenMIDAS.Parameters.filtcoef*. Coefficient for adjusting the cutoff frequency of the filter, in order to avoid aliasing. Type : double.
5. *OpenMIDAS.Parameters.bandfile*. This is the name of the file where the details of the frequency band used in the filtering will be read. Type : character.
6. *OpenMIDAS.Constants.df*. Frequency resolution in Hz. Type : double.

#### **Output arguments :** *OpenMIDAS.Bandfilter*

1. *OpenMIDAS.Bandfilter.Type*. Type of band (*Type* : full band, octave band, third octave band, etc.). Type : character.
2. *OpenMIDAS.Bandfilter.comment*. Comments in the band descriptor file. Type : character.
3. *OpenMIDAS.Bandfilter.fcentre*. Central frequency of each band. Type : vector of double.
4. *OpenMIDAS.Bandfilter.boundary*. Boundaries of each band of the filter. Type : vector of double.
5. *OpenMIDAS.Bandfilter.filter*. Filter. Type : double.
6. *OpenMIDAS.Bandfilter.totband*. Energy per band. Type : double.
7. *OpenMIDAS.Bandfilter.cut*. Cutoff frequency of each band. Type : double.
8. *OpenMIDAS.Bandfilter.nband*. Number of bands of the filter. Type : double.

### B.3.4 ReadCritDescriptor.m

The set of indices to be computed are read from an indices descriptor file<sup>2</sup> selected by the user.

#### **Input arguments :** *OpenMIDAS.Parameters*.

1. *OpenMIDAS.Parameters.critype*. Indicates the filename of acoustic criteria to compute. Type : character.
2. *OpenMIDAS.Parameters.calction*. Calibration state. 1 for 'ON'; 0 for 'OFF'. Type : logical.

---

<sup>2</sup>*OpenMIDAS.Parameters.critype*

**Output arguments :** *OpenMIDAS.CritDescriptor*.

1. *OpenMIDAS.CritDescriptor.crit*. Acoustic criteria chosen by the user. Type : character.
2. *OpenMIDAS.CritDescriptor.ncrit*. Number of acoustic criteria to compute chosen by the user. Type : double.
3. *OpenMIDAS.CritDescriptor.RefAmpli*. Equals 1 if there is a calibration file, 0 if not. Usefulness for the calculation of the amplification. Type : logical.
4. *OpenMIDAS.CritDescriptor.STN1000*. Signal To Noise ratio at 1000 Hz. Type : double.
5. *OpenMIDAS.CritDescriptor.Ref10m*. Equals 1 for the strength calculation. Type : logical.
6. *OpenMIDAS.CritDescriptor.TIweight*. Transmission Index weight of each band. Type : double.
7. *OpenMIDAS.CritDescriptor.Ampli\_ms*. Duration reference, in millisecond, for the amplification calculus. Type : double.
8. *OpenMIDAS.CritDescriptor.criteria*. Initialization of acoustic criteria. Type : matrix of double.
9. *OpenMIDAS.CritDescriptor.critext*. Label of the acoustic criteria to compute. Type : character.

### B.3.5 SetCalibration.m

If calibration is ON<sup>3</sup> and the source is not interrupted noise<sup>4</sup>, the direct sound spectrum is computed from the calibration file<sup>5</sup>. This spectrum is truncated in the frequency domain where it is less than 30dB below its maximum. The weighting curve is divided by this spectrum. The direct sound energy per frequency band is computed. If calibration is OFF or if the source is interrupted noise, go to next function.

**Input arguments :** *OpenMIDAS.Parameters*, *OpenMIDAS.Bandfilter*, *OpenMIDAS.Constants*, *OpenMIDAS.CritDescriptor*.

1. *OpenMIDAS.Parameters.calton*. Calibration state. 1 for 'ON'; 0 for 'OFF'. Type : logical.
2. *OpenMIDAS.Parameters.doublemic*. Doublemic state. 1 for 'ON'; 0 for 'OFF'. Type : logical.
3. *OpenMIDAS.Parameters.cal\_ms*. Calibration window width, in ms. Duration of the reference signal from the calibration file. This duration is also used for testing the direct sound energy repeatability for acquisitions using an impulse response. Type : double.
4. *OpenMIDAS.Parameters.dt*. Time resolution, in second. Type : double.

---

<sup>3</sup>*OpenMIDAS.Parameters.calton = 1*

<sup>4</sup>*OpenMIDAS.Parameters.source ≠ 'h'*

<sup>5</sup>*OpenMIDAS.Parameters.calfile*

5. *OpenMIDAS.Parameters.nmicro*. Number of microphones used for measurements. Type : double.
6. *OpenMIDAS.Parameters.dist\_sm*. Distance between source and microphone. Type : vector of double.
7. *OpenMIDAS.Parameters.calfile*. Calibration filename. Type : character.
8. *OpenMIDAS.Parameters.DryTemp*. Dry temperature in °C. Type : double.
9. *OpenMIDAS.Parameters.humid*. Relative humidity in %. Type : double.
10. *OpenMIDAS.Bandfilter.nband*. Number of bands of the filter. Type : double.
11. *OpenMIDAS.Bandfilter.filter*. Filter. Type : vector of double.
12. *OpenMIDAS.Constants.weight*. Weighting curve. Type : double.
13. *OpenMIDAS.Constants.lfft*. Number of time points per Fast Fourier Transform. Type : double.
14. *OpenMIDAS.Constants.nfft*. Half number of time points per Fast Fourier Transform ( $= lfft/2$ ). Type : double.
15. *OpenMIDAS.CritDescriptor.RefAmpli*. Equals 1 if there is a calibration file, 0 if not. Usefulness for the calculation of the amplification. Type : logical.

**Output arguments :** *OpenMIDAS.Calibration*.

1. *OpenMIDAS.Calibration.y\_cal*. Calibration vector. Type : double.
2. *OpenMIDAS.Calibration.DirectE*. Direct sound energy per band. Type : vector of double.
3. *OpenMIDAS.Calibration.fscal*. Sampling frequency of the calibration file, in Hz. Type : double.
4. *OpenMIDAS.Calibration.trig\_ch*. Trigger channel. Type : double.
5. *OpenMIDAS.Calibration.weight*. Weighting curve divided by the spectrum. Type : double.
6. *OpenMIDAS.Calibration.dref*. Reference distance, in meter, between the source and the microphone(s) used for calibration. Type : double.
7. *OpenMIDAS.Calibration.ndir*. Index sample of direct sound. Type : double.

### B.3.6 ComputeLateralEnergy.m

If a lateral fraction measure is selected in the indices descriptor file (see Section B.3.4), the lateral energy fraction is computed. Otherwise, go to next function. This computation is based on the phase of the cross-spectrum between the 2 microphones in the early part of the response of the hall. This method uses the phase of the cross-spectrum of 2 omnidirectional microphones placed as the ears of a listener.

NB : Cross-spectrum and auto-spectrum are obtained from the module Acquisition.

**Input variables :** *OpenMIDAS.Parameters*.

1. *OpenMIDAS.Parameters.dt*. Equivalent full scale time step of the acquisition, in second.  
Type : double.
2. *OpenMIDAS.Parameters.gapmicro*. Distance in mleter between the two microphones.  
Type : double.
3. *OpenMIDAS.Parameters.Lfms*. Lateral Energy time interval in milliseconds. This is the time interval used for computing the lateral energy fraction. Type : double.
4. *OpenMIDAS.Parameters.fmin*. Minimum frequency of analysis in Hz. Type : double.
5. *OpenMIDAS.Constants.lfft*. Number of time points per Fast Fourier Transform. Type : double.
6. *OpenMIDAS.Bandfilter.nband*. Number of bands of the filter. Type : double.
7. *OpenMIDAS.CritDescriptor.criteria*. Matrix of room acoustical criteria. Type : matrix of double.

#### **Output parameters :**

1. *OpenMIDAS.LateralEnergy.criteria*. Matrix of room acoustical criteria. Type : matrix of double.
2. *OpenMIDAS.LateralEnergy.Lfms*. Lateral Energy time interval in milliseconds. This is the time interval used for computing the lateral energy fraction. Type : double.

**The following routines are part of a  
*Loop On Microphones*.**

### **B.3.7 SetAirComp.m**

This function is available only if the air compensation is ON<sup>6</sup>. Otherwise, go to next function. Propagation absorption is too high in models. The method chosen here is to numerically compensate for the excess air absorption. This filtering (variable unstable filter) is performed after acquisition of the impulse response. This compensation has the effect of raising the background noise at the end of the acquisition, and therefore the background noise must be compensated -according to the scale of the model and the atmospheric conditions (pressure, temperature, relative humidity).

**Input arguments :** *OpenMIDAS.Parameters*, *OpenMIDAS.Constants*, *OpenMIDAS.Bandfilter*.

1. *OpenMIDAS.Parameters.aircomp*. Air compensation state. 1 for 'ON' ; 0 for 'OFF'.  
Type : logical.
2. *OpenMIDAS.Constants.dwindow*. Window increment in millisecond. Type : double.
3. *OpenMIDAS.Constants.nfft*. Half number of time points per Fast Fourier Transform ( $= lfft/2$ ). Type : double.

---

<sup>6</sup>*OpenMIDAS.Parameters.aircomp = 1*

4. *OpenMIDAS.Constants.nwindow*. Number of windows used for processing the signal.  
Type : double.
5. *OpenMIDAS.Bandfilter.nband*. Number of bands of the filter. Type : double.
6. *OpenMIDAS.Bandfilter.boundary*. Boundaries of each frequency band of the filter. Type : vector of double.

**Output arguments :** *OpenMIDAS.AirCompensation*.

1. *OpenMIDAS.AirCompensation.air*. Air filter (*air*). Type : vector of double.
2. *OpenMIDAS.AirCompensation.comp*. Air compensation for the distance source-micophone compensation. Type : vector of double.
3. *OpenMIDAS.AirCompensation.MaxSlope*. Air filter slope ( $\text{MaxSlope} = 10 * \log_{10}(\text{air})$ ).  
Type : vector of double.

### B.3.8 ComputeDirectEnergy.m

If air compensation is OFF<sup>7</sup> and the source is not interrupted noise<sup>8</sup>, the direct sound energy per frequency band is computed. If air compensation is ON, go to next function.

**Input arguments :** *OpenMIDAS.Parameters*, *OpenMIDAS.Constants*, *OpenMIDAS.Bandfilter*, *OpenMIDAS.CritDescriptor*, *OpenMIDAS.Calibration*.

1. *OpenMIDAS.Parameters.dt*. Time resolution in second. Type : double.
2. *OpenMIDAS.Parameters.nsamples*. Number of samples of the acquisition. Type : double.
3. *OpenMIDAS.Parameters.dist\_sm*. Distance, in meter, between source and microphone. Type : double.
4. *OpenMIDAS.Parameters.nmicro*. number of microphones used for measurement. Type : double.
5. *OpenMIDAS.Constants.npts*. Number of points per channel. Type : double.
6. *OpenMIDAS.Constants.lfft*. Number of time points per Fast Fourier Transform. Type : double.
7. *OpenMIDAS.Bandfilter.nband*. Number of bands of the filter. Type : double.
8. *OpenMIDAS.Bandfilter.filter*. Filter. Type : double.
9. *OpenMIDAS.CritDescriptor.RefAmpli*. Equals 1 if there is a calibration file, 0 if not. Usefulness for the calculation of the amplification. Type : logical.
10. *OpenMIDAS.CritDescriptor.Amplim\_s*. Duration reference, in millisecond, for the amplification calculus. Type : double.

---

<sup>7</sup>*OpenMIDAS.Parameters.aricomp* = 0

<sup>8</sup>*OpenMIDAS.Parameters.source* ≠ 'h'

11. *OpenMIDAS.Calibration.weight*. Weighting curve divided by the spectrum. Type : vector of double.
12. *OpenMIDAS.Parameters.IR*. Measured impulse response in Pascal. Type : vector or matrix (according to *OpenMIDAS.Parameters.nmicro*) of double.

**Output arguments :** *OpenMIDAS.DirectSoundEnergy*

1. *OpenMIDAS.DirectSoundEnergy.DirectE*. Direct sound energy per band. Type : vector of double.
2. *OpenMIDAS.DirectSoundEnergy.ndir*. Index of direct sound. Type : double.
3. *OpenMIDAS.DirectSoundEnergy.dref*. Reference distance between the source and the microphone(s). Type : vector of double.

### B.3.9 SignalDecay.m

Using short-term FFT, the sound decay is computed for each frequency band and is compensated (if air compensation is ON<sup>9</sup>) for the air absorption according to the scale of the model and the atmospheric conditions (pressure, temperature, relative humidity). The spectra (returned by the FFT) are weighted by the weighting curve.

**Input arguments :** *OpenMIDAS.Parameters*, *OpenMIDAS.Constants*, *OpenMIDAS.Bandfilter*, *OpenMIDAS.Calibration*, *OpenMIDAS.AirCompensation*.

1. *OpenMIDAS.Calibration.weight*. Weighting curve divided by the spectrum. Type : vector of double.
2. *OpenMIDAS.Constants.normal*. Normalization coefficient. Type : double.
3. *OpenMIDAS.Constants.nfft*. Half number of time points per fft (*lfft*) divided by 2 (*nfft* = *lfft*/2). Type : double.
4. *OpenMIDAS.Constants.nppw*. Number of points per window. Type : double.
5. *OpenMIDAS.Constants.nwindow*. Number of windows used for processing the signal. Type : double.
6. *OpenMIDAS.Constants.incpos*. Window increment samples. Type : double.
7. *OpenMIDAS.Parameters.nmicro*. Number of microphones used during measurement. Type : double.
8. *OpenMIDAS.Parameters.nsamples*. Number of samples of the impulse response. Type : double.
9. *OpenMIDAS.Parameters.incr*. Increment between two adjacent windows expressed as a fraction of the window size. Type : double.

