



HAL
open science

Apprentissage automatique et optimisation pour la détermination des propriétés acoustiques d'une salle à partir de signaux audio

Stéphane Dilungana

► **To cite this version:**

Stéphane Dilungana. Apprentissage automatique et optimisation pour la détermination des propriétés acoustiques d'une salle à partir de signaux audio. Traitement du signal et de l'image [eess.SP]. Université de Strasbourg, 2024. Français. NNT : 2024STRAD015 . tel-04769735

HAL Id: tel-04769735

<https://theses.hal.science/tel-04769735v1>

Submitted on 6 Nov 2024

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.

**ÉCOLE DOCTORALE MATHÉMATIQUES, SCIENCES DE L'INFORMATION ET DE
L'INGÉNIEUR - ED 269**

MULTISPEECH - Inria - Loria UMR 7503

UMRAE - Cerema - Strasbourg

IMAGeS - ICube UMR 7357

THÈSE présentée par :

Stéphane DILUNGANA

soutenue le : **4 septembre 2024**

pour obtenir le grade de : **Docteur de l'université de Strasbourg**

Discipline / Spécialité : **Traitement du Signal et des Images**

**Apprentissage automatique et
optimisation pour la détermination des
propriétés acoustiques d'une salle à
partir de signaux audio**

THÈSE dirigée par :

M. FAISAN Sylvain

M. DELEFORGE Antoine

M. FOY Cédric

Maître de conférences HDR, Université de Strasbourg

Chargé de recherche, Inria, Antenne de Strasbourg

Chargé de recherche, UMRAE, Cerema, Antenne de Strasbourg

RAPPORTEURS :

M. MICHEL Olivier

M. RICHARD Gaël

Professeur, Université Grenoble Alpes

Professeur, Télécom Paris

AUTRES MEMBRES DU JURY :

Mme MESLEM Amina

Professeur, Université de Rennes 1

Remerciements

Je souhaite exprimer ma profonde gratitude envers les organismes qui ont contribué à rendre possible cette thèse et envers toutes les personnes qui ont participé à faire des ces quelques années une expérience inoubliable et enrichissante tant sur le plan scientifique que personnel.

Cette thèse a bénéficié d'un financement de la part de l'Inria dans le cadre du projet "action exploratoire" Acoust.IA. Elle s'est déroulée principalement au sein des laboratoires UMRAE et ICube à Strasbourg, qui ont fourni les équipements et ressources de calcul nécessaires à la réalisation de ces travaux.

Je remercie infiniment mes encadrants Antoine Deleforge, Cédric Foy et Sylvain Faisan sans qui cet accomplissement n'aurait pas été possible. Leur supervision et leur enthousiasme scientifique, ainsi que leur bienveillance et leur soutien m'ont été très précieux et ont constitué des moteurs essentiels de ma motivation tout au long de la thèse. Je leur suis reconnaissant de la grande confiance qu'ils m'ont accordée et je garderai longtemps en mémoire nos échanges passionnants et leurs qualités respectives qui continueront de m'inspirer dans mon parcours professionnel.

Je remercie les membres du jury, Gaël Richard, Olivier Michel et Amina Meslem pour leur investissement et leur intérêt à l'égard de ces travaux ainsi que pour les retours et échanges très intéressants et constructifs. Je remercie également les membres du comité de suivi, Arnaud Can et Emmanuel Franck pour leur disponibilité et leurs conseils avisés.

J'ai eu la chance d'évoluer dans des laboratoires de recherche où j'ai pu me sentir à l'aise grâce à l'atmosphère accueillante et conviviale créée par leur personnel. Un grand merci à l'ensemble des collègues de l'UMRAE (Strasbourg et Nantes), de l'ICube et de l'Inria pour leur sympathie et pour m'avoir permis d'en apprendre toujours un peu plus sur de nombreux sujets.

Je tiens enfin à remercier sincèrement mes parents, ma famille et mes proches pour leur présence et leur soutien inestimable. Spécialement, Maman, Bénédicte et Divine, merci de toujours me soutenir, m'encourager, m'inspirer et me rendre fier. Je dédie cette thèse à ma nièce Amaryah, merci d'être une source intarissable d'amour, de bonheur et de joie.

Résumé

L'objectif de cette thèse est de contribuer à améliorer significativement le diagnostic acoustique de salles *in situ* en développant des nouvelles approches à la croisée de l'apprentissage automatique, de l'optimisation, du traitement du signal audio et de la physique acoustique. Les travaux présentés dans ce manuscrit portent sur l'estimation des propriétés d'absorption acoustique de chacune des parois d'une salle à partir de plusieurs réponses impulsionnelles de salle acquises pour des positions libres de sources et de microphones, ainsi que de mesures des paramètres géométriques et des réponses des appareils associés. Plusieurs méthodes sont proposées pour la résolution de ce problème inverse complexe en visant progressivement une description plus fine des propriétés d'absorption des parois, une application à des données simulées plus réalistes et une plus grande simplicité de mise en oeuvre. Ces travaux aboutissent à la présentation d'une approche d'estimation des réponses impulsionnelles des parois d'une salle par optimisation dans le domaine temporel. Celle-ci est fondée sur une nouvelle extension du modèle de réponse impulsionnelle de salle issu de la méthode des sources-images permettant de rendre compte des erreurs sur la mesure des paramètres géométriques, en plus de la dépendance spatio-fréquentielle des réponses des sources et des microphones.

Mots-clés : Apprentissage Automatique, Optimisation, Traitement du signal audio, Acoustique des salles, Problème inverse, Absorption, Réponse impulsionnelle de salle, Réponse impulsionnelle de mur

Abstract

The aim of this thesis is to contribute to significantly improve *in situ* acoustic diagnosis of rooms by developing new approaches at the crossroads of machine learning, optimization, audio signal processing and acoustics. The work presented in this manuscript focuses on estimating the acoustic absorption properties of each wall in a room from multiple room impulse responses acquired for free positions of sources and microphones, along with measurements of geometric parameters and responses of the associated devices. Several methods are proposed to solve this complex inverse problem, progressively aiming for a more detailed description of the absorption properties of the walls, application to more realistic simulated data, and greater ease of implementation. This work culminates in the presentation of an approach for estimating the wall's impulse responses by optimization in the time domain. This approach is based on a new extension of the room impulse response model derived from the image source method, which accounts for errors in the measurement of the geometric parameters, in addition to the spatial frequency dependence of the source and microphone responses.

Keywords : Machine learning, Optimization, Audio signal processing, Room acoustics, Inverse problem, Absorption, Room impulse response, Wall impulse response

Table des matières

Remerciements	iii
Résumé	iv
Abstract	v
1 Introduction	1
1.1 Motivation	1
1.2 Axes de recherche	2
1.3 Structure du manuscrit	5
1.4 Publications et communications	7
2 Concepts fondamentaux	9
2.1 Propagation du son dans une salle	9
2.1.1 Équation des ondes	9
2.1.2 L’acoustique des salles	10
2.1.3 Absorption, réflexion et impédance acoustiques	12
2.2 La réponse impulsionnelle de salle	15
2.2.1 Décomposition d’une réponse impulsionnelle de salle	15
2.2.2 Identification des échos	17
2.2.3 Mélange des échos	17
2.3 La méthode des sources-images	18

2.4	Impact de la chaîne de mesurage	22
2.4.1	Problématique	22
2.4.2	La directivité des sources et des microphones	23
2.4.3	La réponse en fréquence des sources et des microphones	24
2.4.4	Le bruit de mesure et le bruit ambiant	25
2.4.5	Les erreurs géométriques	27
2.5	Modèle acoustique complet	28
2.5.1	Extension du modèle issu de la méthode des sources-images	28
2.5.2	Modèle discretisé	29
2.6	Les problèmes inverses	31
2.6.1	Définition	31
2.6.2	Résolution par optimisation de fonction objectif	32
2.6.3	Résolution par utilisation de réseaux de neurones	33
2.7	Conclusion	35
3	État de l'art	37
3.1	Estimation de l'absorption acoustique d'un matériau	37
3.1.1	Les approches en laboratoire	37
3.1.2	Les approches en conditions de champ libre	40
3.1.3	Les approches <i>in situ</i>	42
3.2	Bases de données de réponses impulsionnelles de salle	44
3.2.1	Bases de données mesurées	44
3.2.2	Bases de données simulées	47
3.2.3	Librairies disponibles pour la simulation en MATLAB et Python	50

3.2.4	Les dernières avancées en matière de réalisme des simulations	51
3.3	Conclusion	53
4	Estimation pour une géométrie spécifique par apprentissage supervisé virtuel	55
4.1	Introduction	55
4.2	Simulation des bases de données de réponses impulsionnelles de salle	57
4.2.1	Paramétrage du simulateur	57
4.2.2	Prétraitement des données	59
4.2.3	Bases de données simulés	59
4.3	Architectures et entraînement	60
4.3.1	Structure des modèles	60
4.3.2	Méthodologie d’entraînement	60
4.4	Expériences et résultats	61
4.4.1	Étude quantitative : performances, robustesse et généralisation des modèles .	61
4.4.2	Étude qualitative : informativité des différentes composantes de la réponse impulsionnelle de salle	64
4.5	Conclusion	66
5	Estimation informée par la géométrie par optimisation sur les spectrogrammes des réponses impulsionnelles de salle	69
5.1	Introduction	69
5.2	Présentation de l’approche	71
5.2.1	Modélisation du signal	71
5.2.2	Fonction objectif	73
5.2.3	Analogie avec la factorisation en matrices non-négatives	74

5.2.4	Modélisation probabiliste de l'incertitude géométrique	75
5.2.5	Algorithme de sélection robuste	78
5.3	Expériences et Résultats	80
5.3.1	Cadre expérimental	80
5.3.2	Performances de la méthode	81
5.3.3	Influence du nombre d'observations	82
5.4	Conclusion	84
6	Approche par optimisation dans le domaine temporel	87
6.1	Introduction	87
6.2	Formulation du problème	89
6.2.1	Modélisation du signal	89
6.2.2	Écriture matricielle du modèle	91
6.2.3	Problème d'optimisation général	93
6.2.4	Estimation des délais entre les temps d'arrivée erronés et exacts	96
6.2.5	Estimation des réponses impulsionnelles des murs	101
6.3	Expériences et résultats	104
6.3.1	Objectifs des expériences	104
6.3.2	Paramètres de simulation	104
6.3.3	Paramètres de la méthode	107
6.3.4	Métriques d'évaluation	107
6.3.5	Estimation des délais correctifs	108
6.3.6	Estimation des réponses impulsionnelles des murs avec connaissance exacte des temps d'arrivée des échos	113

6.3.7	Correction des temps d'arrivée et estimation des réponses impulsionnelles des murs	117
6.3.8	Influence des réponses des sources et des microphones	123
6.3.9	Influence du nombre d'observations	124
6.3.10	Influence du niveau de bruit géométrique	126
6.4	Conclusion et perspectives	128
6.5	Discussions	130
7	Conclusion	131
7.1	Bilan	131
7.2	Perspectives de travaux	135
7.3	Perspectives d'applications	136
8	Annexes	137
8.1	Notions de base en traitement du signal	137
8.1.1	Signaux continus	137
8.1.2	Signaux discrets	138
	Liste des figures	141
	Liste des tableaux	147
	Liste des acronymes	149
	Notations	150
	Références	151

1.1 Motivation

Qu'elles soient professionnelles, domestiques, éducatives, artistiques ou encore récréatives, la plupart des activités se déroulent aujourd'hui dans des environnements intérieurs. Il s'agit généralement de bâtiments, composés de salles qui possèdent des fonctions spécifiques. La fonction d'une salle est souvent indiquée par sa dénomination. Une salle de classe est destinée à accueillir des élèves et des enseignants afin de permettre la transmission de savoir et l'échange de connaissances. Une salle de concert a pour fonction de réunir des musiciens et une audience afin d'assurer la tenue d'une expérience musicale dans des conditions d'écoute optimales. Une salle de réunion a pour mission de regrouper des collaborateurs professionnels présents sur place et éventuellement à distance autour de discussions de travail. Généralement, et à juste titre, une attention particulière est accordée à l'équipement ainsi qu'au mobilier nécessaires à assurer la fonction de la salle en même temps que le confort de ses usagers. Cependant, un autre aspect très important lui permettant de remplir correctement son rôle est souvent ignoré ou négligé. Il s'agit de sa qualité sonore, communément désignée comme son *acoustique*.