---

<sup>9</sup>*OpenMIDAS.Parameters.aircomp* = 1

10. *OpenMIDAS.Parameters.aircomp*. Air compensation state. 1 for 'ON'; 0 for 'OFF'. Type : logical.
11. *OpenMIDAS.AirCompensation.air*. Air filter. Type : vector of double.
12. *OpenMIDAS.AirCompensation.comp*. Air compensation for the distance source-micophone compensation. Type : vector of double.
13. *OpenMIDAS.Bandfilter.filter*. Filter. Type : double.
14. *OpenMIDAS.Bandfilter.nband*. Number of bands of the filter. Type : double.
15. *OpenMIDAS.Bandfilter.boundary*. Boundaries of each frequency band of the filter. Type : vector of double.
16. *OpenMIDAS.Parameters.IR*. Measured impulse response in Pascal. Type : vector or matrix (according to *OpenMIDAS.Parameters.nmicro*) of double.

**Output arguments :** *OpenMIDAS.Decays*.

1. *OpenMIDAS.Decays.Decays*. Sound decay computed for each frequency band. Type : vector of double.
2. *OpenMIDAS.Decays.LastWindow*. Window for calculating the energy of the IR "beyond" the last window (*LastWindow* = *nwindow* + 1). Type : double.

### B.3.10 NoiseDetectionCompensation.m

For each frequency band the point in the decay curve where it reaches the background noise is determined and the decay curves are truncated at the chosen signal-to-noise margin above this point.

**Input arguments :** *OpenMIDAS.Parameters*, *OpenMIDAS.Constants*, *OpenMIDAS.Bandfilter*, *OpenMIDAS.Decays*, *OpenMIDAS.AirComp*.

1. *OpenMIDAS.Parameters.extra\_ms*. Low time-boundary, in millisecond, for decay extrapolation, used for noise compensation. It is equivalent to the mixing time. Type : double.
2. *OpenMIDAS.Parameters.margin*. Signal to noise margin in dB. This defines the range at the end of the decay curve on which measurements are not made because of the contamination by background noise. Type : double.
3. *OpenMIDAS.Constants.dwindow* : window increment in millisecond. Type : double.
4. *OpenMIDAS.Constants.nwindow*. Number of windows used for processing the signal. Type : double.
5. *OpenMIDAS.Bandfilter.nband*. Number of bands of the filter. Type : double.
6. *OpenMIDAS.Decays.decays*. Sound decay computed for each frequency band. Type : vector of double.
7. *OpenMIDAS.AirCompensation.MaxSlope*. Air filter slope ( $MaxSlope = 10 * \log_{10}(air)$ ). Type : vector of double.

**Output arguments :** *OpenMIDAS.NoiseComp*

1. *OpenMIDAS.NoiseComp.tonoise*. For each band, indicates in which window noise has been detected. Type : vector of double.
2. *OpenMIDAS.NoiseComp.som*. Energy of the IR, i.e. ( $\int_0^\infty IR^2(t)dt$ ). Type : vector of double.
3. *OpenMIDAS.NoiseComp.Tsom*. Time weighted energy, i.e. ( $\int_0^\infty t.IR^2(t)dt$ ). Type : vector of double.
4. *OpenMIDAS.NoiseComp.decays*. Sound decay. Type : matrix of double.  
(Dimensions : *OpenMIDAS.Constants.nwindow*  $\times$  *OpenMIDAS.Bandfilter.nband*).

### B.3.11 ComputeTransIndex.m

If the Transmission Index is selected in the relevant indices descriptor file, the modulation transfer function (MTF) is computed in octave bands and the transmission index calculated. The RASTI is then calculated from the octave band values and stored as the full-band TI.

**Input arguments :** *OpenMIDAS.Constants*, *OpenMIDAS.Bandfilter*, *OpenMIDAS.CritDescriptor*, *OpenMIDAS.LateralEnergy*, *OpenMIDAS.NoiseComp*.

1. *OpenMIDAS.CritDescriptor.STN1000*. Signal To Noise ratio at 1000Hz, in dB. Type : double.
2. *OpenMIDAS.CritDescriptor.TIweight*. Transmission Index weight of each band. Type : double.
3. *OpenMIDAS.Constants.nwindow*. Number of windows used for processing the signal. Type : double.
4. *OpenMIDAS.Bandfilter.nband*. Number of bands of the filter. Type : double.
5. *OpenMIDAS.Bandfilter.Type*. Type of frequency band( 'octave', third octave', etc.). Type : double.
6. *OpenMIDAS.Bandfilter.fcentre*. Band frequency centre in Hz. Type : vector of double.
7. *OpenMIDAS.NoiseComp.tonoise*. For each band, indicates in which window noise has been detected. Type : vector of double.
8. *OpenMIDAS.NoiseComp.decays*. Sound decay. Type : matrix of double (Dimensions : *OpenMIDAS.Constants.nwindow*  $\times$  *OpenMIDAS.Bandfilter.nband*).
9. *OpenMIDAS.LateralEnergy.criteria*. Calculated acoustical criteria. Type : matrix of double.  
(Dimensions : *OpenMIDAS.Parameters.nmicro*  $\times$  *OpenMIDAS.Bandfilter.nband*  $\times$  *OpenMIDAS.CritDescriptor.ncrit*).

**Output arguments :** *OpenMIDAS.TransIndex*

1. *OpenMIDAS.TransIndex.criteria* : *criteria*. Type : matrix of double.  
(Dimensions : *OpenMIDAS.Parameters.nmicro* × *OpenMIDAS.Bandfilter.nband* × *OpenMIDAS.CritDescriptor.ncrit*).

### B.3.12 ComputeDecayCurve.m

The decay curves are backwards integrated (Schroeder integration), according to the international standard ISO 3382 [41].

**Input arguments :** *OpenMIDAS.Constants*, *OpenMIDAS.Bandfilter*, *OpenMIDAS.NoiseComp*.

1. *OpenMIDAS.Bandfilter.nband*. Number of bands of the filter. Type : double.
2. *OpenMIDAS.Constants.nwindow*. Number of windows used for processing the signal. Type : double.
3. *OpenMIDAS.NoiseComp.decays*. Sound decay. Type : matrix of double (Dimensions : *OpenMIDAS.Constants.nwindow* × *OpenMIDAS.Bandfilter.nband*).
4. *OpenMIDAS.NoiseComp.som*. Energy of the IR, i.e. ( $\int_0^\infty IR^2(t)dt$ ). Type : vector of double.

**Output arguments :** *OpenMIDAS.DecayCurve*.

1. *OpenMIDAS.DecayCurve.RIC* : Backward Integrated Curve. Type : matrix of double.  
(Dimensions : *OpenMIDAS.Constants.nwindow* × *OpenMIDAS.Bandfilter.nband*).
2. *OpenMIDAS.DecayCurve.LogRIC* LogRIC (=  $10 \log_{10}(RIC)$ ). Type : matrix of double.  
(Dimensions : *OpenMIDAS.Constants.nwindow* × *OpenMIDAS.Bandfilter.nband*).

### B.3.13 ComputeIIA.m

Acoustic indices selected from the criteria descriptor file<sup>10</sup> are calculated from decay curves.

**Input arguments :** *OpenMIDAS.Parameters*, *OpenMIDAS.Constants*, *OpenMIDAS.Bandfilter*, *OpenMIDAS.CritDescriptor*, *OpenMIDAS.Calibration*, *OpenMIDAS.LateralEnergy*, *OpenMIDAS.NoiseComp*, *OpenMIDAS.TransIndex*, *OpenMIDAS.DecayCurve*.

---

<sup>10</sup>*OpenMIDAS.Parameters.critype*

1. *OpenMIDAS.Parameters.mindyn*. Minimum dynamic in dB. Type : double.
2. *OpenMIDAS.Constants.dwindow*. Window increment in millisecond. Type : double.
3. *OpenMIDAS.Constants.nwindow*. Number of windows used for processing the signal. Type : double.
4. *OpenMIDAS.Bandfilter.nband*. Number of bands of the filter chosen by the user. Type : double.
5. *OpenMIDAS.CritDescriptor.crittext*. Label of the acoustic criteria to compute. Type : character.
6. *OpenMIDAS.CritDescriptor.ncrit*. Number of acoustic criteria chosen by the user. Type : double.
7. *OpenMIDAS.CritDescriptor.Ref10m*. Equals 1 for the strength calculation. Type : logical.
8. *OpenMIDAS.Calibration.dref*. Reference distance, in meter, between the source and the microphone(s) used for calibration. Type : vector of double.
9. *OpenMIDAS.Calibration.DirectE*. Energy of the direct sound. Type : vector of double.
10. *OpenMIDAS.Calibration.ndir*. Index of the sample of the direct sound. Type : double.
11. *OpenMIDAS.LateralEnergy.Lfms*. Lateral Energy time interval in milliseconds. This is the time interval used for computing the lateral energy fraction. Type : double.
12. *OpenMIDAS.NoiseComp.tonoise*. Indicates in which band noise has been detected. Type : vector of double.
13. *OpenMIDAS.NoiseComp.som*. Energy of the IR, i.e.  $(\int_0^\infty IR^2(t)dt)$ . Type : vector of double.
14. *OpenMIDAS.NoiseComp.Tsom*. Time weighted energy, i.e.  $(\int_0^\infty t.IR^2(t) \times dt)$ . Type : vector of double.
15. *OpenMIDAS.TransIndex.criteria*. Calculated acoustic criteria. Type : matrix of double. (Dimensions : *OpenMIDAS.Constants.nwindow* × *OpenMIDAS.Bandfilter.nband* × *OpenMIDAS.CritDescriptor.ncrit*).
16. *OpenMIDAS.DecayCurve.RIC*. Reverse Integrated Curve. Type : matrix of double. (Dimensions : *OpenMIDAS.Constants.nwindow* × *OpenMIDAS.Bandfilter.nband*).

#### **Output arguments : *OpenMIDAS.Index***

1. *OpenMIDAS.Index.criteria*. Computed acoustic criteria. Type : matrix of double. (Dimensions : *OpenMIDAS.Parameters.nmicro* × *OpenMIDAS.Bandfilter.nband* × *OpenMIDAS.CritDescriptor.ncrit*).

**End of Loop On Microphones.**

## B.4 Functions needed for computation

This section lists all MatLab functions needed by module P, including those described in ??.

1. IndexComputation.m
2. SetConstants.m
3. SetBandFilter.m
4. ReadCritDescriptor.m
5. SetCalibration.m
6. ComputeLateralEnergy.m
7. SetAirComp.m
8. ComputeDirectEnergy.m
9. SignalDecay.m
10. NoiseDetectionCompensation.m
11. ComputeTransIndex.m
12. ComputeDecayCurve.m
13. ComputeIIA.m

Other functions :

14. MIDASread.m
15. bandfilter.m
16. ReadCritFile.m
17. compair.m
18. interpolation.m
19. leastsquaresM.m
20. reducefft.m
21. setweight.m
22. IIA.m

### **MIDASread.m**

Called by *IndexComputation.m* (see Section B.3.1). It takes one input argument. This is the file-name of the acquisition, in MIDAS format. *MIDASread.m* outputs one structure incorporating all acquisition parameters useful for the first brick of module P, and also for the lecture of the calibration file (if calibration is ON<sup>11</sup>).

### **bandfilter.m**

Called by *SetBandfilter.m* (see Section B.3.3). This function returns the energy of the spectrum (variable *summ*) in band, defined by the filter (*filter*) and the array *cut*.

### **ReadCritFile.m**

Called by *ReadDescriptor.m* (see Section B.3.4). This function basically reads the textfile chosen

---

<sup>11</sup>*OpenMIDAS.Parameters.calton = 1*

by the user and corresponding to the acoustic criteria to calculate.

#### **compair**

Called by *SetAirComp.m* (see Section B.3.7). This function compensates for the air absorption.

#### **interpolation.m**

Called by *IIA.m*. This function performs a linear interpolation based on points  $(x1,y1)$  and  $(x2,y2)$ .

#### **leastsquaresM.m**

Called by *ComputeIIA.m* (see B.3.13). This function performs a linear regression by the least square method.

#### **reducefft.m**

This function reduces an array (after an FFT calculation) from  $lfft2$  to  $lfft1$  points, conserving the energy of the signal.

#### **setweight.m**

This function affects to the array *weight* the curve defined by the variable *weitype*.

#### **IIA.m**

Called by *ComputeIIA.m* (see Section B.3.13). This function returns the value of the Integrated Impulse Answers (IIA) curve of the bandfilter. Also calls *interpolation.m*.

#### **NB :**

Several variables, derived from Section B.3.1 are set as global variables, such as :

*DryTemp, dist\_sm, humid, fs, HumidRef, TempRef, press, scale, PressRef, cut, boundary, filimit*



## Annexe C

# Annexe : Estimation des retours et du temps de mélange à partir de deux corpus de réponses de salles

Cette Annexe présente les temps de mélange estimés dans deux salles de l'Université Queen Mary de Londres, au Royaume-Uni. Pour chaque salle, les réponses ont été mesurées par Rebecca Stewart *et al.* et les valeurs des indices acoustiques ont été présentés dans [121].