L'acoustique d'une salle désigne son effet sur les propriétés du son qui s'y propage et qui interagit avec ses parois. Elle dépend de la forme de la salle, de sa taille et des matériaux qui composent ses parois. Il est dit d'une salle qu'elle dispose d'une bonne acoustique lorsqu'elle permet à ses usagers de réaliser convenablement et confortablement les activités qui lui sont attribuées et peut être qualifiée de mauvaise dans le cas contraire. Il s'agit d'une caractéristique relative à la fonction de la salle et sujette à la subjectivité, bien qu'il soit en général possible de s'entendre sur la qualité acoustique d'une salle. Lorsqu'il s'agit de salles destinées à des activités pour lesquelles la qualité acoustique est fondamentale telles que des salles de spectacle ou des amphithéâtres, cet aspect est généralement placé au centre des considérations dès les premières étapes de la conception. Pour le reste des applications, la question de l'acoustique est souvent reléguée à l'arrière plan. Cependant, les conséquences d'une mauvaise acoustique peuvent être multiples et très impactantes. En effet, elle peut entraîner un déficit de l'intelligibilité, de la concentration, de l'apprentissage ou encore de la productivité. De plus, elle favorise les nuisances sonores qui sont désignées par les populations comme une des premières sources de gêne, avec des conséquences manifestes sur la santé telles

que le stress, l'anxiété, les troubles du sommeil et les pathologies cardio-vasculaires entre autres. La gêne sonore vécue par les usagers d'une salle peut être provoquée par un son provenant de l'extérieur comme de l'intérieur de la salle. Lorsque la source à l'origine de la gêne est située à l'extérieur de la salle (bruit des transports, bruit de voisinage...), le problème identifié est lié à un défaut d'isolement acoustique par rapport à l'extérieur. Lorsqu'un problème d'acoustique est constaté pour un son produit à l'intérieur de la salle, celui-ci est lié au phénomène de réverbération. Dans cette thèse, nous nous intéressons exclusivement à cette seconde situation.

La *réverbération* désigne la persistance du son dans la salle après que sa source, pouvant être une personne, un instrument de musique, ou encore un haut-parleur, ait arrêté d'en émettre. Lorsque l'acoustique d'une salle ne permet pas de contrôler la réverbération, une accumulation de l'information sonore peut se produire et entraîner une gêne chez ses usagers. Dans une telle situation, il est nécessaire de faire intervenir des acousticiens dont le rôle est de concevoir et mettre en place une solution de réhabilitation acoustique de l'existant qui soit efficace. Un des principaux leviers à disposition des acousticiens pour répondre à cette problématique est d'agir sur les propriétés acoustiques des parois de la salle, notamment en y installant précisément des éléments acoustiques spécialisés absorbeurs ou diffuseurs. Afin qu'elle puisse être efficace, cette solution doit donc nécessairement reposer sur un diagnostic acoustique précis et détaillé de l'existant. Seulement, le diagnostic acoustique reste encore aujourd'hui long, coûteux, empirique et global. En pratique, l'acousticien procède à la mesure d'indices sonores globaux qu'il utilise ensuite pour déterminer, au regard d'observations de terrain et de données de laboratoire sur les matériaux, des valeurs des paramètres acoustiques d'intérêt permettant de reproduire ces indices à partir de modèles analytiques ou numériques. Il s'agit d'une procédure manuelle, itérative, qui repose grandement sur l'expérience et les connaissances de l'expert et qui nécessite des outils hautement spécialisés. Par conséquent, le développement d'une méthode permettant de dresser *in situ* le diagnostic acoustique d'une salle de manière fiable, automatique, rapide et simple constituerait une avancée majeure dans le domaine. Un tel outil pourrait permettre à l'acousticien d'obtenir simplement une description complète et objective des propriétés acoustiques de la salle. Ainsi, il pourrait focaliser davantage son travail sur l'analyse et l'interprétation de celle-ci afin de concevoir une solution qui répondent spécifiquement aux attentes vis-à-vis de la salle. Aussi, un tel outil pourrait faciliter l'expansion du diagnostic et de la réhabilitation acoustique au-delà des applications spécialisées habituelles. L'objectif de cette thèse est de contribuer à améliorer significativement le diagnostic acoustique des salles *in situ* en développant des nouvelles approches à la croisée de l'apprentissage automatique, de l'optimisation, du traitement du signal audio et de la physique acoustique.

1.2 Axes de recherche

L'acoustique des salles est le domaine de recherche qui étudie spécifiquement le phénomène de propagation du son au sein des salles à partir des équations de la physique des ondes. Ce domaine

dispose d'une littérature mature et de pointe au sujet de la modélisation analytique et numérique de l'évolution dans le temps et l'espace du son émis dans une salle, à partir de ses caractéristiques acoustiques et géométriques ainsi que celles des sources et récepteurs utilisés. Cette évolution temporelle du champ sonore, défini comme le niveau de pression acoustique en tout point de la salle, résulte des multiples interactions entre le son et les surfaces réfléchissantes de la salle au cours de sa propagation. C'est ce qu'on appelle le *problème direct*. Cependant, de nombreux obstacles restent encore à ce jour à surmonter dans ce domaine concernant la détermination des caractéristiques acoustiques ou géométriques d'une salle à partir de mesures du champ sonore. Cet état de fait est valable en particulier pour la tâche du diagnostic acoustique, d'intérêt dans cette thèse, qui vise à déterminer les propriétés acoustiques des parois acoustiquement réfléchissantes d'une salle (murs, sol, plafond,...) à partir de mesures du champ sonore et de mesures géométriques. La tâche correspondante répond à la définition d'un *problème inverse*, qui consiste à déterminer les causes d'un phénomène à partir d'observations de ses effets. Celle-ci est communément abordée à partir d'une mesure acoustique particulière, appelée la *réponse impulsionnelle de salle*. La réponse impulsionnelle de salle est le signal audio qui caractérise l'empreinte d'une salle sur le son subséquente aux multiples interactions avec ses parois lors de sa propagation depuis une source sonore vers un récepteur à des emplacements donnés. Il existe une littérature importante en physique acoustique au sujet du calcul des principaux indices acoustiques à partir de la réponse impulsionnelle de salle et de leurs liens avec les propriétés acoustiques des parois de la salle. Cependant, les formules de référence concernent uniquement des descriptions globales des propriétés acoustiques et ne sont valables seulement pour des conditions acoustiques idéalisées et restreintes. En dehors de ce cadre limité, le problème inverse considéré reste totalement ouvert.

En parallèle, le traitement du signal audio a connu un essor considérable ces dernières années encore, notamment en bénéficiant des avancées récentes et majeures en apprentissage automatique ou intelligence artificielle, à l'instar du traitement automatique d'images et du langage naturel. Principalement, ces progrès concernent des sujets autour de l'automatisation à partir de signaux audio de l'analyse du contenu sémantique et musical, de la classification des scènes et des événements, de l'identification de sources, de la séparation de sources ou encore de la localisation de sources. Cependant, la question spécifique de la détermination des propriétés acoustiques d'un environnement et en particulier d'une salle à partir de signaux audio a reçu relativement peu d'attention jusqu'à ce jour dans la littérature concernée. Au-delà de son intérêt propre développé précédemment, il s'agit d'un sujet important pour l'ensemble des applications mentionnées qui sont impactées significativement par l'effet acoustique. Toute avancée sur ce sujet est donc susceptible de bénéficier plus largement à un large éventail d'applications en traitement du signal audio.

Cette thèse s'inscrit dans le projet ACOUST.IA : L'Intelligence Artificielle au Service de l'Acoustique du Bâtiment. Ce projet a pour mission de permettre le rapprochement et la collaboration entre les communautés de recherche en physique acoustique et en traitement du signal audio, évoluant historiquement en parallèle, afin de simplifier et d'améliorer radicalement le diagnostic acoustique.

Dans ce sens, cette thèse matérialise une collaboration entre les équipes MULTISPEECH de l'Inria, UMRAE du Cerema et de l'UGE ainsi que IMAGEs de l'ICube, réunissant des expertises en optimisation, apprentissage automatique, traitement du signal audio et image ainsi que physique acoustique. L'objectif de cette thèse est de développer une approche novatrice et transversale inspirée de ces différentes disciplines pour répondre à la problématique suivante :

Est-il possible d'obtenir automatiquement une description précise et fiable des propriétés acoustiques des parois d'une salle uniquement à partir de signaux audio mesurés à des emplacements libres, et de mesures géométriques ?

Les travaux présentés dans ce manuscrit portent sur l'estimation des propriétés acoustiques de chacune des parois d'une salle à partir de plusieurs réponses impulsionnelles de salle pour des positions arbitraires de sources et de microphones ainsi que des paramètres géométriques et de mesures des réponses des appareils. Plus précisément, ils s'intéressent aux propriétés d'absorption acoustique des parois, qui jouent un rôle majeur pour le contrôle du phénomène de réverbération dans une salle. Au fil du manuscrit, les différentes approches d'estimation proposées visent une richesse de description croissante des propriétés d'absorption des parois. De plus, faute de disponibilité de l'ensemble des données réelles nécessaires à l'évaluation de ces approches, celles-ci sont testées sur des données simulées, avec un réalisme grandissant atteignant un niveau élevé avec des expériences sur des données hybrides simulées à partir de mesures réelles de réponses d'appareils.

1.3 Structure du manuscrit

Le manuscrit de thèse est organisé comme suit. Un préambule au début de chaque chapitre rappelle sa description.