Deux salles différentes ont été considérées. La première est une salle (Great Hall) de représentation, servant aux conférences ou aux petites pièces de théâtres. Le Great Hall peut accueillir jusqu'à 800 personnes et possède un balcon qui fait face à la scène. Ses dimensions sont approximativement de  $18m \times 18m^1$ . Au total 169 réponses sont mesurées dans la salle. La seconde salle possède une géométrie particulière (elle est octogonale et mesure 7.5m de côté et 23m de long) et est assez haute (21m). Son volume est estimé à  $9500m^3$ . Au total, 169 réponses sont mesurées dans chaque salle, selon un pavage régulier au sol : une mesure est réalisée chaque mètre selon un axe  $x$  et un axe  $y$  (Fig. C.1<sup>2</sup>). Le récepteur est un microphone omnidirectionnel, et la source un sinus glissant<sup>3</sup> émis par un haut-parleur omnidirectionnel.

L'estimation des retours est réalisée grâce à Matching Pursuit (Chapitre 5) sur chacune des réponses des deux salles. Le dictionnaire d'atomes est constitué uniquement du son direct, translaté dans le temps. Les mesures que nous considérons ici présentent l'avantage que la source est plus reproductible qu'une impulsion de revolver. Le critère d'arrêt  $SRR$  de la poursuite est estimé de la même manière que dans le Chapitre 5 : nous nous basons sur la comparaison entre les indices acoustiques du signal original et ceux du signal reconstitué après poursuite pour différentes valeurs du  $SRR$ . Dans les deux cas, nous choisissons un critère d'arrêt de  $SRR = 6dB$ .

La Fig. C.2 présente les temps de mélange estimés en fonction de la distance source/récepteur. Comme nous l'avons déjà montré (Chapitre 5) la relation qui lie ces deux grandeurs est linéaire et est directement reliée au temps d'arrivée du son direct au point de mesure. Les temps de mélange que nous estimons pour chacune des salles montre qu'elle est immédiatement, ou presque, diffuse.

La Fig. C.3 représente les temps de mélange dans chaque salle comme des fonctions de l'espace. Lorsque le temps de propagation n'est pas pris en compte (dans ce cas, nous ne pouvons pas rigoureusement parler de temps de mélange), le mélange s'établit à peu près en même temps dans l'espace, et ce dans chacune des deux salles quelque soit le lieu d'observation. Les quelques

<sup>1</sup>Le volume de la salle n'est pas communiqué dans l'acte de congrès [121].

<sup>2</sup>Nous remercions Rebecca Stewart de nous avoir communiqué ces schémas.

<sup>3</sup>Les fréquences balayées ne sont pas mentionnées.

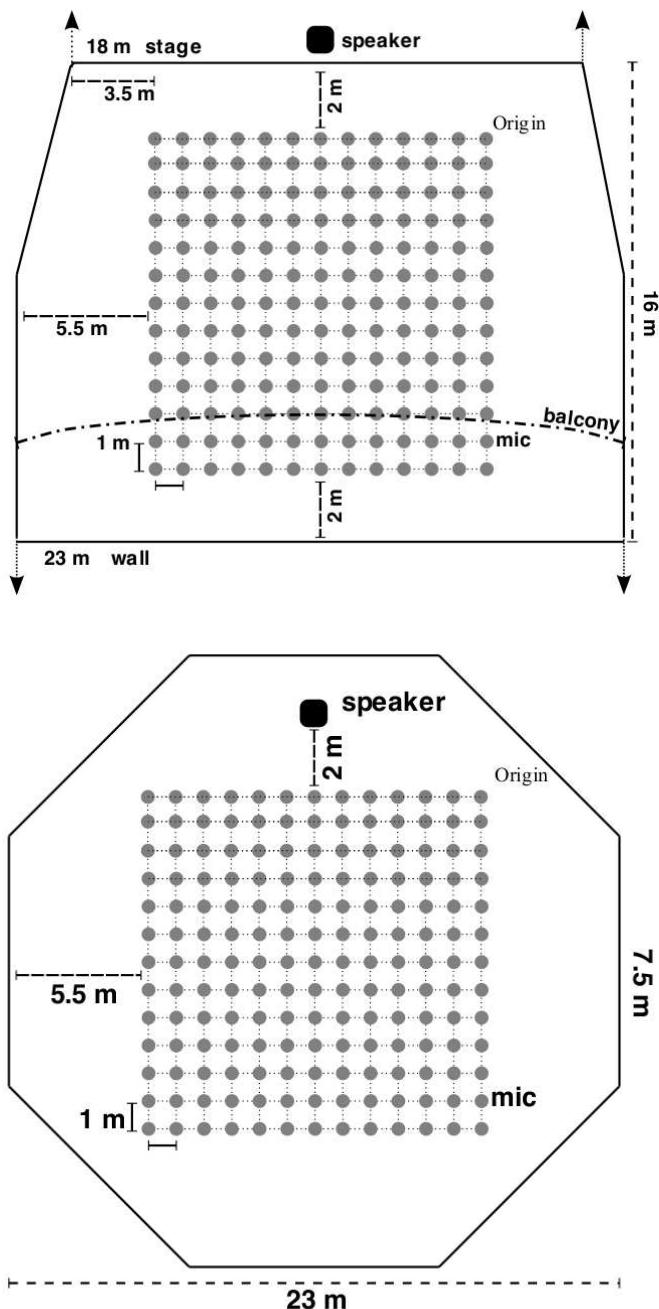


FIG. C.1 – Haut : schéma de la salle Great Hall à l'Université Queen Mary de Londres (Royaume-Uni). Bas : schéma de la salle Octagon à l'Université Queen Mary de Londres (Royaume-Uni). Les points indiquent les emplacements du récepteur (microphone omnidirectionnel).

disparités du temps de mélange (sans le temps de propagation) qui existent pour chaque salle sont de  $4ms$  environ pour Great Hall et de  $1ms$  environ pour la salle Octagon. Relier la qualité d'écoute dans une salle à ce temps auquel le système devient diffus localement pourrait constituer l'objet de futures recherches.

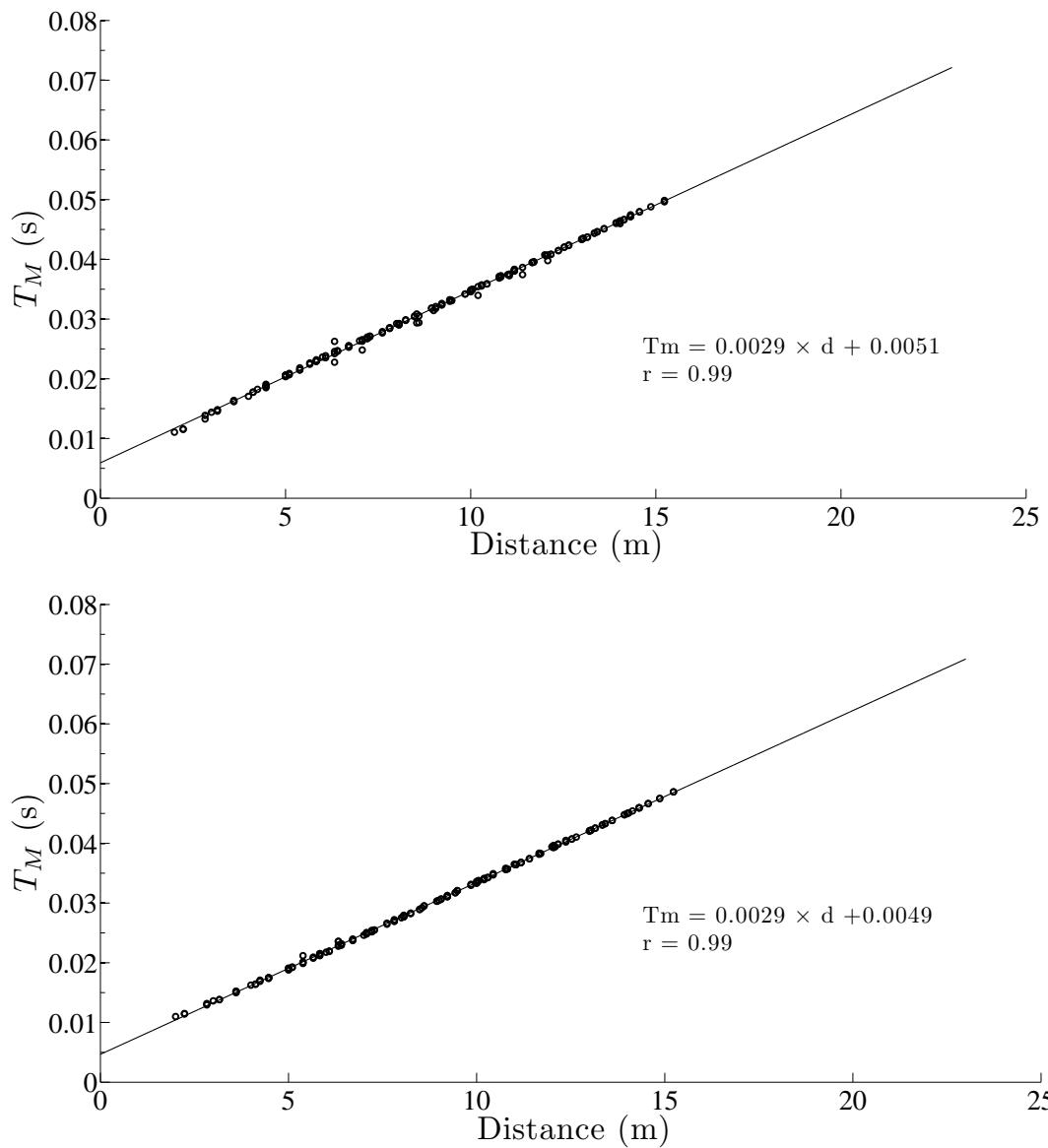


FIG. C.2 – Cercles : temps de mélange estimés sur les salles Great Hall (haut) et Octagon (bas) ( $SRR = 6dB$ ). Traits pleins : régressions linéaires.

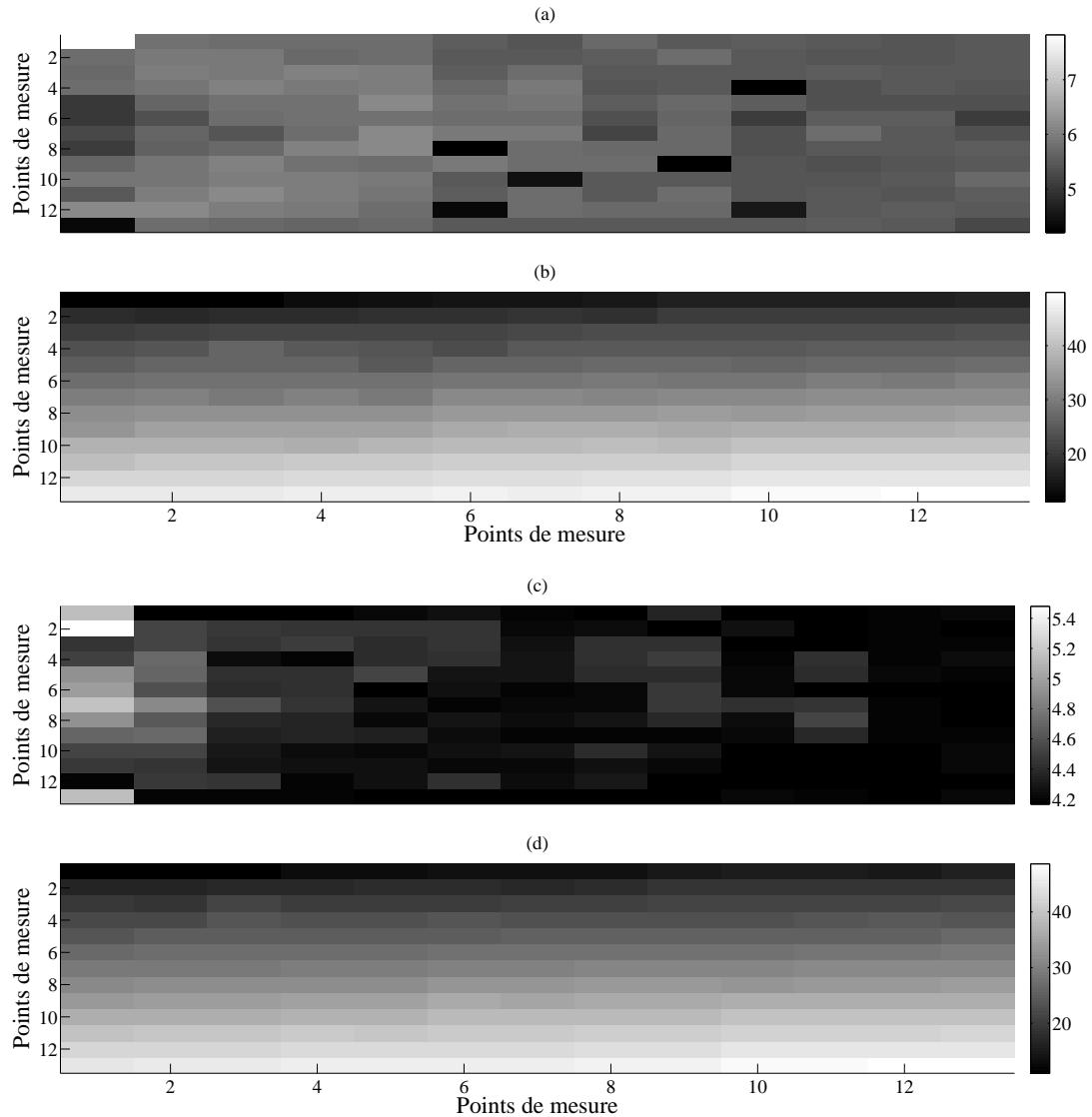


FIG. C.3 – Temps de mélange (en  $ms$ ) estimés sans (a et c) et avec (b et d) la distance de propagation source/récepteur. Deux figures du haut (a) et (b) : Great Hall. Deux figures du bas (c) et (d) : Octagon. ( $SRR = 6dB$ ).

## Annexe D

# Annexe : Tests de Gaussianité appliqués à des réponses impulsionales Gaussiennes

Le Chapitre 4 montre en quoi des tests de Gaussianité tels que le test du  $\chi^2$ , de Lilliefors ou de Kolmogorov-Smirnov ne constituent pas des outils satisfaisants pour détecter le moment auquel le signal de la réponse d'une salles devient distribué normalement. Dans cette Annexe, nous appliquons les mêmes estimateurs à des réponses impulsionales de salles définies par un signal Gaussien exponentiellement décroissant.

Les mêmes remarques que celles émises dans le Chapitre 4 s'appliquent ici : les résultats dépendent de la taille de la fenêtre d'analyse (Fig. D.1-D.3). Ainsi, même un signal gaussien n'est pas estimé comme tel si le nombre d'échantillons considéré est soit trop faible, soit trop élevé.

L'entropie d'une réponse impulsionnelle modélisée par un bruit blanc Gaussien exponentiellement décroissante dont la décroissance est compensée est calculée pour deux tailles de fenêtres différentes (Fig. D.4). Là encore, la le lien entre les résultats et la taille de la fenêtre d'analyse est flagrant.

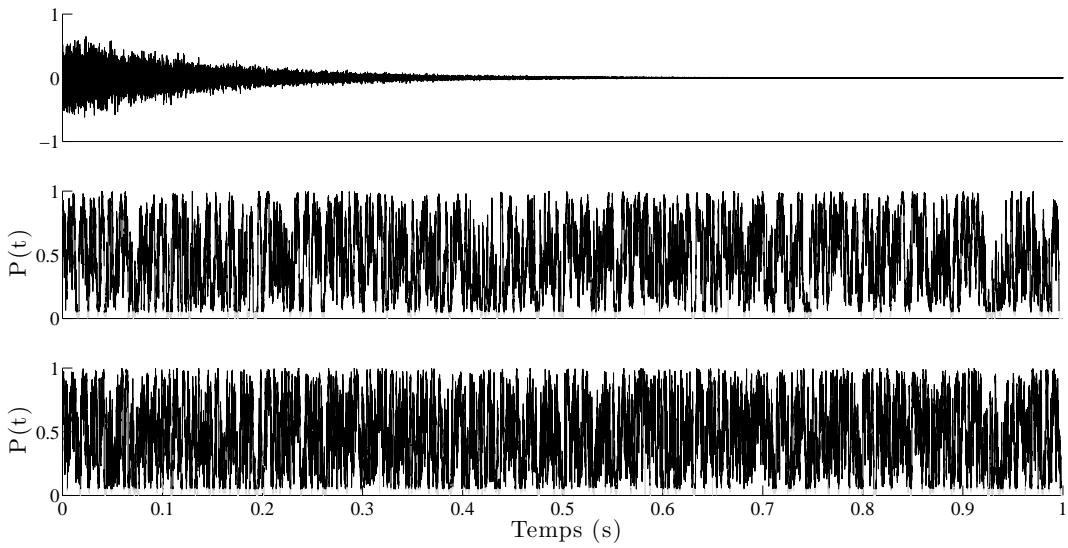


FIG. D.1 – Haut : réponse impulsionale modélisée par un bruit blanc Gaussien (fréquence d'échantillonnage de  $F_s = 48kHz$ ). Milieu : probabilité de validité de l'hypothèse nulle du test de  $\chi^2$  sur 1 fenêtre de taille 200 échantillons. Bas : probabilité de validité de l'hypothèse nulle du test de  $\chi^2$  sur 1 fenêtre de taille 120 échantillons. Note : en noir, les instants auxquels  $P(t) \geq 0.05$ . Remarquons que la validation de l'hypothèse nulle dépend de la taille de la fenêtre d'analyse.

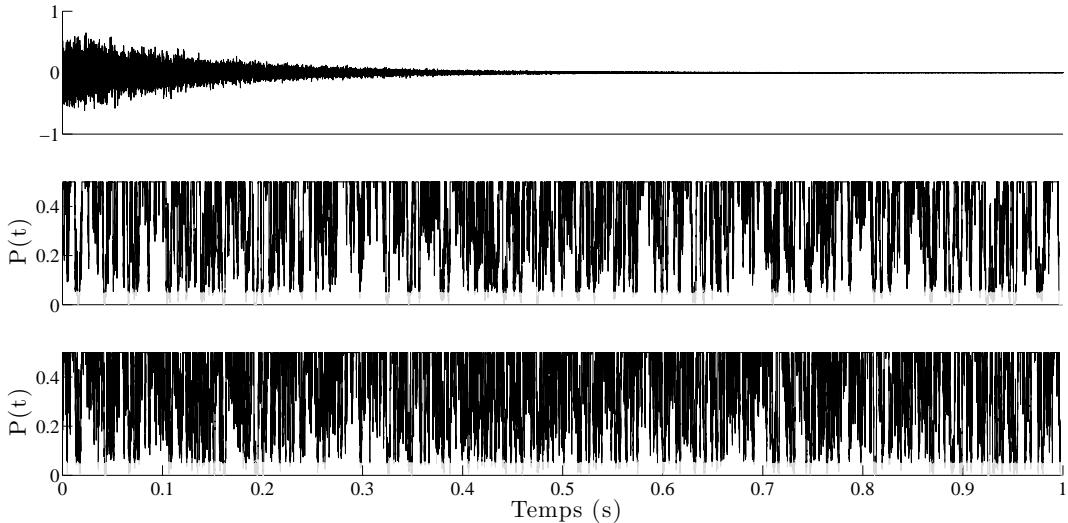


FIG. D.2 – Haut : réponse impulsionale modélisée par un bruit blanc Gaussien (fréquence d'échantillonnage de  $F_s = 48kHz$ ). Milieu : probabilité de validité de l'hypothèse nulle du test de Lilliefors sur des fenêtres de taille 200 échantillons. Bas : probabilité de validité de l'hypothèse nulle du test de Lilliefors sur des fenêtres de taille 120 échantillons. Note : en noir, les instants auxquels  $P(t) \geq 0.05$ . Remarquons que la validation de l'hypothèse nulle dépend de la taille de la fenêtre d'analyse.

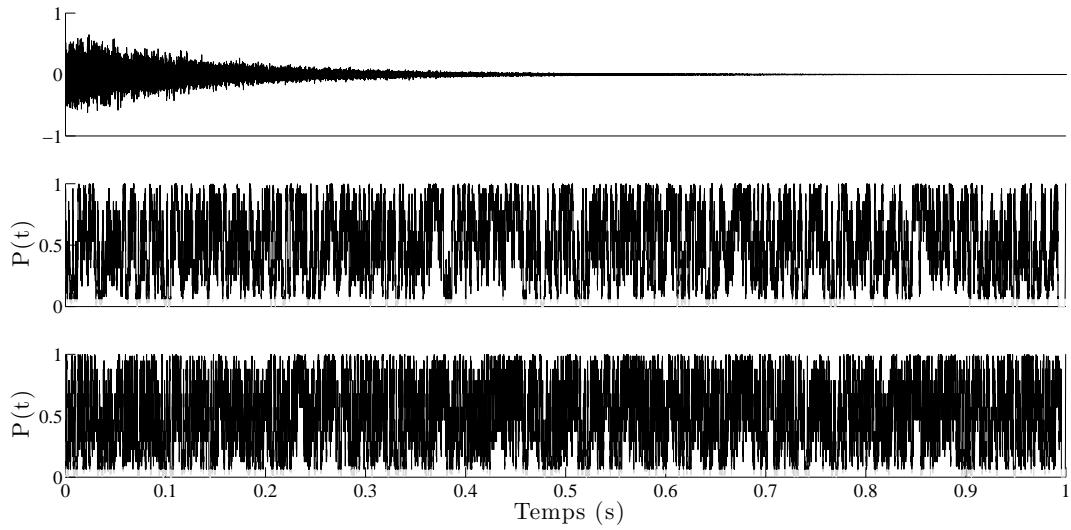


FIG. D.3 – Haut : réponse impulsionnelle modélisée par un bruit blanc Gaussien (fréquence d'échantillonnage de  $F_s = 48kHz$ ). Milieu : probabilité de validité de l'hypothèse nulle du test de Kolmogorov-Smirnov sur 2 fenêtres de taille 200 échantillons. Bas : probabilité de validité de l'hypothèse nulle du test de Kolmogorov-Smirnov sur 2 fenêtres de taille 120 échantillons. Note : en noir, les instants auxquels  $P(t) \geq 0.05$ . Remarquons que la validation de l'hypothèse nulle dépend de la taille de la fenêtre d'analyse.

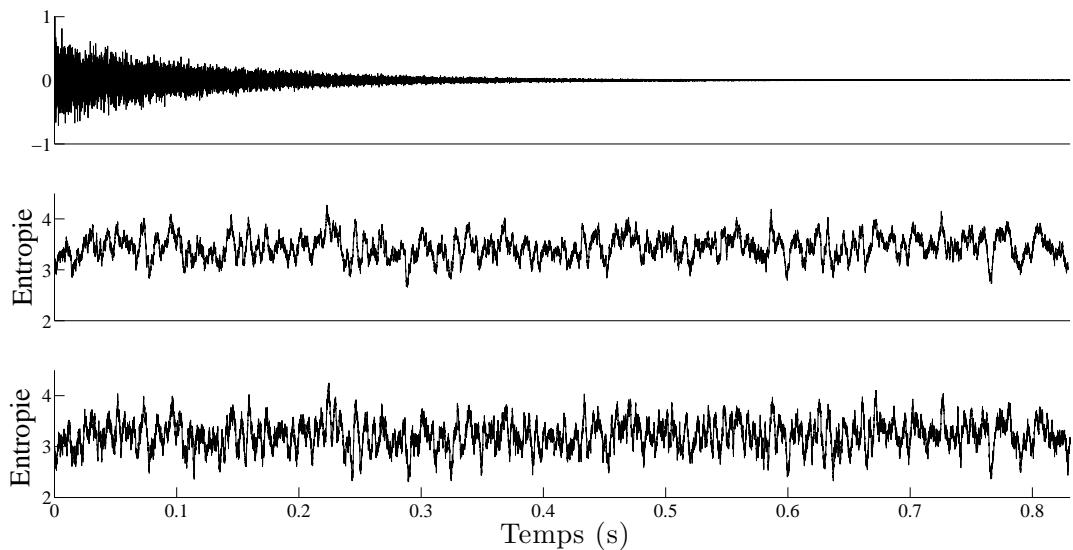


FIG. D.4 – Haut : réponse impulsionnelle modélisée par un bruit blanc Gaussien (fréquence d'échantillonnage de  $F_s = 48kHz$ ). Milieu : entropie calculée sur des fenêtres de taille 200 échantillons. Bas : entropie calculée sur des fenêtres de taille 120 échantillons. Remarquons que la valeur de l'entropie dépend de la taille de la fenêtre d'analyse.



## Annexe E

# Annexe : Evaluation perceptive du modèle stochastique

### E.1 Introduction

J-D. Polack montre que le modèle stochastique qu'il a développé [27] est valide du point de vue des indices utilisés pour caractériser objectivement l'acoustique de salles [7]. Nous avons montré que le modèle est également validé au regard de la statistique des retours [33] (Chapitre 5) et de l'évolution temporelle/fréquentielle de la réponse [106] (Chapitre 6). Afin de procéder à une validation complète, il nous faut maintenant nous intéresser à la validité perceptive du modèle. Cette évaluation perceptive passe dans un premier temps par le choix d'un test d'écoute, et dans une second temps par le choix des paramètres du modèle que nous souhaitons tester.

### E.2 Choix du test

L'idée initiale de l'évaluation perceptive du modèle reposait sur la comparaison de réponses impulsionales de salles de concerts enregistrées et de réponses synthétisées pour les paramètres des salles considérées. Cependant, le nombre de réponses mesurées dans des salles différentes avec un protocole expérimental défini et suivi rigoureusement était trop faible au moment des tests perceptifs. Nous avons donc choisi d'orienter ces tests vers une validation écologique du modèle. En d'autres termes, l'acoustique des salles synthétisées correspond-elle à l'expérience des sujets ? Un moyen efficace d'évaluer cette validité est l'emploi du test de M. Barron [122] largement répandu dans le domaine de l'acoustique des salles et communément admis comme pertinent.

Le test de Barron propose au sujet d'évaluer différents critères lors de l'écoute d'une pièce musicale dans une salle. Ces critères sont regroupés dans le tableau E.3. Nous rajoutons à ces critères l'évaluation du réalisme de la salle synthétisée. Chaque critère est évalué sur une échelle non graduée mais bornée (Tableau E.3, Fig. E.1). Toutes les échelles ont la même dimension. De plus, le sujet a la possibilité de réagir librement en inscrivant un commentaire au sujet de l'acoustique de la salle et de leur ressenti.

Par la suite, nous énumérons les différents critères du questionnaire de Barron (auquel est rajouté l'évaluation du réalisme) :

1. **Précision** : se rapporte à la possibilité de percevoir les détails de la musique.
2. **Réverbérance** : se rapporte au degré de la réverbération perçue de la salle.

3. **Enveloppement** : se rapporte à l'aspect spatial du son perçu, à la sensation d'être entouré par le son.
4. **Intimité** : se rapporte au degré d'identification avec l'exécution , si l'on se sent auditivement impliqué ou détaché.
5. **Intensité** : se rapporte à ce que l'on considère comme acceptable pour la taille des sources sonores.
6. **Equilibre tonal** : décrit l'impression relative des graves et des aigus par rapport aux médiums.
7. **Expérience** : se rapporte à la sensation auditive générale. Les précédents critères sont inclus dans celui-ci, mais ici il s'agit d'une impression globale. L'auditeur est invité à baser son jugement sur ses propres expériences de concerts dans des salles de spectacles. La salle qu'il entend est-elle proche ou non de ses expériences ?