- **Chapitre 2 - Concepts fondamentaux** : Ce chapitre présente les concepts fondamentaux utiles à la bonne compréhension des modèles et approches développés dans cette thèse afin d'apporter une solution au problème de la détermination des propriétés d'absorption acoustique des parois d'une salle à partir de réponses impulsionnelles de salle. Cette thèse pluridisciplinaire aborde des notions à la croisée de la physique acoustique, de l'optimisation, de l'apprentissage automatique et du traitement du signal.
- **Chapitre 3 - État de l'art** : Ce chapitre présente l'état de l'art en matière d'estimation des propriétés acoustiques des parois d'une salle, de bases de données de réponses impulsionnelles de salle, ainsi que de simulateurs d'acoustique de salle disponibles en libre accès.
- **Chapitre 4 - Estimation pour une géométrie spécifique par apprentissage supervisé virtuel** : Ce chapitre présente une méthode pour estimer les coefficients d'absorption dans six bandes d'octave des parois d'une salle donnée à partir d'une seule réponse impulsionnelle de salle et de la connaissance des paramètres géométriques associés, aux erreurs de mesure près. Cette estimation est réalisée avec un réseau de neurones profond entraîné sur une base de données simulée et annotée de réponses impulsionnelles de salle correspondant à cette même géométrie spécifique (*apprentissage supervisé virtuel*).
- **Chapitre 5 - Estimation informée par la géométrie par optimisation sur les spectrogrammes des réponses impulsionnelles de salle** : Ce chapitre présente une méthode d'optimisation pour estimer conjointement les profils d'absorption des parois d'une salle, décrits sur une échelle linéaire des fréquences, à partir des spectrogrammes de plusieurs réponses impulsionnelles de salle. Cette méthode repose sur l'identification au sein des spectrogrammes des fenêtres temporelles de taille fixe qui incluent entièrement un unique écho acoustique isolé afin d'éviter les difficultés causées par leur superposition dans le domaine énergétique. Ces fenêtres sont identifiées en se basant sur la géométrie associée aux réponses impulsionnelles de salle, connue approximativement aux erreurs de mesure près. Une procédure probabiliste inspirée de l'algorithme *Random Sample Consensus* (RANSAC) (Bolles et al., 1981) est mise en place afin de construire un sous-ensemble de fenêtres pertinentes formant un modèle acoustique cohérent. Une fois sélectionnées, les fenêtres temporelles pertinentes sont utilisées pour estimer les paramètres d'intérêt à chaque fréquence en minimisant sous contraintes une fonction objectif définie dans le domaine des spectrogrammes.
- **Chapitre 6 - Approche par optimisation dans le domaine temporel** : Ce chapitre présente une approche d'optimisation pour l'estimation des réponses impulsionnelles des parois d'une salle à partir de réponses impulsionnelles de salle, des paramètres de la géométrie connus aux erreurs de mesure près et des mesures des réponses des sources et des microphones

utilisés. L'estimation est réalisée dans le domaine temporel en définissant deux sous-problèmes d'optimisation complémentaires. Le premier porte sur l'estimation des erreurs induites par l'incertitude géométrique sur les temps d'arrivée des échos calculés avec la méthode des sources images en optimisant sous contraintes un critère défini dans le domaine temporel continu avec l'algorithme ADAM (Kingma et al., 2014). Ce critère est défini à partir d'une extension du modèle de réponse impulsionnelle de salle issu de la méthode des sources images permettant de rendre compte de ces délais sur les temps d'arrivée des échos et suppose initialement des parois parfaitement rigides. Le second sous-problème s'intéresse à l'estimation des réponses impulsionnelles des parois de la salle, formulée comme un problème d'optimisation linéaire et multi-observation dans le domaine temporel, avec la méthode du gradient conjugué (Hestenes et al., 1952).

- **Chapitre 7 - Conclusion**
- **Chapitre 8 - Annexes**

1.4 Publications et communications

Au cours de la thèse, deux articles ont été publiés à l'occasion de conférences internationales :

- DILUNGANA, Stéphane ; DELEFORGE, Antoine ; FOY, Cédric ; FAISAN Sylvain. Learning-based estimation of individual absorption profiles from a single room impulse response with known positions of source, sensor and surfaces. In : INTER-NOISE and NOISE-CON Congress and Conference Proceedings. Institute of Noise Control Engineering, 2021. p. 5623-5630 ;
- DILUNGANA, Stéphane ; DELEFORGE, Antoine ; FOY, Cédric ; FAISAN Sylvain. Geometry-informed estimation of surface absorption profiles from room impulse responses. In : 2022 30th European Signal Processing Conference (EUSIPCO). IEEE, 2022. p. 867-871.

Les travaux présentés dans le chapitre 6 sont à paraître.

Les principales communications des travaux de recherche effectuées au cours de la thèse ont été données à l'occasion des conférences suivantes :

- Inter-noise, Août 2021, Washington D.C., États-Unis (présentation orale à distance) : Learning-based estimation of individual absorption profiles from a single room impulse response with known positions of source, sensor and surfaces ;
- Doctoriales du département Aménagement, Mobilités et Environnement (AME) de l'Université Gustave Eiffel, Octobre 2021, Le Croisic, France (présentation d'un poster en présentiel) : Intelligence artificielle au service de l'Acoustique de Bâtiment ;
- Congrès Français d'Acoustique, Avril 2022, Marseille, France (présentation orale en présentiel) : Estimation jointe des profils d'absorption des surfaces d'une salle à partir de réponses impulsionnelles ;
- European Signal Processing Conference (EUSIPCO), Août 2022, Belgrade, Serbie (présentation orale en présentiel) : Geometry-informed estimation of surface absorption profiles from room impulse responses.

Concepts fondamentaux

Préambule : Ce chapitre présente les concepts fondamentaux utiles à la bonne compréhension des modèles et approches développés dans cette thèse afin d’apporter une solution au problème de la détermination des propriétés d’absorption acoustique des parois d’une salle à partir de réponses impulsionnelles de salle. Cette thèse pluridisciplinaire aborde des notions à la croisée de la physique acoustique, de l’optimisation, de l’apprentissage automatique et du traitement du signal.

La section 2.1 aborde des notions générales liées au phénomène de propagation du son dans une salle. La section 2.2 présente la réponse impulsionnelle de salle. La section 2.3 porte sur la méthode des sources-images pour la modélisation de la réponse impulsionnelle de salle. La section 2.4 détaille l’impact de la chaîne de mesurage. La section 2.5 dévoile le modèle acoustique complet qui rend compte de l’impact de la chaîne de mesurage. La section 2.6 porte sur la résolution des problèmes inverses. La section 2.7 apporte une conclusion au chapitre.

Les notions de base en traitement du signal utilisées dans ce chapitre et dans la suite du manuscrit sont définies en annexes 8.

2.1 Propagation du son dans une salle

2.1.1 Équation des ondes

Il y a ”son” quand le milieu (dans notre cas, il s’agit d’un fluide, en l’occurrence d’air) est le siège d’une perturbation dynamique instaurée par une source acoustique modifiant localement la pression, la masse volumique, la température, l’entropie et la vitesse des particules élémentaires (ensemble de molécules occupant un volume suffisamment petit pour y attacher une grandeur, mais suffisamment grand pour que l’hypothèse de ”milieu continu” soit vérifiée).

Les relations qui lient ces variables entre elles sont obtenues par application de la loi des gaz parfaits, la loi de conservation de la masse (qui traduit la compressibilité du milieu), la loi d’Euler (équilibre des forces s’exerçant sur la particule), et la loi de conservation de l’énergie thermomécanique (qui traduit la perte de l’énergie acoustique en chaleur).

Par calcul, en combinant ces lois entre elles, en supposant la perturbation petite, il est possible d’obtenir une équation portant uniquement sur la pression acoustique, plus précisément, la variation de pression acoustique autour de la pression atmosphérique de référence. Il s’agit l’équation de propagation acoustique (ou encore équation des ondes) :

$$\nabla^2 p(\mathbf{r}, t) - \frac{1}{c^2} \frac{\partial^2 p(\mathbf{r}, t)}{\partial t^2} = -f(\mathbf{r}, t), \quad (2.1)$$

avec :

- ∇^2 l’opérateur du Laplacien,
- $p(\mathbf{r}, t)$ la pression acoustique en fonction du temps t et de la position spatiale $\mathbf{r} \in \mathbb{R}^3$,
- c la vitesse de propagation du son dans le milieu,
- $f(\mathbf{r}, t)$ la source acoustique.

Cette équation (2.1) témoigne de l’évolution dans le temps et l’espace du son, c’est-à-dire de l’onde mécanique engendrée par la source sonore, à une vitesse qui dépend des caractéristiques physiques du milieu. Si la source est impulsionnelle en temps et en espace $f(\mathbf{r}, t) = \delta(\mathbf{r} - \mathbf{r}^{\text{src}})\delta(t - t_s)$, avec \mathbf{r}^{src} et t_s respectivement la position et le temps d’émission de celle-ci, alors, sous l’hypothèse de rayonnement de Sommerfeld, la solution est la fonction de Green (notée G) suivante :

$$G(|\mathbf{r} - \mathbf{r}^{\text{src}}|, t - t_s) = \frac{1}{4\pi|\mathbf{r} - \mathbf{r}^{\text{src}}|} \delta\left(\frac{|\mathbf{r} - \mathbf{r}^{\text{src}}|}{c} - (t - t_s)\right), \quad (2.2)$$

avec $|\cdot|$ la norme euclidienne.

Cette thèse considère uniquement la propagation du son dans l’air pour des conditions atmosphériques standard et une température de 20°C. Dans ces conditions, le son se propage dans l’air à $c = 343$ m/s. Le terme $\frac{1}{4\pi|\mathbf{r} - \mathbf{r}^{\text{src}}|}$ témoigne de l’atténuation géométrique, qui correspond à une atténuation de l’énergie du son au cours de sa propagation inversement proportionnelle au carré de la distance parcourue.

2.1.2 L’acoustique des salles

Une salle est un volume rempli d’air, clos par des murs¹ d’épaisseurs finies et vide de tout objet. Lorsqu’une source sonore est placée à l’intérieur d’une salle et émet un son, celui-ci se propage dans toutes les directions de l’espace en produisant une multitude d’interactions avec les murs

1. Par abus de langage, dans toute la thèse, le terme “mur” est utilisé sans différence avec le terme “paroi” pour désigner les murs, le sol ou le plafond d’une salle.