Critères	De...	... à
Précision	confus	précis
Réverbérance	court	long
Enveloppement	large	étriqué
Intimité	distant	proche
Intensité	fort	faible
Equilibre tonal	trop	pas assez
Selon votre expérience	éloigné	proche

TAB. E.1 – Critères du questionnaire de M.Barron [122] et les bornes d'évaluation.

### E.3 Choix des paramètres du modèle

Le Chapitre 2.5 présente le modèle stochastique tel qu'il a été élaboré par J-D. Polack, et les différentes *améliorations*<sup>1</sup> proposées : ajout de la diffusion et du filtrage de la salle. Rappelons ici les différents paramètres du modèle :

1. le volume de la salle en  $m^3$  ;
2. la distance source/récepteur en  $m$  ;
3. le temps de réverbération moyen à  $30dB$  en  $s$  ;
4. l'absorption moyenne en % ;
5. la synthèse de la diffusion ;
6. la synthèse du filtrage de la salle (trois familles de filtres sont proposées) ;
7. les angles d'incidence des rayons sonores au point de réception.

Au regard de cette liste de paramètres, nous comprenons immédiatement que le nombre de combinaisons possibles peut vite devenir important sachant que l'on peut attribuer un grand nombre de valeurs (ou modalités) aux paramètres. Il nous faut donc réduire l'espace des possibles.

---

<sup>1</sup>C'est sous ce nom que nous les avons présentées au Chapitre 2.5. Le présent Chapitre montre la pertinence perceptive d'un tel nom.

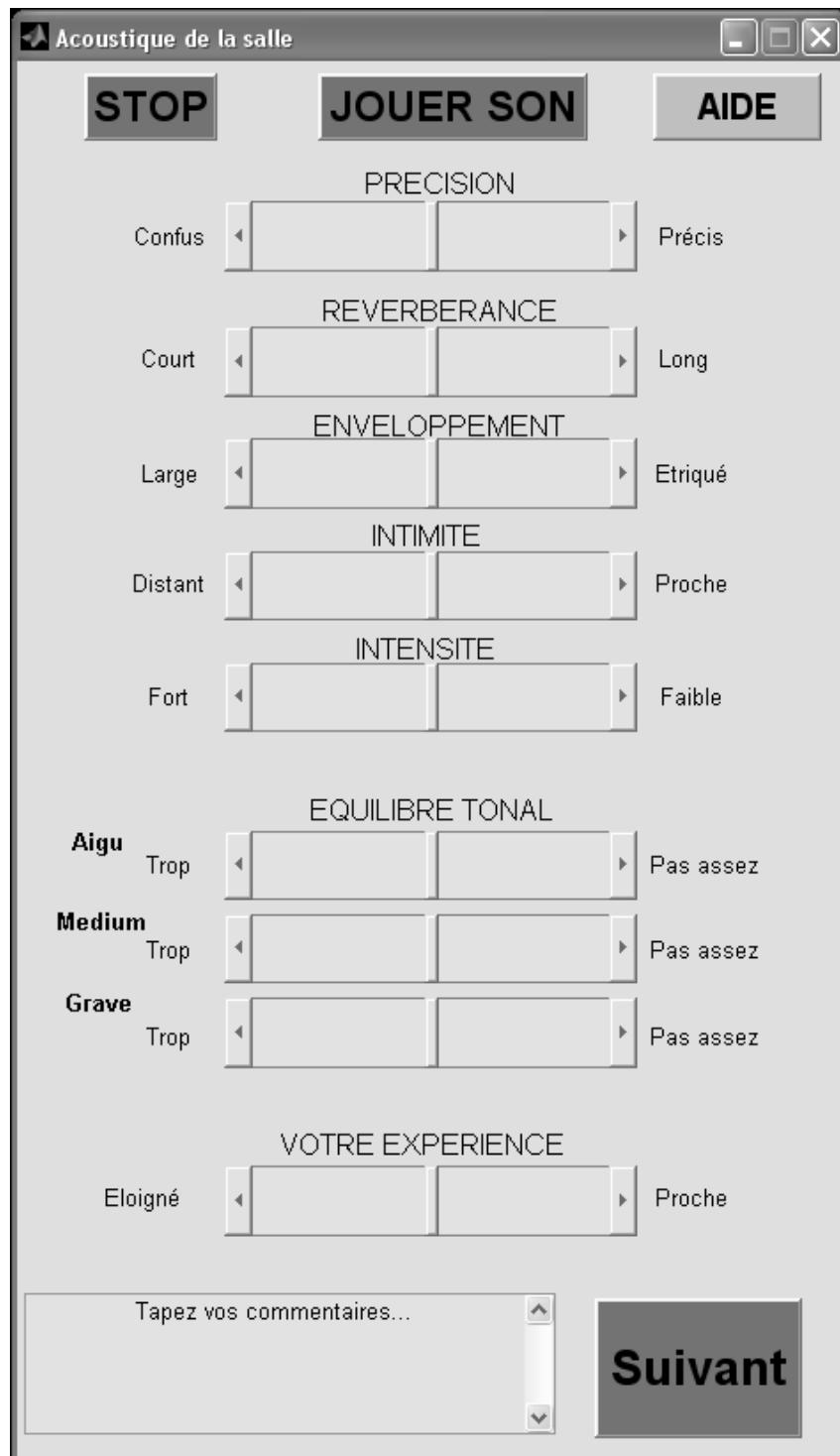


FIG. E.1 – Interface de test sous MatLab inspirée du test de Barron [122].

### E.3.1 Test de catégorisation libre

Un premier test d'écoute est réalisé à l'aide de quelques personnes, afin d'identifier les paramètres les plus importants perceptivement. Ce test, dit de catégorisation libre, propose au sujet de créer des familles avec les sons qu'il écoute. Nous avons choisi de synthétiser des réponses

impulsionnelles avec les paramètres suivants puis de les convoluer par un seul extrait de musique classique<sup>2</sup> d'environ 1min30s<sup>3</sup> :

1. volume :  $V_1 = 15000m^3$ ,  $V_2 = 25000m^3$  ;
2. distance source/récepteur :  $d_1 = 5m$ ,  $d_2 = 30m$  ;
3. temps de réverbération moyen :  $RT_{30} = 1.8s$ ,  $RT_{30} = 2.5s$  ;
4. diffusion : *synthétisée, non synthétisée* ;
5. filtrage : *bandes d'octave* (1ère famille de filtres), *type matériaux* (2ème famille de filtres), *Variations Spectrales Aléatoires -VSA* (3ème famille de filtres) ;
6. distribution angulaire d'incidence des rayons au point d'écoute : elle est fixée arbitrairement. Il s'agit d'une distribution normale centrée en 0, dont l'écart-type est égal à 0.3.

Afin de reproduire la sensation d'espace (avec toutes les limitations que cela implique<sup>4</sup>), les séquences sonores (24 au total) sont diffusées sur le système Ambisonic de la salle sèche du LAM. Le sujet fait face à un écran d'ordinateur sur lequel est représenté une page avec autant de pastilles qu'il y a de séquences sonores à écouter. La lecture d'un son est activée en cliquant sur une pastille et peut être interrompue. De plus, chaque son peut être écouté autant de fois souhaitées. Le sujet regroupe les sons qui lui semblent se ressembler en déplaçant les pastilles, et donne un nom à chaque groupe. Une fois le test passé (environ 30 minutes), il est invité à exprimer librement ses impressions. Afin de ne pas biaiser les résultats du test, seules des personnes ayant une expérience des concerts dans de grandes salles, et n'ayant pas de représentation physique préalable des phénomènes acoustiques ont passé le test. Précisons que le caractère virtuel des salles à l'écoute n'est jamais mentionné avant le test.

### E.3.2 Résultats du test de catégorisation libre

Au total, six personnes ont accepté de passer le test de catégorisation libre<sup>5</sup>. Contre toute attente pour un physicien, la distance source/récepteur n'est pas un élément discriminant. En effet, la majorité des groupes sont constitués en fonction du type de filtrage. De plus, la présence ou l'absence de la diffusion n'est pas ressentie. L'Annexe ?? recense les catégories sonores créées par les sujets et les commentaires associés.

Notons que les séquences dont le filtrage est celui de la 3ème famille (filtrage VSA) sont celles qui sont préférées et jugées les plus proches de l'expérience, du vécu des sujets. En revanche, le filtrage en bandes d'octave (1ère famille) est jugé désagréable et très éloigné de la réalité. C'est également le cas du filtrage type matériaux (2ème famille), mais dans une moindre mesure. Par conséquent, les filtrages par bandes d'octave et type matériaux sont abandonnés pour la suite de l'évaluation perceptive du modèle.

Enfin, les différents volumes ne sont que peu ressentis, car l'attention des sujets se portent principalement sur la couleur de la salle. Cependant, quelques sujets ont tout de même mentionné les différences de taille de salles (Annexe ??).

<sup>2</sup>Anechoic Orchestral Music Recording, Pure Gold Collection – GLINKA, Ouverture, *Ruslan and Lyudmila*.

<sup>3</sup>Cet extrait musical présente la particularité de regrouper en peu de temps des passages percussifs, des passages d'ensembles d'instruments à cordes, et des passages lents et rapides.

<sup>4</sup>En effet, il est certainement illusoire de demander à des sujets de se sentir dans une salle de concert, alors qu'ils se trouvent dans une salle de  $20m^3$  aux parois grises sans autres bruits parasites que les leurs. Les informations visuelles (couleurs, dimensions de la salle), olfactives (parfums) et tactiles (les sièges sont souvent en velours, avec des accoudoirs en bois) participent au sentiment d'être dans une salle de concert.

<sup>5</sup>Il est intéressant de noter qu'aucun des sujets n'a écouté plus d'une fois une séquence sonore dans sa totalité. Passée la première écoute, les sujets se concentraient sur les dix premières secondes de chaque séquence.

## E.4 Protocole expérimental

Sachant que le modèle stochastique est un modèle hautes fréquences, on considère ici trois filtrages possibles de la base de retours synthétisée par le modèle :

1. filtrage par Variations Spectrales Aléatoires (VSA) ;
2. filtrage par une impulsion à spectre plat ;
3. filtrage par une impulsion à spectre plat, puis application du filtre VSA.

Le test de catégorisation libre montrant que la présence ou l'absence de la diffusion n'a perceptivement pas d'impact, elle n'est pas retenue comme paramètre du test suivant. En revanche, les paramètres de volume, de distance source/récepteur et de filtrages sont retenus, et ont pour modalités (Table E.4) :

Paramètres	Modalité 1	Modalité 2	Modalité 3
Volume ( $m^3$ )	15000	20000	25000
Distance source/récepteur (m)	10	20	30
Filtrage	VSA	impulsion	impulsion + VSA

TAB. E.2 – Paramètres du test d'évaluation perceptive du modèle stochastique et leurs modalités.  
Note : VSA = Variations Spectrales Aléatoires.

La pièce de musique utilisée lors du test de catégorisation (Section E.3.1) est convoluée par chacune des 27 réponses impulsionales synthétisées à l'aide des paramètres du tableau E.4. Le questionnaire de Barron est mis sous forme d'une interface graphique. Le sujet peut écouter autant de fois qu'il le souhaite une séquence sonore et l'interrompre avant sa fin. De plus, les séquences sonores sont présentées dans un ordre aléatoire à chaque test.

## E.5 Discussion

Au moment de la rédaction de ce document, les tests sont en cours de réalisation. Nous devrions les faire paraître prochainement dans une publication.

Si les tests préliminaires ne valident pas immédiatement perceptivement le modèle, il faut néanmoins retenir la réaction des sujets. D'une part, ceux-ci ne sont pas sensibles au niveau du son direct, c'est-à-dire à la distance source/récepteur, mais bien plus à la couleur de la salle. Alors que les filtres en bandes d'octave et type matériaux ne sont pas jugés proches de l'expérience des sujets, le filtrage par Variations Spectrales Aléatoires se rapproche des attentes des sujets. Un autre aspect qui devrait retenir l'attention des futures recherches est la direction d'incidence des rayons sonores au cours du temps. Une première étude basée sur l'estimation des retours à partir de mesures Ambisonics (Chapitre 5) est sans doute nécessaire. Des tests perceptifs renseigneront sur la pertinence d'une telle approche. Bien entendu, si l'on souhaite modéliser "exactement" la réponse d'une salle, la connaissance "parfaite" de l'incidence des retours est primordiale. Mais la question que nous posons ici est : est-ce que cette "parfaite" connaissance est nécessaire et utile d'un point de vue perceptif ?

## E.6 Résultats du test de catégorisation libre

Sont regroupés dans cette Annexe les groupes et les commentaires faits par les six sujets ayant passé le test de catégorisation libre. Nous rappelons que ce test est préliminaire à l'évaluation

perceptive du modèle stochastique.

La Tableau E.3 ci-dessous recense les différents paramètres du modèle testés :

Paramètres	Modalité 1	Modalité 2	Modalité 3
Volume ( $m^3$ )	19000 ( $V_1$ )	22000 ( $V_2$ )	-
Distance source/récepteur (m)	5 ( $d_5$ )	30 ( $d_{30}$ )	-
Absorption (%)	30%	-	-
Diffusion	synthétisée ( $D_{on}$ )	non synthétisée ( $D_{off}$ )	-
Filtrage	bandes d'octave ( $F_1$ )	type matériaux ( $F_2$ )	VSA ( $F_3$ )
Distribution angulaire des rayons	$\mathcal{N}(\mu, \sigma^2)$ , $\mu = 0$ , $\sigma = 0.3$	-	-

TAB. E.3 – Paramètres et modalités du modèle testés lors du test de catégorisation libre. Note : VSA=Variations Spectrales Aléatoires.

A partir des modalités du Tableau E.3, il nous est possible de lister toutes les combinaisons possibles de réponses impulsionales (Tableau E.4) :

Numéro de la séquence sonore	Combinaison des paramètres
1	$F_1 V_1 D_{off} d_5$
2	$F_1 V_1 D_{off} d_{30}$
3	$F_1 V_1 D_{on} d_5$
4	$F_1 V_1 D_{on} d_{30}$
5	$F_1 V_2 D_{off} d_5$
6	$F_1 V_2 D_{off} d_{30}$
7	$F_1 V_2 D_{on} d_5$
8	$F_1 V_2 D_{on} d_{30}$
9	$F_2 V_1 D_{off} d_5$
10	$F_2 V_1 D_{off} d_{30}$
11	$F_2 V_1 D_{on} d_5$
12	$F_2 V_1 D_{on} d_{30}$
13	$F_2 V_2 D_{off} d_5$
14	$F_2 V_2 D_{off} d_{30}$
15	$F_2 V_2 D_{on} d_5$
16	$F_2 V_2 D_{on} d_{30}$
17	$F_3 V_1 D_{off} d_5$
18	$F_3 V_1 D_{off} d_{30}$
19	$F_3 V_1 D_{on} d_5$
20	$F_3 V_1 D_{on} d_{30}$
21	$F_3 V_2 D_{off} d_5$
22	$F_3 V_2 D_{off} d_{30}$
23	$F_3 V_2 D_{on} d_5$
24	$F_3 V_2 D_{on} d_{30}$

TAB. E.4 – Combinaisons des paramètres pour créer les réponses impulsionales du test.

Nous reportons ci-dessous les groupes et les commentaires associés à chaque groupe créés par chaque sujet.

### **Sujet 1 :**

1. Groupe 1 : 10, 12, 14, 16

Sons assourdis.

2. Groupe 2 : 5, 17, 21, 23

Assez peu de basses fréquences. Plus de volume<sup>6</sup>.

3. Groupe 3 : 4, 6, 8, 18, 20, 22, 24

Peu de basses fréquences. Impression de grésillement.

4. Groupe 4 : 2

Assourdi, lointain.

5. Groupe 5 : 9

Volume sonore élevé. Percussions bizarres.

6. Groupe 6 : 11, 13, 15

Percussions au son de casserole.

7. Groupe 7 : 7, 19

Sons les plus vivants.

8. Groupe 8 : 3

Sons les plus métalliques.

### **Sujet 2 :**

1. Groupe 1 : 2, 4, 6, 8, 18

Aigrelet, fluet. Comme une musique lue en avance rapide. La salle est très grande, et froide, comme recouverte de carrelage. On est très loin.

2. Groupe 2 : 11, 13, 15

Dans une grotte.

3. Groupe 3 : 17, 21

Musique étouffée par un voile.

4. Groupe 4 : 20, 22, 24

Musique étouffée par un rideau de velours. Mat. Salle pour musique de chambre.

5. Groupe 5 : 9, 14

Prétentieux. Très fort<sup>7</sup>, imposant. La salle est une cathédrale.

6. Groupe 6 : 10, 12, 16

Grave.

7. Groupe 7 : 3, 5

Liquide, cristallin, comme une église.

8. Groupe 8 : 1, 7, 19, 23

L'équilibre. Ces séquences sont les plus attendues pour une salle de concert.

---

<sup>6</sup>Comprendre ici, niveau sonore.

<sup>7</sup>Comprendre ici qu'il s'agit du niveau sonore.

### **Sujet 3 (anglophone) :**

1. Groupe 1 : 4, 6, 8, 22

Far, far away in an open space. This does not sound like a real concert hall. There is no reverberation.

2. Groupe 2 : 18, 20, 21, 24

Far, far away, in a large but closed space.

3. Groupe 3 : 9, 11, 12, 13, 15

Closer, but under the orchestra. A lot of bass.

4. Groupe 4 : 10, 14, 16

Underwater drowning. Too much bass. Not realistic at all for a concert hall.

5. Groupe 5 : 1, 2, 3, 5, 7

Sitting inside a trombone. We feel like backstage. There is a strong resonance. These are the worst sounds, not realistic at all for a concert hall.

6. Groupe 6 : 19, 21, 23

Still far away, but in a sweeter spot. These are my favorite sounds. We can see and listen to the orchestra. These are the more realistic. We stand in the centre of a quite big hall. Tones are balanced.

### **Sujet 4 (anglophone) :<sup>8</sup>**

1. Groupe 1 : 1, 3

Tin-like strings. This does not sound like a real hall.

2. Groupe 2 : 2, 4, 6, 8, 17, 24

Distant. Bad seat. I feel like I am listening to a recording of an orchestra which played in another hall than the original one. I am facing the orchestra.

3. Groupe 3 : 9, 11, 13, 15

Muddy under-water. Empty large hall. Envelopment. I feel like being into the orchestra, in a real hall.

4. Groupe 4 : 5, 7, 18, 20, 22

Crisp but distant. Bad seat in a good hall. Envelopment. We feel in a real hall.

5. Groupe 5 : 10, 12, 14, 16

Close, but less clean/clear than group 3.

6. Groupe 6 : 19, 21, 23

Grainy string. Same remark than group 2 about the recording.

7. General remark : The six groups define six different concert halls.

---

<sup>8</sup>Ce sujet est flûtiste semi-professionnel, et a l'habitude jouer dans un orchestre ainsi que de faire des représentations.

**Sujet 5 :**

1. Groupe 1 : 1, 7, 9, 11, 13, 15

Les sons les plus équilibrés en terme d'aigus et de graves. Plus englobant, plus proche de moi.

2. Groupe 2 : 2, 4, 6, 8, 24

Très loin devant, métalliques. Peu de basses.

3. Groupe 3 : 3, 17, 18, 20, 22

Loin devant (mais moins que le groupe 2). Moins métallique aussi que le groupe 2.

4. Groupe 4 : 12, 14

Plus de basses, de profondeur.

5. Groupe 5 : 5, 19, 21, 23

Assez équilibrés mais moins bien que le groupe 1.

6. Groupe 6 : 10, 16

Les pires.

7. Remarque générale : le groupe 1 correspond à une salle qui me paraît vraie, comparativement aux autres groupes (autres salles) dont le son ne me paraît pas naturel, pas plausible.



# Table des figures

2.1	Phénomènes de diffusion et d'absorption des ondes sonores dans un espace clos [11].	19
2.2	Représentation bidimensionnelle de l'évolution du front d'onde créé par une source omnidirectionnelle dans une salle parallélépipédique [11]. . . . .	19
2.3	Représentation bidimensionnelle de l'évolution du trajet d'un rayon sonore dans la salle parallélépipédique de la Fig. 2.2 [11]. . . . .	20
2.4	Représentation bidimensionnelle du trajet de l'onde directe et des premiers retours dans la salle parallélépipédique de la Fig. 2.2 dont les parois ont une impédance infinie [11]. $S$ est la source sonore, $M$ le point de réception, $D$ le trajet direct de l'onde entre $S$ et $M$ , $R_1$ est un rayon sonore subissant une réflexion. . . . .	21
2.5	Représentation schématique de la distribution temporelle des retours captés en un point $M$ dans une salle [11]. La source émet au temps $t = 0$ une impulsion. . . . .	22
2.6	Croissance et décroissance de l'énergie sonore dans une salle [11]. La source émet au temps $t = 0$ un signal stationnaire de niveau constant. . . . .	22
2.7	Représentation bidimensionnelle d'une densité temporelle de sources-images. $S'_i$ $S''_i$ $S''_{ij}$ ... sont les sources-images que donnent les parois de la source $S$ . Il est possible de tracer tous les rayons virtuels $S'_i M$ , $S''_i M$ , ... et de définir le nombre $\Delta N$ de sources comprises entre les cercles de rayons $c_0 t$ et $c_0(t + \Delta t)$ [11]. . . . .	28
2.8	Réflexion d'un rayon suivant la loi de Lambert. . . . .	30
2.9	Spectre (module et phase) du filtre type matériaux passe-bas. Fréquence de coupure : $700\text{Hz}$ . Spectre calculé sur $N = 2048$ points. . . . .	37
2.10	Spectre (module et phase) du filtre type matériaux passe-bande. Bande passante : $1000 - 4000\text{Hz}$ . Spectre calculé sur $N = 2048$ points. . . . .	37
2.11	Spectre (module et phase) du filtre type matériaux passe-haut. Fréquence de coupure : $6000\text{Hz}$ . Spectre calculé sur $N = 2048$ points. . . . .	38
2.12	Spectres (modules) du filtre <i>Variations Spectrales Aléatoires (VSA)</i> . Trait plein : son direct (impulsion de revolver). Trait interrompu : première réflexion. Trait interrompu pointillé : deuxième réflexion. Spectres calculés sur $N = 2048$ points. Ici, l'écart-type des variations est fixé à 0.1. . . . .	38
2.13	Spectres (phases) du filtre <i>Variations Spectrales Aléatoires (VSA)</i> . Trait plein : son direct (impulsion de revolver). Trait interrompu : première réflexion. Trait interrompu pointillé : deuxième réflexion. Spectres calculés sur $N = 2048$ points. Ici, l'écart-type des variations est fixé à 0.1. . . . .	39
2.14	Signal temporel d'une réponse impulsionale synthétisée par le modèle (sans diffusion, sans filtrage). . . . .	40
2.15	Haut : signal temporel d'une réponse impulsionale synthétisée par le modèle (avec diffusion et filtrage type matériaux). Milieu et bas : spectrogramme de la réponse impulsionale synthétisée ( $N_{FFT} = 2048$ pts). Haut : signal temporel entier ( $M=512$ pts temporel). Bas : zoom sur les premiers instants de la figure du haut ( $M=128$ pts temporel). L'échelle des couleurs est en décibels. . . . .	41

2.16	Haut : signal temporel d'une réponse impulsionnelle synthétisée par le modèle (avec diffusion et filtrage en bandes d'octave). Milieu et bas : spectrogramme de la réponse impulsionnelle synthétisée ( $N_{FFT} = 2048$ pts). Milieu : signal temporel entier ( $M=512$ pts temporel). Bas : zoom sur les premiers instants de la figure du haut ( $M=128$ pts temporel). L'échelle des couleurs est en décibels. . . . .	42
2.17	Haut : signal temporel d'une réponse impulsionnelle synthétisée par le modèle (avec diffusion et filtrage type VSA). Milieu et bas : spectrogramme de la réponse impulsionnelle synthétisée ( $N_{FFT} = 2048$ pts). Milieu : signal temporel entier ( $M=512$ pts temporel). Bas : zoom sur les premiers instants de la figure du haut ( $M=128$ pts temporel). L'échelle des couleurs est en décibels. . . . .	43
3.1	Variations of onset times estimation as a function of $SNR$ . . . . .	56
3.2	Variations of $C_{80}$ (top) and $T_C$ (bottom) for the 11 methods, $D_{SLP}$ being taken as reference. . . . .	57
3.3	Example of three different RIRs (note the different time scales). a) Simple case onset ; b) Scattering effect ; c) Precursor ; d) Details of a precursor presented in (c). . . . .	58
3.4	Experimental RIR for which determining the direct sound is not obvious. . . . .	60
3.5	Smoothed Wigner-Ville distributions of the direct sound of three experimental RIRs measured with : - a balloon burst (top) ; - a gun shot (middle) ; - a cracker (bottom) (at $48kHz$ ). . . . .	65
3.6	Matching Pursuit ran on an experimental RIR ( $SRR = 20dB$ ) [ (a) : experimental RIR, (b) : direct sound, (c) : linear set of coefficients ( $\gamma_{opt}^{(n)}$ ), (d) : residual]. Note this is a slightly modified version of the original figure. . . . .	66
3.7	Impulse duration of balloon bursts, gun shots, and crackers measured in a damped room. . . . .	67
3.8	Impulse duration calculated on RIRs using the <i>STFT</i> method. . . . .	67
3.9	Impulse duration calculated on RIRs using Matching Pursuit ( $SRR = 20dB$ ). . . . .	67
4.1	Haut (2 premiers graphiques) : Ecart-type ( $R$ ) (noir) d'une réponse impulsionnelle expérimentale (gris) calculé sur $N = 200$ points (fréquence d'échantillonnage de $F_s = 48kHz$ ) (signal entier et détails). Pour une meilleure lisibilité, on a tracé $R - 2$ . Bas (2 derniers graphiques) : Kurtosis (noir) d'une réponse impulsionnelle expérimentale (gris) calculé sur $N = 200$ points ( $F_s = 48kHz$ ) (signal entier et détails). Note : les signaux sont normalisés et les estimateurs passant par 0 marquent la réalisation de l'hypothèse de Gaussianité. On visualise nettement les retours de plus fortes énergie, ainsi que la présence de bouffées de cohérence une fois le champ diffus établi. Enfin, le signal utilisé ici est le même sur le graphique du haut et du bas. . . . .	72
4.2	Haut : réponse impulsionnelle expérimentale (fréquence d'échantillonnage de $F_s = 48kHz$ ) (signal entier et détails). Milieu : probabilité de validité de l'hypothèse nulle du test de $\chi^2$ sur des fenêtres de taille 200 échantillons. Bas : probabilité de validité de l'hypothèse nulle du test de $\chi^2$ sur des fenêtres de taille 120 échantillons. Note : en noir, les instants auxquels $P(t) \geq 0.05$ . Remarquons que la validation de l'hypothèse nulle, contre celle non-nulle, dépend de la taille de la fenêtre d'analyse. De plus, quelque soit le niveau de détails auquel nous pourrions observer les graphiques, nous visualiserions toujours une zone de rejet de l'hypothèse nulle. . . . .	75