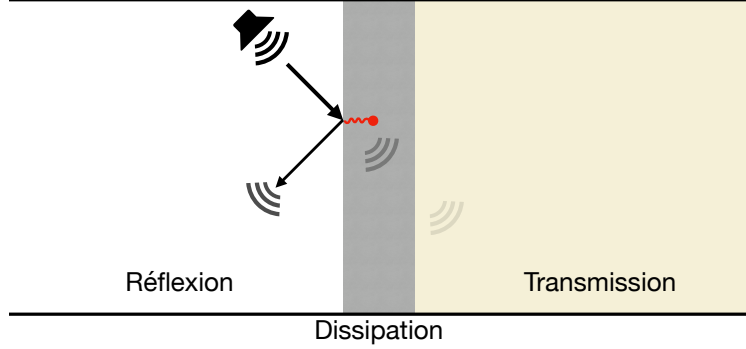


FIGURE 2.1 – Schéma illustrant l'interaction entre le son émis par une source sonore dans un milieu homogène (en blanc) et un mur d'épaisseur finie (en gris). Lors de cette interaction, le son est réfléchi, dissipé et transmis par le mur au milieu suivant (en beige).

de la salle. En effet, lorsqu'une onde sonore qui se propage dans la salle rencontre un mur, une partie du son incident est réfléchi au niveau de ce mur vers l'intérieur de la salle, tandis que le reste de son énergie est partiellement dissipée par le mur et transmise vers l'extérieur de la salle. La figure 2.1 schématise la réflexion, la dissipation et la transmission par un mur de l'énergie du son incident. Les fractions d'énergie réfléchie, dissipée et transmise au niveau d'un mur dépendent de ses caractéristiques physiques, aussi appelées *propriétés acoustiques*. En acoustique des salles, l'*absorption acoustique* d'un mur est définie par rapport à la fraction d'énergie du son incident qui n'est pas réfléchi par le mur pour un son émis dans la salle. Les propriétés d'absorption d'un mur agrègent donc les phénomènes de dissipation et de transmission.

Soit $\Omega \subset \mathbb{R}^3$ l'espace à l'intérieur d'une salle (murs exclus) et $\partial\Omega$ la frontière formée par ses murs. Afin que le problème soit bien posé, l'équation de propagation des ondes présentée précédemment doit être complétée par une condition aux limites, caractérisant les échanges acoustiques aux parois de la salle, et une condition initiale :

$$\begin{cases} [\frac{\partial^2}{\partial t^2} - c^2 \Delta]p(\mathbf{r}, t) = \delta(t)\delta(\mathbf{r} - \mathbf{r}^{\text{src}}) & \mathbf{r} \in \Omega, t \in (t_i, \infty) \\ [\frac{\partial}{\partial n} + \frac{1}{c} \frac{\partial \beta(\mathbf{r}, t)}{\partial t} *]p(\mathbf{r}, t) = 0 & \mathbf{r} \in \partial\Omega, t \in (t_i, \infty) \\ \frac{\partial p(\mathbf{r}, t)}{\partial t} = p(\mathbf{r}, t) = 0 & \mathbf{r} \in \Omega, t = t_i, \end{cases} \quad (2.3)$$

où $*$ désigne un produit de convolution en temps. La première équation est l'équation de propagation des ondes pour une source sonore impulsionnelle. La deuxième équation est la condition aux limites, avec $\beta(\mathbf{r}, t)$ l'admittance normalisée caractérisant la capacité d'absorption acoustique du matériau au point \mathbf{r} sur la paroi de la salle. La troisième équation correspond aux conditions initiales de pression à l'instant initial t_i . Dans la pratique, ce système d'équation est la plupart du temps impossible à résoudre de façon exacte. En appliquant la transformée de Fourier temporelle au système d'équation (2.3) on obtient :

$$\begin{cases} (\Delta + k^2)\underline{p}(\mathbf{r}, f) = -\delta(\mathbf{r} - \mathbf{r}^{\text{src}}) \\ [\frac{\partial}{\partial n} + jk\underline{\beta}(f)]\underline{p}(\mathbf{r}, f) = 0, \end{cases} \quad (2.4)$$

où $k = 2\pi f/c$ est le nombre d'onde et $\underline{u}(f)$ désigne la Transformée de Fourier de toute fonction $u(t)$. Cette notation est adoptée dans la suite de la thèse.

L'admittance normalisée $\underline{\beta}(f)$ est reliée à l'impédance acoustique de surface $\underline{Z}(f)$ par :

$$\underline{\beta}(f) = \frac{\rho c}{\underline{Z}(f)}, \quad (2.5)$$

où ρ est la masse volumique de l'air ($\rho = 1,293 \text{ kg/m}^3$). L'impédance, comme l'admittance, rendent compte de l'interaction physique entre le son et le matériau. De par sa définition, l'impédance correspond au rapport de la pression incidente sur la vitesse normale des particules du matériau soumis à l'onde acoustique. Elle peut être vue comme la résistance acoustique du matériau. Précisons que l'ensemble des grandeurs introduites ici sont généralement de nature complexe. Il existe plusieurs méthodes d'estimation de cette impédance basées sur la mesure de la pression face au matériau, en un ou deux points de mesures, en milieu ouvert, ou fermé. De même, lorsque les matériaux sont poreux, il existe plusieurs modèles empiriques explicitant cette impédance acoustique en fonction de différents critères mesurables et caractérisant la structure du poreux comme la résistance au passage de l'air, la tortuosité ou la porosité (Allard et al., 2009).

Malheureusement, il n'existe pas d'expression analytique générale pour la solution du système d'équation (2.3) qui permettrait d'aboutir à un modèle acoustique exploitable dans le développement de nos méthodes inverses. Seul le cas Neumann correspondant au cas particulier d'une salle dont les murs sont parfaitement réfléchissants, c'est-à-dire, parfaitement rigides ($\underline{\beta}(f) = 0$), possède une solution analytique bien connue (Allen et al., 1979). Celle-ci est obtenue à partir des modes propres de salles et de la formulation d'une onde sphérique en une somme (intégrale) d'ondes planes. Dans certains cas, la forme de cette solution est telle qu'elle peut être vue comme une somme d'ondes sphériques émises au même instant que la source sonore réelle par des sources virtuelles, les *sources-images*, construites par symétrie avec les parois de la salle. Cette solution sera discutée plus en détail en section 2.3.

2.1.3 Absorption, réflexion et impédance acoustiques

Le lien entre l'impédance acoustique de paroi $\underline{Z}(f)$ introduite précédemment et le coefficient d'absorption de celle-ci (noté par suite $\alpha(f)$), paramètre d'intérêt dans cette étude, est établi par la théorie en s'intéressant à la réflexion d'une onde plane acoustique sur cette paroi. Soit une interface

(paroi) en $x = 0$, une onde plane incidente $\underline{p}_i(\mathbf{r}, f) = p_0 e^{j(2\pi ft - k \cos \theta_i x)}$ (p_0 est une constante liée à la source) arrive en incidence oblique d'angle θ_i sur cette interface pour donner naissance à une onde plane réfléchie $\underline{p}_r(\mathbf{r}, f) = \underline{R} p_0 e^{j(2\pi ft - k \cos \theta_r x)}$ dans la direction de réflexion θ_r . Le respect de la condition aux limites donnée par l'équation 2.4 pour la pression $\underline{p}(\mathbf{r}, f) = \underline{p}_i(\mathbf{r}, f) + \underline{p}_r(\mathbf{r}, f)$ permet d'aboutir aux équations suivantes :

$$\begin{aligned} \theta_i &= \theta_r = \theta, \\ \underline{R}(f, \theta) &= \frac{\underline{p}_i}{\underline{p}_r} = \frac{\underline{Z}(f) \cos \theta - \rho c}{\underline{Z}(f) \cos \theta + \rho c}. \end{aligned} \quad (2.6)$$

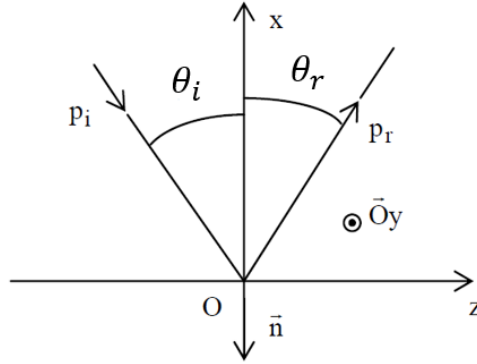


FIGURE 2.2 – Onde plane incidente \underline{p}_i et onde plane réfléchie \underline{p}_r pour une réflexion spéculaire ($\theta_i = \theta_r$)

La première équation signifie que la réflexion est de nature *spéculaire*, signifiant que l'angle de réflexion de l'onde réfléchie est égal à celui de l'onde incidente (illustré en figure 2.2). La deuxième expression relie le coefficient de réflexion complexe en onde plane, noté $\underline{R}(f, \theta)$, à l'impédance acoustique de la paroi considérée. Il dépend de l'impédance, de la fréquence et de l'angle d'incidence. La représentation temporelle équivalente du coefficient de réflexion complexe d'un mur est appelée *réponse impulsionnelle de mur* (notée **RIM** dans la suite du document). Elle est obtenue par application de la Transformée de Fourier Inverse :

$$R(t, \theta) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \underline{R}(f, \theta) \cdot e^{j2\pi ft} df. \quad (2.7)$$

Par définition, l'énergie réfléchie $E_r(f, \theta)$ par la paroi lors de la réflexion de l'énergie incidente $E_i(f, \theta)$ est égal au produit de cette dernière par $\omega(f, \theta) = |\underline{R}(f, \theta)|^2$, aussi communément appelé *coefficient de réflexion (énergétique)*. Le *coefficient d'absorption* $\alpha(f, \theta)$ de la paroi, paramètre d'intérêt dans ce travail, correspond au rapport de l'énergie absorbée $E_a(f, \theta)$ sur l'énergie incidente

$E_i(f)$ et se définit de la façon suivante :

$$\begin{aligned}\alpha(f, \theta) &= 1 - |\underline{R}(f, \theta)|^2 \in [0, 1] \\ &= 1 - \omega(f, \theta),\end{aligned}\tag{2.8}$$

Du fait que l'énergie est proportionnelle au carré de la pression, ce coefficient d'absorption se définit également comme le module au carré de la pression absorbée sur le module au carré de la pression incident. Le coefficient d'absorption $\alpha(f)$ est ensuite obtenu en moyennant sur toutes les directions θ . Il est compris entre 0 et 1 (respectivement 0% et 100% absorbant). Il s'agit du paramètre acoustique le plus utilisé dans les normes et la réglementation en acoustique du bâtiment et des salles lorsque l'on aborde la notion de correction acoustique.

En acoustique des salles, il est d'usage de considérer la gamme de fréquence allant de 100Hz à 10kHz. Il est clair que, même après échantillonnage, ceci reviendrait à devoir définir un trop grand nombre d'inconnues pour notre problème. Un filtrage passe-bande adéquat et un moyennage sont donc réalisés sur les fréquences afin d'explicitier le coefficient d'absorption $\alpha(f_b)$ portant sur six bandes d'octave centrées sur les fréquences $f_b = 125\text{Hz}$, $f_b = 250\text{Hz}$, $f_b = 500\text{Hz}$, $f_b = 1\text{kHz}$, $f_b = 2\text{kHz}$, $f_b = 4\text{kHz}$, bandes d'octave couramment considérées dans la réglementation en acoustique du bâtiment. Pour une salle possédant six parois, le nombre d'inconnues atteint dès lors 36 (6×6).

2.2 La réponse impulsionnelle de salle

2.2.1 Décomposition d’une réponse impulsionnelle de salle

Théoriquement, si les réponses de la source et du récepteur sont *idéales*, c’est à dire instantanées, indépendantes de l’angle et de la fréquence, et à bande illimitée, alors la solution de l’équation (2.3) est appelée *réponse impulsionnelle de salle*. La réponse impulsionnelle de salle (notée **RIS** dans la suite du document) est le signal particulier qui caractérise l’effet acoustique de la salle sur le son qui se propage depuis une source sonore vers un microphone tous deux positionnés dans la salle en interagissant avec ses murs. Les dimensions de la salle et les emplacements spécifiques de la source et du microphone associés à la mesure d’une RIS sont désignés comme *la géométrie* ou encore *les paramètres géométriques* associés à la RIS. En théorie, la RIS correspond pour une géométrie donnée au signal mesuré par un microphone de réponse idéale suite à l’émission d’une impulsion de Dirac par une source de réponse idéale. Dans le langage du traitement du signal, elle correspond à la réponse impulsionnelle du système linéaire invariant dans le temps (“filtre”) constitué par la salle et qui s’applique sur le son qui se propage depuis la source vers le microphone pour une géométrie donnée *via* de multiples réflexions sur ses murs. Idéalement, la RIS devrait intégrer uniquement les propriétés géométriques et acoustiques qui lui ont donné naissance. Dans la pratique, ce n’est jamais le cas, elle est parasitée pour différents signaux sonores extérieures (bruit ambiant, bruit des appareils) ou par les réponses des appareils de mesure. La figure 2.3 montre une RIS mesurée issue de la base de données dEchorate (Di Carlo et al., 2021).

La RIS est constituée d’une somme de “pics”, appelés *échos* par la suite, associés à différents trajets du son entre la source et le microphone dans la salle, impliquant des interactions multiples avec ses parois. Un écho est donc un signal temporel qui résulte du cumul des réponses temporelles des parois impliquées sur son trajet. Cet écho est de très courte durée en raison de la très brève oscillation d’un mur induite par une interaction avec le son. Étant donné que la vitesse du son est constante, les *temps d’arrivée* des échos au sein de la RIS sont proportionnels à la longueur de leurs trajets associés. Par définition, l’*ordre* d’un écho désigne le nombre de parois rencontrées sur son trajet (par exemple, un écho d’ordre 3 signifie que le son a subi trois réflexions avec les parois entre le moment de son émission et celui de son arrivée au récepteur considéré). De ce fait, les amplitudes des échos d’une RIS contiennent des informations sur les propriétés acoustiques des parois, qui nous intéressent dans cette thèse.

Classiquement, une RIS est décomposée en 3 parties : *le son direct*, *les premiers échos* (partie précoce de la RIS), et *les échos tardifs*. La figure 2.4 schématise cette décomposition. Le son direct correspond au son s’étant propagé directement de la source vers le microphone sans aucune interaction avec les murs de la salle. Les premiers échos correspondent aux échos de faibles ordres, le son ayant subi peu de réflexion avec les parois de la salle durant son trajet entre la source et

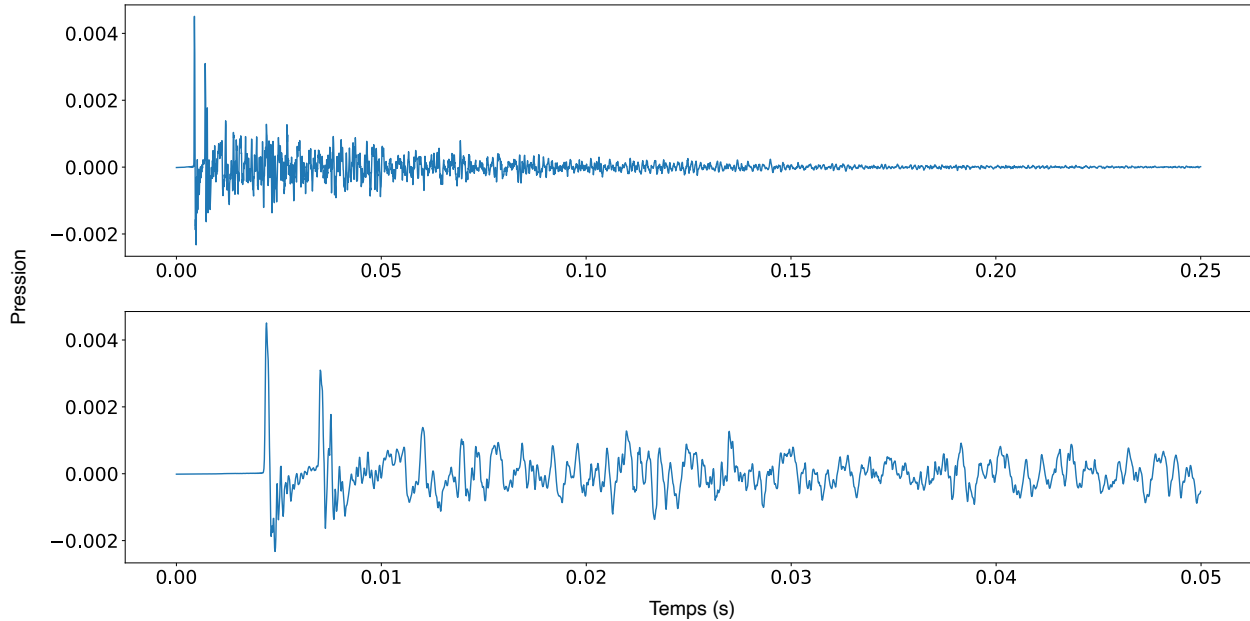


FIGURE 2.3 – Exemple d’une RIS mesurée issue de la base de données dEchorate (Di Carlo et al., 2021). Le premier graphe représente la RIS entière. Le second graphe représente les premières 50ms de la RIS.

le microphone. Les échos tardifs correspondent à la partie du champ sonore pour lequel la densité des échos dans la RIS est telle qu’ils ne sont plus distinguables les uns des autres. Ces échos sont ceux d’ordre élevé, ayant donc subi un grand nombre de réflexions avec les parois avant de parvenir au microphone. La limite entre les premières échos et les échos tardifs est marquée par le *temps de mélange*. Les modèles acoustiques les plus connus permettant de faire le lien entre la RIS et des propriétés acoustiques globale de la salle, la théorie de Sabine (Sabine, 1922) et celle d’Eyring (Eyring, 1930), reposent sur l’hypothèse cette partie tardive est issue d’un champs acoustique diffus, c’est-à-dire, identique en toutes directions et en tous points de la salle. Il faut cependant noter que cette hypothèse est en général valable seulement pour des formes de salles standards, fréquemment rencontrées, par exemple convexes avec des dimensions relativement homogènes, mais qu’elle n’est pas raisonnable pour certaines salles aux dimensions inhomogènes comme des couloirs (H. G. Davies, 1973) ou pour des formes plus atypiques.

Au regard de cette décomposition, il apparaît logique de penser que les propriétés “fines” acoustiques (absorption) et géométriques (positions des appareils, dimensions des salles (longueur, largeur, hauteur) pourront être retrouvées grâce au champ direct et au champ précoce (premiers échos) des RISs. En complément, les parties tardives des RISs qui ont un aspect très similaire entre elles, seront plus exploitable afin de retrouver des propriétés “globales” géométriques (volume et surface totale des parois) et acoustiques (absorption et diffusion moyenne de la salle).

Dans les travaux de cette thèse, nous cherchons à retrouver le coefficient d’absorption des pa-

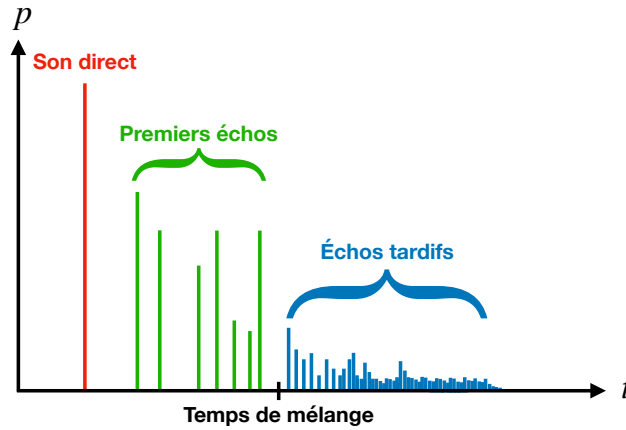


FIGURE 2.4 – Schéma de la composition d’une réponse impulsionnelle de salle.

rois au sein de ces RISs, grandeur comprise entre 0 et 1, en développant des méthodes inverses spécifiquement dédiées à cet objectif.

2.2.2 Identification des échos

Un premier verrou scientifique peut déjà être identifié ici. Il sera nécessaire de mettre en place une méthode capable de localiser automatiquement les échos dans la RIS puis d’identifier le jeu de parois réfléchissantes qui leurs sont respectivement associées afin de pouvoir estimer les profils d’absorption des murs de la salle.

2.2.3 Mélange des échos

Un deuxième verrou émerge du fait que la densité des échos augmente très rapidement comme il peut être observé en figure 2.3. Leur nombre augmente proportionnellement à t^3 (Allen et al., 1979). Ceci rend ces échos plus difficilement séparables et ce, d’autant plus qu’ils sont larges temporellement. Cette superposition des échos et cette densité complexifient fortement l’accès à leur information, nécessaire à l’estimation des profils d’absorption des parois. Dans la pratique, seuls les premiers échos sont susceptibles d’être isolés. En conséquence, il est nécessaire d’avoir recours à des techniques spécifiques capables d’extraire les échos exploitables de la RIS afin de pouvoir traiter l’information qu’ils véhiculent.

2.3 La méthode des sources-images

Comme il a été introduit en section 2.1.2, la résolution analytique du système d'équation (2.3) est difficile excepté pour le cas particulier d'une salle dont les murs sont parfaitement réfléchissants ($\beta = 0$). Dans ce cas particulier, [Allen et al., 1979](#) montrent alors que cette solution analytique reviendrait finalement à considérer un réseau infini de salles virtuelles, chacune possédant une source virtuelle impulsionnelle, appelée *une source-image*, d'amplitude égale à celle de la source réelle et émettant en même temps que celle-ci. La propagation du son dans la salle (milieu fermé), *via* les réflexions aux parois est alors assimilée à celle d'un jeu de sources-images virtuelles émettant en champ libre (milieu infini) une impulsion similaire à celle de la source réelle, ce qui s'écrit :

$$\frac{1}{c^2} \frac{\partial^2}{\partial t^2} p(\mathbf{r}, t) - \Delta p(\mathbf{r}, t) = \sum_{k=0}^{\infty} \delta(t) \delta(\mathbf{r} - \mathbf{r}_k^{src}), \mathbf{r} \in \Omega, \quad (2.9)$$

où $\{\mathbf{r}_k^{src}\}_{k \in \mathbb{N}}$ désigne la position de la source-image numérotée k , calculables par la méthode des sources-images ([Borish, 1984](#); [Ribeiro et al., 2011](#)). La solution de cette équation est alors identique à celle du système d'équation (2.3) avec $\beta = 0$.