4.3	Haut : réponse impulsionale expérimentale (fréquence d'échantillonnage de $F_s = 48kHz$ ) (signal entier et détails). Milieu : probabilité de validité de l'hypothèse nulle du test de Lilliefors sur des fenêtres de taille 200 échantillons. Bas : probabilité de validité de l'hypothèse nulle du test de Lilliefors sur des fenêtres de taille 120 échantillons. Note : en noir, les instants auxquels $P(t) \geq 0.05$ . Remarquons que la validation de l'hypothèse nulle, contre celle non-nulle, dépend de la taille de la fenêtre d'analyse. La réponse impulsionale étudiée ici est la même que celle de la Fig. 4.2. . . . .	76
4.4	Haut : réponse impulsionale expérimentale (fréquence d'échantillonnage de $F_s = 48kHz$ ) (signal entier et détails). Milieu : probabilité de validité de l'hypothèse nulle du test de Kolmogorov-Smirnov sur 2 fenêtres de taille 200 échantillons. Bas : probabilité de validité de l'hypothèse nulle du test de Kolmogorov-Smirnov sur 2 fenêtres de taille 120 échantillons. Note : en noir, les instants auxquels $P(t) \geq 0.05$ . Remarquons que la validation de l'hypothèse nulle, contre celle non-nulle, dépend de la taille de la fenêtre d'analyse. La réponse impulsionale étudiée ici est la même que celle de la Fig. 4.2. . . . .	77
4.5	Haut : réponse impulsionale modélisée par un bruit blanc Gaussien (fréquence d'échantillonnage de $F_s = 48kHz$ ). Milieu : probabilité de validité de l'hypothèse nulle du test de $\chi^2$ sur des fenêtres de taille 200 échantillons. Bas : probabilité de validité de l'hypothèse nulle du test de $\chi^2$ sur des fenêtres de taille 120 échantillons. Note : en noir, les instants auxquels $P(t) \geq 0.05$ . Remarquons que la validation de l'hypothèse nulle, contre celle non-nulle, dépend de la taille de la fenêtre d'analyse. . . . .	78
4.6	Haut : réponse impulsionale expérimentale (fréquence d'échantillonnage de $F_s = 48kHz$ ) Bas : entropie calculée sur des fenêtres de 120 échantillons. Note : les instants où l'entropie du système est proche de celle initiale se distinguent nettement. La décroissance de l'énergie de la réponse est compensée pour le calcul de l'entropie, mais n'est pas montrée sur la figure du haut. . . . .	80
5.1	Early times of an experimental RIR (encapsuled graphic), measured in Salle Pleyel.	83
5.2	Experimental RIR for which determining the direct sound is obvious. Plain line : direct sound. . . . .	87
5.3	Signed variations in % of $EDT_{10}$ , $RT_{20}$ , $RT_{30}$ and $T_C$ versus $SRR$ in dB. . . . .	88
5.4	Matching Pursuit ran on an experimental RIR ( $SRR = 20dB$ ). -(top) : Experimental RIR -(bottom) : Linear set of coefficients, which correspond to the arrivals.	89
5.5	Detail of an experimental RIR for which determining the direct sound is not obvious. The direct sound estimated by MP is delimited by dashed lines. . . . .	90
5.6	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. . . . .	91
5.7	$CDF$ , for an experimental RIR (without compensation of the energy decay). Dashed line : a kink in the distribution of arrivals is noticeable. The encapsuled graphic is a zoom on the early times of the $CDF$ . . . . .	91
5.8	Plain : Logarithm of $(1 - CDF)$ of the arrivals of 21 experimental RIRs (without compensation of the energy decay). Dash : mean logarithmic decay of the RIRs. .	92
5.9	$CDFs$ . 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 $CDF$ . . . . .	92

5.10	<i>CNAs</i> of a synthesized RIR. Dash : theoretical <i>CNA</i> . Plain : estimated <i>CNA</i> (the atom is a pistol shot) -bold : estimated <i>CNA</i> (the atom is a Dirac) (with compensation of the energy decay). . . . .	93
5.11	Detail of a cubic ( $t < T_M$ ) and a linear ( $t > T_M$ ) fits made on a <i>CNA</i> . Plain bold line : <i>CNA</i> . Dashed bold line : cubic fit. Dashed line : linear fit. Square : mixing time. . . . .	94
5.12	Circles : Slope coefficients of the linear fits made on the 21 <i>CNAs</i> ( <i>mean</i> = 1160 arrivals/second, <i>std</i> = 5%). Plain line : mean slope coefficients. Dashed lines : standard deviation. . . . .	95
5.13	a) : Mixing time (circle) detected on a <i>CDF</i> (bold line) for an experimental RIR and the set of estimated arrivals. ( <i>SRR</i> = 5dB). Plot (b) details the beginning of plot (a). . . . .	96
5.14	a) : Mixing time (circle) detected on a <i>CDF</i> (bold line) for a synthesized RIR convolved with a pistol shot and the set of estimated arrivals ( <i>SRR</i> = 5dB). Plot (b) details the beginning of plot (a). . . . .	97
5.15	Mixing times estimated on experimental RIRs using different values of <i>SRR</i> (dB) (with compensation of energy decay). . . . .	98
5.16	Variations in % of $EDT_{10}$ , $RT_{20}$ , $RT_{30}$ , $T_C$ versus the <i>SRR</i> in dB (with compensation of the energy decay). . . . .	98
5.17	Estimated mixing times as a function of distance (without (triangles) and with (circles) time propagation between the source and the receiver). . . . .	99
5.18	Estimated mixing times (dots) as a function of distance (with time propagation between the source and the receiver). Dashed line : linear regression. . . . .	100
5.19	Mixing times (dots) detected for 20 RIRs synthesized at different distances (the time of propagation is taken into account) ( <i>SRR</i> = 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 $d \geq 25m$ ). . . . .	101
5.20	Mean mixing times of the model (triangles) and of the experiments (squares). Dashed line : linear regression made on the minima mixing times of the model. Plain line : linear regression of experimental RIRs. . . . .	102
5.21	Mean mixing times from the model (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. . . . .	103
5.22	Superposition des résultats expérimentaux et numériques (et régressions linéaires). La diffusion n'est pas synthétisée par le modèle. Points : temps de mélange du modèle. Carrés : temps de mélange moyens du modèle. Cercles : temps de mélange expérimentaux ( <i>SRR</i> = 5dB). . . . .	104
5.23	Superposition des résultats expérimentaux et numériques (et régressions linéaires). La diffusion est synthétisée par le modèle. Points : temps de mélange du modèle. Carrés : temps de mélange moyens du modèle. Cercles : temps de mélange expérimentaux ( <i>SRR</i> = 5dB). . . . .	105
5.24	Résultats expérimentaux et numériques (et régressions linéaires). Le filtrage (en bandes d'octaves) est synthétisé par le modèle. Points : temps de mélange du modèle. Carrés : temps de mélange moyens du modèle. Etoiles : temps de mélange expérimentaux. Ligne pleine : régression linéaire sur les temps de mélange moyens du modèle. Ligne pointillée : régression linéaire sur les temps de mélange expérimentaux ( $T_m = 0.0026 \times d + 0.026$ ) ( <i>SRR</i> = 5dB). . . . .	106

5.25 Résultats expérimentaux et numériques (et régressions linéaires). Le filtrage (type matériaux) est synthétisé par le modèle. Points : temps de mélange du modèle. Carrés : temps de mélange moyens du modèle. Etoiles : temps de mélange expérimentaux. Ligne pleine : régression linéaire sur les temps de mélange moyens du modèle. Ligne pointillée : régression linéaire sur les temps de mélange expérimentaux ( $Tm = 0.0026 \times d + 0.026$ ) ( $SRR = 5dB$ ). . . . .	107
5.26 Résultats expérimentaux et numériques (et régressions linéaires). Le filtrage (Variations Spectrales Aléatoires -VSA) est synthétisé par le modèle. Points : temps de mélange du modèle. Carrés : temps de mélange moyens du modèle. Etoiles : temps de mélange expérimentaux. Ligne pleine : régression linéaire sur les temps de mélange moyens du modèle. Ligne pointillée : régression linéaire sur les temps de mélange expérimentaux ( $Tm = 0.0026 \times d + 0.026$ ) ( $SRR = 5dB$ ). . . . .	108
5.27 Résultats expérimentaux et numériques (et régressions linéaires). Le filtrage (en bandes d'octaves) et la diffusion sont synthétisés par le modèle. Points : temps de mélange du modèle. Carrés : temps de mélange moyens du modèle. Etoiles : temps de mélange expérimentaux. Ligne pleine : régression linéaire sur les temps de mélange moyens du modèle. Ligne pointillée : régression linéaire sur les temps de mélange expérimentaux ( $Tm = 0.0026 \times d + 0.026$ ) ( $SRR = 5dB$ ). . . . .	109
5.28 Résultats expérimentaux et numériques (et régressions linéaires). Le filtrage (type matériaux) et la diffusion sont synthétisés par le modèle. Points : temps de mélange du modèle. Carrés : temps de mélange moyens du modèle. Etoiles : temps de mélange expérimentaux. Ligne pleine : régression linéaire sur les temps de mélange moyens du modèle. Ligne pointillée : régression linéaire sur les temps de mélange expérimentaux ( $Tm = 0.0026 \times d + 0.026$ ) ( $SRR = 5dB$ ). . . . .	110
5.29 Résultats expérimentaux et numériques (et régressions linéaires). Le filtrage (VSA) et la diffusion sont synthétisés par le modèle. Points : temps de mélange du modèle. Carrés : temps de mélange moyens du modèle. Etoiles : temps de mélange expérimentaux. Ligne pleine : régression linéaire sur les temps de mélange moyens du modèle. Ligne pointillée : régression linéaire sur les temps de mélange expérimentaux ( $Tm = 0.0026 \times d + 0.026$ ) ( $SRR = 5dB$ ). . . . .	111
5.30 Haut : nombre de retours par seconde en fonction de l'emplacement de réception choisi dans la Salle Pleyel (mesures omnidirectionnelles) pour trois points d'émission sur la scène. Bas : nombre maximum de retours captés au point de réception pour trois points d'émission sur la scène. Note : PF= Parterre-Face ; PL=Parterre Latéral ; PFF=Parterre Face au Fond ; BF1=Balcon Face 1er étage ; BL1=Balcon Latéral 1er étage ; BF2=Balcon Face 2nd étage ; BL2=Balcon Latéral 2nd étage.	117
5.31 Nombre de retours par seconde (dans le champ diffus) en fonction de l'emplacement de réception choisi dans la Salle Pleyel et de la direction d'incidence des rayons (mesures B-format). Ici, une seule position de source est considérée. Note : W : captation omnidirectionnelle ; X : directivité avant/arrière ; Y : directivité latérale ; Z : directivité dessus/dessous. On rappelle que lors de mesures, l'axe X est dirigé vers la source situé à la place du chef d'orchestre. . . . .	118
5.32 Les 4 CNAs correspondant aux 4 canaux de B-format. Mesure au premier rang du parterre (source S1, réception PF). La source est en champ direct à 8m. Trait plein gras : canal W. Trait pointillé gras : canal X. Trait pointillé : canal Y. Trait plein : canal Z. Note : le temps de propagation entre la source et le récepteur n'est pas pris en compte. . . . .	118

5.33	Les 4 <i>CNA</i> s correspondant aux 4 canaux de B-format. Mesure (source <i>S3</i> , réception <i>BL1</i> ) sur un balcon à 18m de la source qui est cachée. Trait plein gras : canal <i>W</i> . Trait pointillé gras : canal <i>X</i> . Trait pointillé : canal <i>Y</i> . Trait plein : canal <i>Z</i> . Note : le temps de propagation entre la source et le récepteur n'est pas pris en compte. . . . .	119
5.34	Interface de test inspiré des tests MUSHRAM [86]. . . . .	119
5.35	Temps de mélange estimé dans la Salle Pleyel à l'aide Matching Pursuit pour un critère d'arrêt $SRR = 8dB$ . . . . .	120
5.36	Temps de mélange estimé dans la Salle Pleyel à l'aide Matching Pursuit pour un critère d'arrêt $SRR = 6dB$ . . . . .	120
5.37	En haut : nombre de retours estimés (fenêtre de 24ms) dans la Salle Pleyel à l'aide Matching Pursuit pour des critères d'arrêt allant de 0.5 à 20dB. En bas : nombre de retours estimés dans la Salle Pleyel à l'aide Matching Pursuit pour des critères d'arrêt allant de 4 à 10dB. . . . .	121
5.38	Haut : azimuts des retours estimés dans la Salle Pleyel à l'aide Matching Pursuit pour un critère d'arrêt $SRR = 5dB$ au premier rang. Bas : sites des retours estimés dans la Salle Pleyel à l'aide Matching Pursuit pour un critère d'arrêt $SRR = 5dB$ au premier rang. . . . .	122
5.39	Décomposition d'une réponse impulsionale expérimentale en 1/3 de bandes d'octave ( $SRR = 5dB$ ). Chaque trait blanc marque l'occurrence d'un retour dans le plan temps/fréquence. On retrouve la coloration particulière de l'impulsion de revolver [47] ( voir aussi Chapitre 3). . . . .	123
6.1	Detail of the early times of an experimental room impulse response. The encapsulated plot corresponds to the full record length RIR. . . . .	128
6.2	Scheme of the principle of the XFT. For each window, a Discrete Fourier Transform of the signal $x$ is computed. In that case, the window $w$ is a Hann window, which starts from the time of arrival of the direct sound. The latter is detected as described in [55]. . . . .	130
6.3	XFT applied to an experimental room impulse response signal for which the energy decay is compensated (but not shown here for an easier readability) (top). Middle : magnitude of spectra (linear scale). Bottom : unwrapped phases of spectra (linear scale in $n\pi$ ). The dashed line marks a noticeable change in the statistics of the signal. Note the different frequency scales (middle and bottom graphs). The time of propagation between the source and the receiver is not shown here. The XFT is calculated from the time of occurrence of the direct sound. The RIR is the same as Fig.6.2. . . . .	132
6.4	(a) : Zoom on the early times of a synthesized RIR (the energy compensation is applied, but not shown for an easier readability). (b) : Time elapsed between successive arrivals. (c)-(h) : Unwrapped phase at different times starting from the time of arrival of the direct sound. (c) : $0 \leq t \leq 60 ms$ . (d) : $0 \leq t \leq 80 ms$ . (e) : $0 \leq t \leq 90 ms$ . (f) : $0 \leq t \leq 110 ms$ . (g) : $0 \leq t \leq 120 ms$ . (h) : $0 \leq t \leq 130 ms$ . Note that phases scales are the same for each graph from (c) to (h). . . . .	144
6.5	Visualization of the unwrapped phases of the XFT computed on a RIR measured in Salle Pleyel (source : pistol shot). Note that the compensation of the energy decay of the RIR is applied, but not shown for a better readability. Furthermore, the time of propagation between the source and the receiver positions is not taken into account. First window width : $L_0 = 128$ samples. The estimated mixing time is materialized by a dashed line superposed to the RIR. . . . .	145