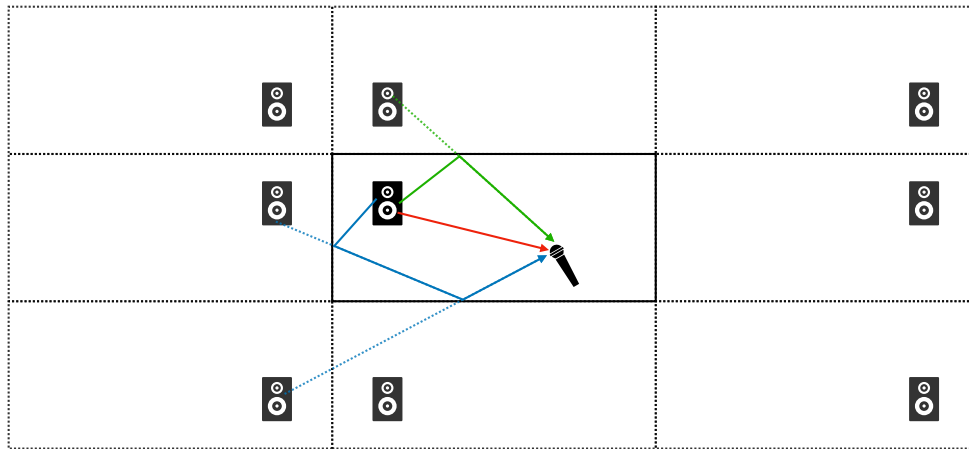


FIGURE 2.5 – Illustration en deux dimensions de la méthode des sources-images faisant apparaître le son direct, une réflexion spéculaire d'ordre 1 et une réflexion spéculaire d'ordre 2 avec les sources-images associées.

La source-image numéro k est associée à un certain ordre q signifiant qu'elle est associée à un trajet sonore allant de la source réelle au récepteur *via* q réflexions aux parois. La figure 2.5 illustre la méthode des sources-images en montrant la construction d'une source-image d'ordre $q = 1$ (en vert) et d'une source-image d'ordre $q = 2$ (en bleu). Bien qu'illustré ici pour une salle en forme de pavé droit, la méthode des sources-images s'applique pour des salles en forme de polyèdres 3D quelconques ([Borish, 1984](#)). Pour une salle à S parois, la position de chacune des S sources-images

d'ordre $q = 1$ est obtenue par symétrie de la source-réelle par rapport à chacune des S parois. Par suite, les positions des sources-images d'ordre $q = k$ sont obtenues par symétrie des sources-images d'ordre $q = k - 1$ par rapport à la prochaine paroi sur chacun de leurs trajets de réflexion spéculaires. De par cette construction, le terme $|\mathbf{r} - \mathbf{r}_k^{src}|$ de l'équation (2.2) correspondant à la distance entre la source-image k et le récepteur associée à l'équation (2.9), est alors égal à longueur du trajet sonore associé.

Par linéarité, en considérant l'équation (2.2), la solution de (2.9), notée ici x , est la somme de fonctions de Green suivante :

$$x(\mathbf{r}, t) = \sum_{k=0}^{\infty} \frac{1}{4\pi \|\mathbf{r} - \mathbf{r}_k^{src}\|_2} \delta\left(t - \frac{\|\mathbf{r} - \mathbf{r}_k^{src}\|_2}{c}\right). \quad (2.10)$$

Pour le cas où l'admittance aux parois est non nulle (ce qui est toujours le cas en pratique), une solution analytique approximative est extrapolée à partir de la solution de Neumann précédente ($\beta = 0$). De façon à rendre compte de l'énergie absorbée et du déphasage de l'onde incidente lors des q réflexions, l'onde sphérique émise par cette source-image numéro k d'ordre q est équivalente à celle émise par la source réelle convoluée avec les RIMs $\{R_i(t)\}$ des parois rencontrées lors de ces q réflexions (2.7). En notant $v_k(t)$ le cumul des convolutions temporelles des RIMs qui filtrent l'écho associé à la source-image k , on a :

$$\begin{aligned} v_k(t) &= \left(\bigotimes_{i=1}^S \bigotimes_{j=1}^{q_i} R_i \right) (t) \\ &= \left(\bigotimes_{j=1}^{q_1} R_1 \right) * \cdots * \left(\bigotimes_{j=1}^{q_S} R_S \right) (t) \\ &= \underbrace{R_1 * \cdots * R_1}_{q_1 \text{ fois}} * \cdots * \underbrace{R_S * R_S * \cdots * R_S}_{q_S \text{ fois}}(t), \end{aligned} \quad (2.11)$$

où \bigotimes est ici un produit cumulé de convolutions, S est le nombre de parois de la salle et q_i est le nombre de réflexions sur le mur i pour le trajet du son associé à la source-image k , avec $q = q_1 + \dots + q_S$. Ce produit de convolution de l'ensemble des RIMs associées à la source-image numéro k , noté $v_k(t)$, est appelé *RIM cumulée* associée à la source-image numéro k .

Afin de soulager l'écriture, dans la suite du manuscrit, les sources-images seront indexées par un uplet multidimensionnel \mathbf{k} défini comme suit :

- $\mathbf{k} = ()$ (l'uplet vide) représente la source-image associée au son direct (il s'agit de la source elle-même) ;

- $\mathbf{k} = (k_1)$ représente la source-image associée à un trajet comprenant une unique réflexion (ordre $q = 1$) sur la paroi k_1 ;
- $\mathbf{k} = (k_1, k_2)$ représente la source-image associée à un trajet comprenant deux réflexions (ordre $q = 2$), d’abord sur la paroi k_1 , puis sur la paroi k_2 ;
- $\mathbf{k} = (k_1, k_2, k_1, k_3)$ représente la source-image associée à un trajet comprenant quatre réflexions (ordre $q = 4$), d’abord sur la paroi k_2 , puis la paroi k_1 , puis de nouveau la paroi k_1 , puis la paroi k_3 ;
- etc.

Si le nombre de sources-images dans le modèle est fini et en introduisant le terme \mathcal{K} désignant l’ensemble des sources-images allant jusqu’à l’ordre fini $q = Q_{\max}$, la RIM cumulée associée à la source-image $\mathbf{k} \in \mathcal{K}$ s’écrit plus simplement :

$$v_{\mathbf{k}}(t) = \left(\bigotimes_{k \in \mathbf{k}} R_k \right) (t). \quad (2.12)$$

Par conséquent, la solution $x(\mathbf{r}, t)$ extrapolée pour le cas de cette salle absorbante s’écrit encore comme une somme de distributions de Dirac, désormais convoluée avec les RIMs cumulées pour chaque source-image :

$$x(\mathbf{r}, t) = \sum_{\mathbf{k} \in \mathcal{K}} \frac{v_{\mathbf{k}}}{4\pi \|\mathbf{r} - \mathbf{r}_{\mathbf{k}}^{\text{src}}\|_2} * \delta \left(t - \frac{\|\mathbf{r} - \mathbf{r}_{\mathbf{k}}^{\text{src}}\|_2}{c} \right) + e(t). \quad (2.13)$$

Le terme $e(t)$ est un terme résiduel capturant l’effet des sources-images d’ordre supérieur à Q_{\max} non modélisées. Pour le problème considéré dans cette thèse, il est alors attendu que les sources-images modélisent les échos d’une RIS introduits en section 2.2. D’ailleurs, dès lors qu’il ne peut y avoir de confusion, l’acronyme source-image pourra aussi être utilisé pour désigner l’écho qui lui est associé au sein de la RIS modélisée par la méthode des sources-images. La somme en équation (2.13) rend alors compte du mélange des échos souligné en section 2.2.3.

Note sur les occlusions acoustiques : Pour des géométries de salles générales, il existe des successions de murs pour lesquels il est impossible de construire un trajet de réflexions spéculaires depuis une source vers un microphone à des positions données dans la salle. Dans ce cas, la source-image correspondante n’est pas “visible” depuis la position du microphone et ne doit pas être prise en compte dans le modèle (2.13). Ce phénomène est qualifié d’*occlusion acoustique*. Dans cette thèse, nous considérons uniquement des salles en forme de pavé droit pour lesquelles la visibilité d’une source-image se vérifie aisément par le calcul à partir de la géométrie à l’aide d’un test tel qu’illustré en figure 2.6. Les méthodes proposées dans cette thèse peuvent cependant facilement

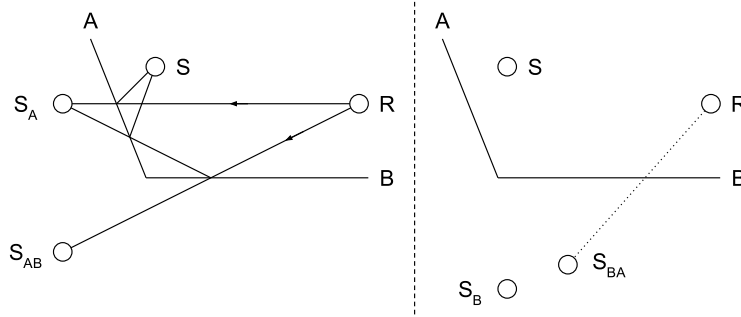


FIGURE 2.6 – Illustration d’une situation d’occlusion acoustique. Les lettres A et B représentent deux murs, R le récepteur, S la source, S_A , S_B des sources-images d’ordre 1 et S_{AB} , S_{BA} des sources-images d’ordre 2. (Gauche) Les trajets $S \rightarrow A \rightarrow R$ et $S \rightarrow A \rightarrow B \rightarrow R$ sont tous les deux valides. (Droite) Le trajet spéculaire $S \rightarrow A \rightarrow B \rightarrow R$ est invalide car $R \rightarrow S_{BA}$ n’intersecte pas A. Source : https://reuk.github.io/wayverb/image_source.html

s’étendre pour prendre en compte de telles occlusions dans le cas d’autres formes de salles (Borish, 1984).

Note sur la diffusion acoustique : Dans la pratique, les murs peuvent posséder des irrégularités (murs texturés) ou des éléments architecturaux (colonnes, sculptures) qui induisent un effet de *diffusion acoustique* du son, celui-ci pouvant alors être réfléchi en partie dans des directions aléatoires. Le modèle (2.13) évoqué ci-dessus ne prend en compte que des réflexions spéculaires. Dans cette thèse la diffusion est considéré comme une partie du terme d’erreur $\epsilon(t)$ dans l’équation (2.13). D’autres modèles d’acoustique géométrique stochastique par tirs de rayons peuvent prendre en compte le phénomène de la diffusion acoustique (Savioja et al., 2015). Le modèle général de réflexion du son sur un mur considérant les réflexions spéculaires et diffuses est illustré par la figure 2.7.

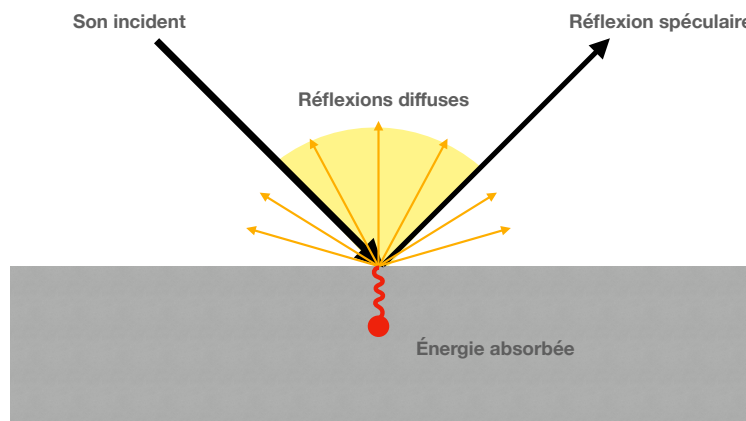


FIGURE 2.7 – Modèle général de la réflexion acoustique. Une partie du son incident est absorbée par le mur (en gris), l’autre partie est divisée entre une réflexion spéculaire et une réflexion diffuse.

2.4 Impact de la chaîne de mesurage

2.4.1 Problématique

Précédemment, nous avons introduit les équations de l’acoustique des salles dont la solution est la RIS. Étant donné que le cas général n’admet pas de solution analytique, nous avons introduit celle extrapolée et reposant sur la méthode des sources-images. Le passage de cette RIS théorique à une RIS réelle mesurée sur site nécessite de présenter ici l’ensemble des aléas existant lors du mesurage sur site. Lors des campagnes de mesurage, les principaux facteurs impactant les RISs sont la chaîne d’acquisition comprenant principalement une source sonore et un microphone, le bruit ambiant du site et les erreurs géométriques.

Comme il a déjà été souligné, si les réponses temporelles des instruments et des parois étaient instantanées et parfaites, une RIS se composerait d’échos infiniment étroits et serait solution de l’équation (2.3) Cependant, ce n’est jamais le cas dans la réalité. Ces instruments disposent de réponses propres du fait de leur structure et de leur technicité.

Chaque instrument (source et microphone) possède sa réponse propre en terme :

- de directivité, c’est-à-dire sa capacité à émettre ou capter le signal en fonction de la direction de l’espace considérée ;
- de réponse en fréquence, c’est-à-dire sa capacité à émettre ou capter le signal en fonction de la fréquence (bande passante),
- de bruit propre,
- et de niveau maximal, c’est-à-dire le niveau maximal de pression sonore au-delà duquel l’instrument ne peut plus gérer les sons de façon acceptable (forte distorsion).

Nous nous intéressons principalement à la directivité, à la réponse en fréquence et au bruit propre. Nous supposons les instruments ont été choisis de telle sorte qu’il n’y ait pas de distorsion. Par suite, nous parlerons d’“instrument idéal” lorsque sa réponse n’a aucun impact sur la RIS (instrument omnidirectionnel et indépendant de la fréquence sur la gamme de fréquence considérée).

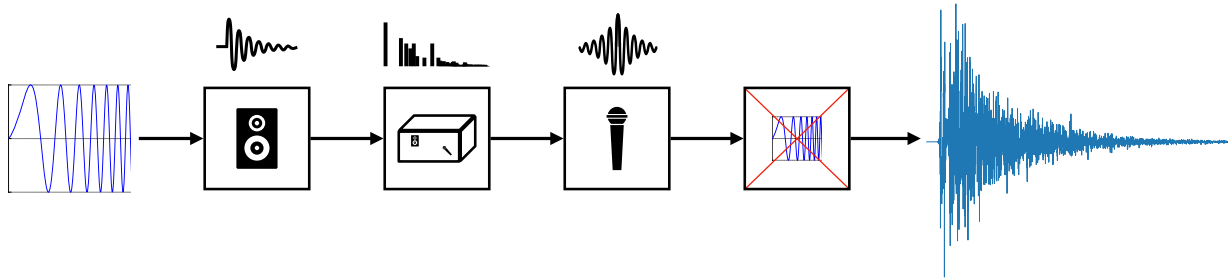


FIGURE 2.8 – Schéma d’une chaîne d’acquisition de RIS par émission d’un signal sonore “sine sweep”. Avant d’être enregistré par le microphone, le signal émis par la source est affecté successivement par la source, la salle puis le microphone. La RIS est obtenue à partir du signal mesuré en déconvoluant le sine sweep.

Les appareils de réponses idéales et les impulsions de Dirac sont des outils théoriques qui n’existent pas. Dans la réalité, les RISs sont mesurées à partir de signaux sources et d’appareils qui ont une influence manifeste. En pratique, il est possible de limiter l’influence du signal source en utilisant un signal puissant et bref afin d’“imiter” une impulsion de Dirac, tel qu’un claquement de mains ou l’éclatement d’un ballon baudruche. Il est également d’usage d’utiliser des signaux particuliers aux propriétés intéressantes et dont il est commode d’annuler l’influence sur le signal mesuré, tels que les *sine sweeps* (Angelo Farina, 2000). La figure 2.8 schématise une chaîne d’acquisition réaliste de RIS. Cependant, il est nettement plus difficile de s’affranchir de l’effet des appareils réels, dont les réponses dépendent de la direction ainsi que de la fréquence. L’effet des appareils a tendance à allonger considérablement la durée des réflexions qui se retrouvent alors mélangées dans les RISs mesurées. Pour rappel, la figure 2.3 représente une RIS réelle mesurée issue de la base de données dEchorate (Di Carlo et al., 2021).

En résumé, La RIS porte l’information des directivités des appareils, de la géométrie et des propriétés acoustiques des murs de la salle. L’absence dans la RIS de toute information sur un signal source particulier en fait le signal privilégié pour accéder aux propriétés géométriques et acoustiques d’une salle.

2.4.2 La directivité des sources et des microphones

Le son émis par une source qui se situe dans une salle peut être mesuré par un microphone qui converti l’information de variation de pression au niveau de sa membrane, induite par le son qui lui parvient, en signal électrique. Ce signal électrique peut ensuite être converti en signal numérique pour un traitement par ordinateur. Ces appareils (la source et le microphone) possèdent des réponses propres qui impactent le son et qui varient en fonction de la direction d’émission ou de réception

considérée ainsi que de la fréquence.

La *directivité* d'un appareil désigne la variation de sa réponse propre en fonction de la direction de l'espace. Un appareil dont la réponse est considérée identique dans toutes les directions de l'espace est qualifié d'*omnidirectionnel*. Dans le cas contraire, il est qualifié de *directionnel*. La figure 2.9 illustre la directivité d'un appareil pour chacun de ces cas.

Les expériences menées dans cette thèse considèrent d'une part des instruments parfaitement omnidirectionnel n'ayant pas d'effets sur le son et d'autre part des instruments réels possédants une directivité.

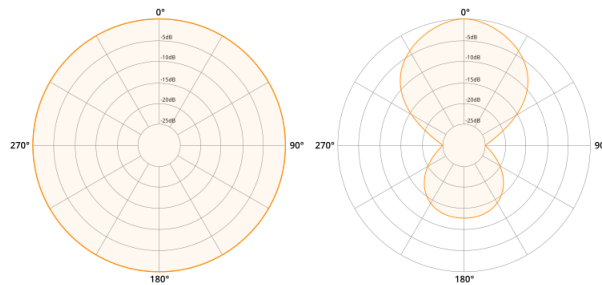


FIGURE 2.9 – Représentation de la directivité pour le cas d'un instrument omnidirectionnel (gauche) et directionnel (droite).

2.4.3 La réponse en fréquence des sources et des microphones

Une autre difficulté provient de l'effet de la source et du microphone qui s'applique à chaque écho dans la RIS, qui dépend de la fréquence. Ces filtrages apportent une information supplémentaire aux échos venant ainsi masquer et complexifier l'accès à celle liée aux parois. De plus, ces réponses en fréquence des instruments ont en général pour effet d'allonger considérablement les réflexions dans le temps et ainsi d'accentuer le mélange des échos dans la RIS, évoqué dans la section 2.2.3.

À titre de comparaison, la figure 2.10a représente une RIS simulée pour une géométrie donnée dans le cas où les réponses des instruments sont idéales et identifie le son direct et les échos d'ordre 1. La figure 2.10c représente une RIS simulée pour la même géométrie en utilisant des mesures réelles de réponses à dépendance spatiale et fréquentielle d'un haut-parleur Genelec 8020 et d'un microphone AKG C414 omnidirectionnel (publiées par Brandner et al., 2018) à l'aide d'une extension du simulateur Pyroomacoustics (Scheibler et al., 2018) développée par Srivastava et al., 2022. Ceci permet d'observer un exemple de la distorsion des réflexions provoquée par l'effet source-microphone et son impact sur le niveau de mélange dans la RIS. Dans cette situation, les superpositions entre les échos sont omniprésentes dans la RIS. La forme de la RIS simulée en utilisant des mesures réelles de réponses de source et de microphone s'apparente davantage à celle de la RIS mesurée représentée en figure 2.3.