- 6.6 Grey : experimental RIR (the energy decay compensation is applied, but not shown for a better readability). Plain black line : curve  $\delta\Phi(p)$  (Eq.(6.34)) (normalized). Black bold dashed line : cubic fit. Note that  $\delta\Phi$  is over sampled for a better readability. The graphic included shows the details of the selected region within the dashed rectangle. The RIR is the same as Fig.6.2. . . . . 146
- 6.7 Top : an experimental RIR (the energy decay compensation is applied, but not shown for a better readability). Bottom : Eq.(6.34) in grey ; Eq.(6.35) in black. The breaking point (i.e., the mixing time) in the statistics of the RIR is materialized by a dashed line. Note that bottom graphic is over sampled for a better readability. Notice the other beaking points after the mixing time (other peaks). The time of propagation between the source and the receiver positions is not shown here. On bottom graphic, curves are normalized for an easier readability. The RIR is the same as Fig.6.2. . . . . 146
- 6.8 Estimated mixing times on 42 RIRs measured in Salle Pleyel (sources : pistol shots = circles, balloon bursts = stars, cracker impulses = crosses). Note the superposition of some data. Dashed line : linear regression calculated on mixing times for the three types of sources. . . . . 147
- 6.9 Dots : estimated mixing times on 420 synthesized RIRs for the Salle Pleyel (source : Dirac). First analysis window width :  $L_0 = 4$  samples. In that case, the model synthesizes diffusion. Squares : mean mixing times. Dashed line : linear regression calculated on mean mixing times. . . . . 147
- 6.10 Top : mean mixing times  $\bar{T}_m$  (s) estimated on RIRs synthesized with diffusion (stars) and without diffusion (squares). Bottom : standard deviations  $\sigma$  (%) of mixing times estimated on synthesized RIRs with diffusion (stars) and without diffusion (squares). The source is a Dirac. First analysis window width :  $L_0 = 4$  samples. . . . . 148
- 6.11 Dots : estimated mixing times on 420 synthesized RIRs for the Salle Pleyel (source : balloon burst). Stars : experimental mixing times (using balloon bursts). First window width :  $L_0 = 256$  samples. Squares : mean mixing times. Dashed line : linear regression calculated on mean mixing times. . . . . 149
- 6.12 Dots : estimated mixing times on 420 synthesized impulse responses for the Salle Pleyel (source : pistol shot). Stars : experimental mixing times (using pistol shots). First window width :  $L_0 = 128$  samples. Squares : mean mixing times. Dashed line : linear regression calculated on mean mixing times. . . . . 149
- 6.13 Dots : estimated mixing times on 420 synthesized impulse responses for the Salle Pleyel (source : cracker impulse). Stars : experimental mixing times (using cracker imlpulses). First window width :  $L_0 = 128$  samples. Squares : mean mixing times. Dashed line : linear regression calculated on mean mixing times. . . . . 150
- 6.14 Variations in percentage (interquartiles) of mixing time versus the signal to noise ratio  $SNR$  (dB). 420 synthesized RIRs are considered (the source is a pistol shot). First analysis window width :  $L_0 = 128$  samples. The reference mixing time is estimated for  $SNR = 80$  dB. Dots stand for extreme values. . . . . 150
- 6.15 Variations in percentage (interquartiles) of mixing time versus the signal to noise ratio ( $SNR$  (dB)). 21 RIRs measured in Salle Pleyel are considered (source : pistol shots). First analysis window width :  $L_0 = 128$  samples. The reference mixing time is the one estimated at  $SNR = 80$  dB (experimental conditions [52]). Dots stand for extreme values. . . . . 151

6.16	Signed variations in percentage (interquartiles) of mixing time $\Delta T_M$ (%) versus variations $\Delta L_0$ (%) of the analysis window width ( $L_0$ ). 420 synthesized RIRs are considered (the source is a pistol shot). The reference mixing times are the ones estimated for $\Delta L_0 = 0\%$ , that is, $L_0 = 128$ samples. Bottom : zoom on the top graphic. Dots stand for extreme values. . . . .	152
6.17	Top : signed variations in percentage (interquartiles) of mixing time $\Delta T_M$ (%) versus variations $\Delta L_0$ (%) of the analysis window width ( $L_0$ ). 21 experimental RIRs (source : pistol shots) are considered (with compensation of the energy decay). The reference mixing times are the ones estimated for $\Delta L_0 = 0\%$ , that is, $L_0 = 128$ samples. Bottom : zoom on the top graphic. Dots stand for extreme values. . . . .	153
6.18	Median variations in percentage of mixing time $\Delta T_M$ (%) versus variations $\Delta L_0$ (%) of the analysis window width ( $L_0$ ). Black line : 420 synthesized RIRs are considered. Grey : 21 experimental RIRs. The source is a pistol shot. First analysis window width : $L_0 = 128$ samples. . . . .	154
6.19	Temps de mélange estimé dans la Salle Pleyel à l'aide Matching Pursuit [ $SRR = 5dB$ ] (étoiles) et la XFT (cercles). Source : pistolet. . . . .	154
7.1	Schéma d'une salle aux parois diffusantes. La diffusion s'établit progressivement dans l'espace. . . . .	157
7.2	Schéma d'une salle et d'une antenne de microphones. . . . .	159
7.3	Observation schématique en deux dimensions du nombre de particules captées dans un volume de l'espace au cours du temps ( $t_1 \dots t_m$ ). Ici, $t_m$ correspond au temps à partir duquel la densité de particules est uniforme sur tout l'espace observé.	160
7.4	Photographies de la propagation d'ondes dans des maquettes à l'aide la technique de Schlieren [12]. . . . .	160
A.1	Source-image du premier ordre ( $S_1$ ). $S$ : point source. $R$ point de réception. . . . .	171
A.2	Exemple de source-image ne contribuant pas à la réponse impulsionnelle de la salle. La source $S_1$ contribue à la réponse au récepteur $R_a$ mais pas à celle du récepteur $R_b$ [38]. . . . .	172
A.3	Exemple d'estimation d'une source-image par rétro-propagation [38]. . . . .	175
A.4	A gauche : Cône, dont l'origine est la source $S$ , et sa fonction de pondération en $\cos^2$ . A droite : Recouvrement des cônes sur un élément de surface de la sphère entourant le point source ( $S$ ) [38]. . . . .	176
C.1	Haut : schéma de la salle Great Hall à l'Université Queen Mary de Londres (Royaume-Uni). Bas : schéma de la salle Octagon à l'Université Queen Mary de Londres (Royaume-Uni). Les points indiquent les emplacements du récepteur (microphone omnidirectionnel). . . . .	200
C.2	Cercles : temps de mélange estimés sur les salles Great Hall (haut) et Octagon (bas) ( $SRR = 6dB$ ). Traits pleins : régressions linéaires. . . . .	201
C.3	Temps de mélange (en $ms$ ) estimés sans (a et c) et avec (b et d) la distance de propagation source/récepteur. Deux figures du haut (a) et (b) : Great Hall. Deux figures du bas (c) et (d) : Octagon. ( $SRR = 6dB$ ). . . . .	202

D.1	Haut : réponse impulsionale modélisée par un bruit blanc Gaussien (fréquence d'échantillonnage de $F_s = 48kHz$ ). Milieu : probabilité de validité de l'hypothèse nulle du test de $\chi^2$ sur 1 fenêtre de taille 200 échantillons. Bas : probabilité de validité de l'hypothèse nulle du test de $\chi^2$ sur 1 fenêtre de taille 120 échantillons. Note : en noir, les instants auxquels $P(t) \geq 0.05$ . Remarquons que la validation de l'hypothèse nulle dépend de la taille de la fenêtre d'analyse. . . . .	204
D.2	Haut : réponse impulsionale modélisée par un bruit blanc Gaussien (fréquence d'échantillonnage de $F_s = 48kHz$ ). Milieu : probabilité de validité de l'hypothèse nulle du test de Lilliefors sur des fenêtres de taille 200 échantillons. Bas : probabilité de validité de l'hypothèse nulle du test de Lilliefors sur des fenêtres de taille 120 échantillons. Note : en noir, les instants auxquels $P(t) \geq 0.05$ . Remarquons que la validation de l'hypothèse nulle dépend de la taille de la fenêtre d'analyse. . . . .	204
D.3	Haut : réponse impulsionale modélisée par un bruit blanc Gaussien (fréquence d'échantillonnage de $F_s = 48kHz$ ). Milieu : probabilité de validité de l'hypothèse nulle du test de Kolmogorov-Smirnov sur 2 fenêtres de taille 200 échantillons. Bas : probabilité de validité de l'hypothèse nulle du test de Kolmogorov-Smirnov sur 2 fenêtres de taille 120 échantillons. Note : en noir, les instants auxquels $P(t) \geq 0.05$ . Remarquons que la validation de l'hypothèse nulle dépend de la taille de la fenêtre d'analyse. . . . .	205
D.4	Haut : réponse impulsionale modélisée par un bruit blanc Gaussien (fréquence d'échantillonnage de $F_s = 48kHz$ ). Milieu : entropie calculée sur des fenêtres de taille 200 échantillons. Bas : entropie calculée sur des fenêtres de taille 120 échantillons. Remarquons que la valeur de l'entropie dépend de la taille de la fenêtre d'analyse. . . . .	205
E.1	Interface de test sous MatLab inspirée du test de Barron [122]. . . . .	209

## Résumé

Nous étudions les propriétés statistiques de réponses impulsionales de salles. Si la qualité acoustique d'une salle est décrite par ses réponses impulsionales, ces dernières renseignent également sur le système physique considéré. En 1975, Joyce émet l'hypothèse selon laquelle les grandes salles sont mélangeantes ou tout au moins ergodiques. Afin de vérifier cette dernière hypothèse, ce document se concentre sur l'estimation du temps auquel le système devient mélangé : le temps de mélange. Plusieurs types d'estimateur sont mis en œuvre.

L'algorithme Matching Pursuit permet d'estimer les temps d'occurrence des retours. L'étude de leur distribution temporelle permet de définir un temps à partir duquel la densité de retours augmente d'un rapport constant avec le temps. Nous avons développé l'eXtensible Fourier Transform (XFT) qui permet une visualisation temporelle pas-à-pas des composantes spectrales du signal. La XFT permet d'estimer le temps à partir duquel les composantes spectrales du signal sont distribuées statistiquement. Les deux estimateurs s'accordent tant sur des réponses expérimentales que synthétisées par un modèle stochastique : le temps de mélange est une fonction de la distance source/récepteur. Nous montrons que le temps que nous estimons n'est pas le temps de mélange au sens dynamique du terme. Nous proposons l'emploi du nom de *temps de transition* à la place. Finalement, les estimateurs ouvrent la voie à l'estimation de la diffusion, du filtrage d'une salle ou de l'entropie comme des fonctions du temps.

### Mots-clé :

Acoustique des salles, systèmes dynamiques, mélange, billards, processus stochastiques, traitement du signal, propriétés statistiques des signaux.

## Abstract

In this dissertation, we focus on the statistical properties of room impulse responses (RIRs). RIRs are measured not only to describe the acoustics of a hall, but also to provide information about the physical system. In 1975, Joyce assumed that large halls are mixing, or at least ergodic. In order to verify this assumption, we concentrate on the estimation of what is called the mixing time, for which we develop and study several estimators. Among these, one algorithm allows the estimation of the time of arrivals. By studying their time distribution, we are able to define a time at which the number of arrivals increases at a constant rate (the mixing time). We also develop the eXtensible Fourier Transform (XFT) which allows the visualization of the time evolution of spectral components of a RIR. The XFT is able to find the time at which spectral components are statistically distributed. We evaluate both estimators on experimental and synthesized RIRs using a stochastical model. We found that the mixing time is a linear function of the source/receiver distance. We show that the time we estimate is not the mixing time according to the theory of dynamical systems. Instead, we propose the term of *cross-over time*. Finally, our estimators allow for future study of diffusion, filtering of the room and the entropy as functions of time.

### Keywords :

Room acoustics, dynamical systems, mixing, billiards, stochastic processes, signal processing, statistical properties of signals.