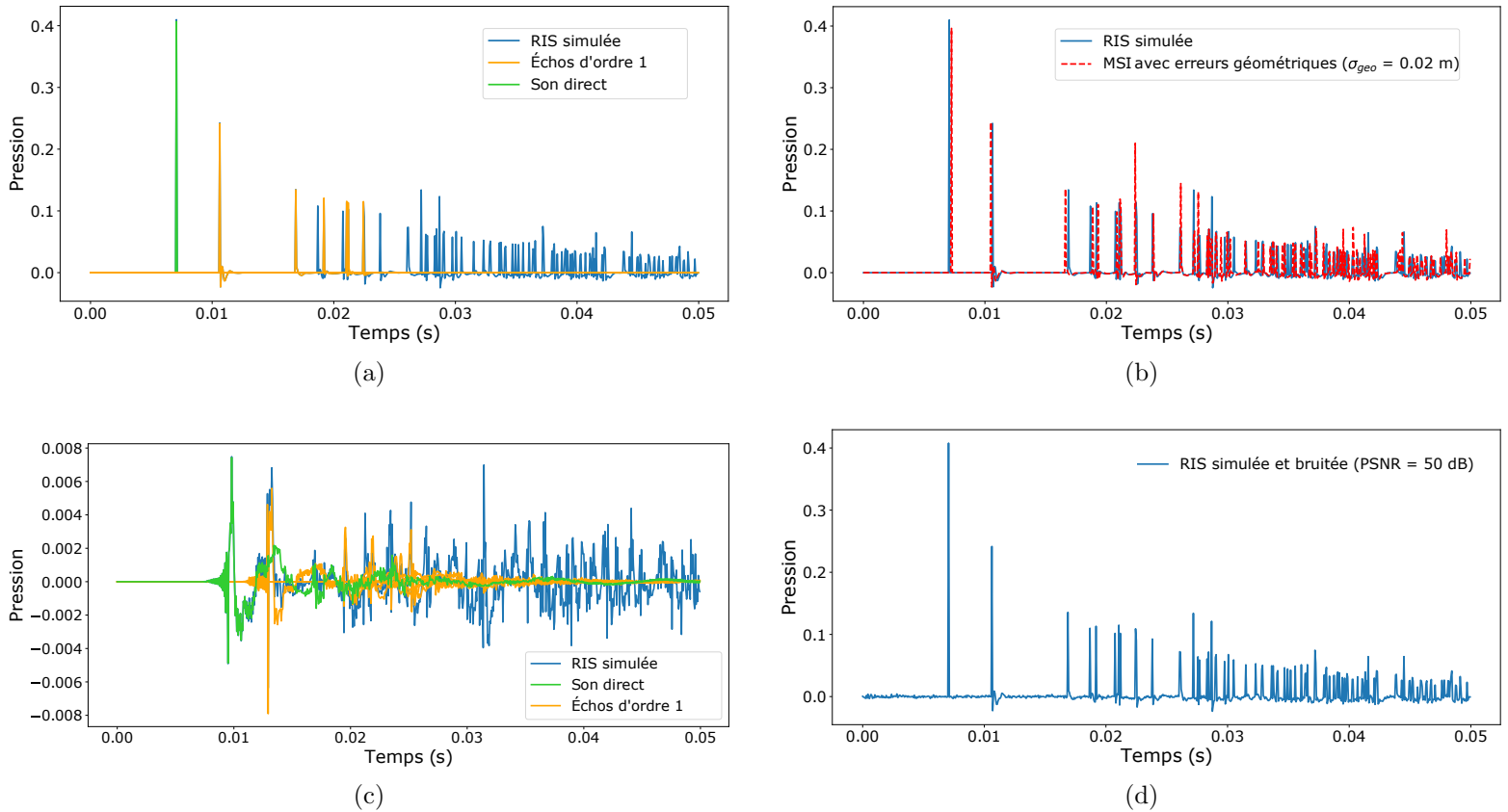


FIGURE 2.10 – Illustration par une RIS simulée (50ms), pour une salle et une géométrie fixée, de différentes problématiques intervenant dans l’estimation des profils d’absorption des surfaces d’une salle à partir de RISs. La figure 2.10a correspond à une situation “idéale” (géométrie connue exactement et appareils de réponses idéales) et sert de référence. La figure 2.10b illustre l’impact des erreurs géométriques. La figure 2.10c représente l’effet causé par la source et le microphone en utilisant des réponses mesurées d’une source et d’un microphone directionnels pour la simulation. La figure 2.10d rend compte du bruit de mesure.

2.4.4 Le bruit de mesure et le bruit ambiant

Il regroupe différents phénomènes responsables d’une distorsion du signal enregistré, généralement indépendante du signal émis, et qui se manifeste principalement par une variation aléatoire de la pression. Dans le contexte de cette thèse, plusieurs causes de bruit peuvent être identifiées. Premièrement, les appareils électroniques qui interviennent dans la mesure d’une RIS, en particulier la source et le microphone, émettent des sons principalement en lien avec les aspects thermiques de leur fonctionnement propre. Ces sons *parasites* qui s’apparentent à des vibrations sont qualifiés de *bruit propre* et viennent s’ajouter au signal mesuré. De plus, l’environnement de la salle (intérieur et extérieur) produit un son stable qualifié de *bruit ambiant* qui s’ajoute également à l’enregistrement. Dans l’expression (2.13), la modélisation du bruit de mesure est gérée par le terme résiduel $e(t-)$.

L'importance du bruit dans le signal peut être quantifiée par le *rapport signal sur bruit* (SNR pour *signal-to-noise ratio* en anglais), souvent exprimé en décibel par :

$$\text{SNR}_{\text{dB}} = 10 \log_{10} \left(\frac{P_{\text{signal}}}{P_{\text{bruit}}} \right), \quad (2.14)$$

où SNR_{dB} désigne le SNR du signal en décibels (dB), P_{signal} désigne la puissance totale du signal mesuré et P_{bruit} désigne la puissance totale du bruit mesuré en l'absence d'émission volontaire d'un signal source. Le SNR est d'autant plus faible que l'importance du bruit augmente.

Les RISs sont des signaux relativement *parcimonieux*, dont l'essentiel de l'énergie est reparti sur quelques échantillons dispersés et varie fortement en fonction des propriétés acoustiques des murs de la salle. Pour un niveau de bruit fixé, une RIS associée à une géométrie donnée peut donc présenter un SNR nettement plus élevé pour une salle globalement réfléchissante par rapport à une salle absorbante. Par conséquent, pour refléter au mieux et de manière plus stable le niveau de bruit dans la salle, nous utiliserons le *rapport signal sur bruit maximal* (PSNR pour *peak signal-to-noise ratio* en anglais) calculé par rapport à la valeur maximale du signal :

$$\text{PSNR}_{\text{dB}} = 10 \log_{10} \left(\frac{\max_{\text{signal}}^2}{P_{\text{bruit}}} \right), \quad (2.15)$$

où PSNR_{dB} désigne le SNR du signal en décibels (dB), \max_{signal} désigne le maximum du signal en valeur absolue, P_{bruit} désigne la puissance totale du bruit mesuré en l'absence d'émission volontaire d'un signal source et \log désigne le logarithme en base 10. Le PSNR est d'autant plus faible que l'importance du bruit augmente.

Dans le cas d'études portant sur des simulations, le bruit de mesure est communément modélisé en additionnant du *bruit blanc gaussien additif* aux RISs. Cette modélisation du bruit consiste à ajouter à chaque échantillon de la RIS une perturbation aléatoire de loi $\mathcal{N}(0, \sigma_{\text{bruit}}^2)$, c'est-à-dire gaussienne centrée et de variance σ_{bruit}^2 . La puissance P_{bruit} d'un bruit blanc gaussien additif est égale à sa variance σ_{bruit}^2 . Par conséquent, en utilisant (2.15), il est possible de générer un bruit blanc gaussien additif de sorte à atteindre un PSNR_{dB} fixé en choisissant la variance σ_{bruit}^2 telle que :

$$\sigma_{\text{bruit}}^2 = \max_{\text{signal}}^2 \cdot 10^{-\frac{10}{\text{PSNR}_{\text{dB}}}}. \quad (2.16)$$

Les RISs sont en général mesurées dans un cadre contrôlé ou sont traitées a posteriori pour limiter l'influence du bruit de mesure. Il n'est donc pas nécessaire en principe de mettre en place une approche de *débruitage* spécifique afin de pouvoir les traiter et les analyser. Néanmoins, il

semble important de veiller à la robustesse au bruit des approches d'estimation proposées dans le cas de niveaux de bruit modérés.

Comme évoqué précédemment en section 2.2, la diffusion acoustique peut être considérée dans une certaine mesure comme une forme de bruit aléatoire. En effet, ce phénomène peut être vu comme une redistribution aléatoire d'une fraction de l'énergie de chaque réflexion au reste de la RIS. Dans ce cas, sa modélisation peut également être intégrée au terme résiduel $e(t)$ dans (2.13), à l'instar du bruit de mesure.

La figure 2.10d représente une RIS affectée par un bruit blanc gaussien pour un PSNR égal à 50 dB. Des niveaux de bruits similaires seront considérés pour les expériences menées dans cette thèse.

2.4.5 Les erreurs géométriques

La géométrie associée à une RIS (position des murs, de la source et du microphone) n'est jamais connue exactement dans la réalité en raison de l'imprécision du processus de mesure des dimensions et des distances. Cependant, étant donnée l'importance de la vitesse du son (environ 343 m/s dans l'air), de légers écarts de la géométrie de mesurage engendrer une différence notable entre les RISs. Ces erreurs sur les paramètres géométriques affectent également la modélisation de la RIS par la méthode des sources-images qui les utilisent pour calculer les positions des sources-images et puis les temps d'arrivée des échos (2.13).

L'incertitude géométrique sera prise en compte en appliquant une erreur aléatoire de loi normale centrée, d'écart-type noté $\sigma_{\text{géo}}$, aux paramètres géométriques exacts. La figure 2.10b illustre la différence entre une RIS simulée pour une géométrie donnée (en bleu) et RIS simulée pour cette géométrie affectée par une erreur aléatoire de loi normale centrée, d'écart-type $\sigma_{\text{géo}} = 0.02$ m. Nous constatons alors qu'une incertitude sur la mesure des paramètres géométriques même faible suffit à provoquer une erreur sur la localisation des échos et sur la forme de la RI. En effet, ces différences de temps d'arrivée induites par l'erreur géométrique peuvent conduire certains échos à s'additionner, modifiant alors significativement les amplitudes.

Finalement, face à cette analyse globale, il apparaît qu'un objectif ce travail de thèse est de retenir une approche capable de pouvoir extraire dans les RISs l'information utile à la recherche des profils d'absorption, parasitée par l'information relative au mesurage. Pour cela, les approches inverses proposée doivent pouvoir se reposer sur un modèle analytique de propagation réaliste, capable d'intégrer au mieux l'impact du mesurage. La section suivante présente le modèle retenu.

2.5 Modèle acoustique complet

Cette section présente ici le modèle acoustique “direct” retenu afin de reproduire au mieux une RIS réelle compte tenu de l’ensemble des aléas soulignés précédemment.

2.5.1 Extension du modèle issu de la méthode des sources-images

Lors de ses travaux portant sur la modélisation de la RIS, [Schröder, 2011](#) propose une extension de l’expression précédente de la méthode des sources images (2.13) tenant compte de l’atténuation liée à l’air (atténuation atmosphérique), de la réponse en direction et en fréquence des instruments, ainsi que des RIMs, s’écrivant :

$$x(\mathbf{r}, t) = \sum_{\mathbf{k} \in \mathcal{K}} \left(\frac{\xi_{\text{air}}}{4\pi c t_{\mathbf{k}}} * \delta(\cdot - t_{\mathbf{k}}) * g^{\text{mic}}(\theta_{\mathbf{k}}^{\text{src}}, \psi_{\mathbf{k}}^{\text{src}}) * g^{\text{src}}(-\theta_{\mathbf{k}}^{\text{src}}, -\psi_{\mathbf{k}}^{\text{src}}) * v_{\mathbf{k}} \right) (t) + e(t), \quad (2.17)$$

avec $*$ l’opérateur de convolution, ξ_{air} l’atténuation liée à l’air, $t_{\mathbf{k}} = \frac{\|\mathbf{r} - \mathbf{r}_{\mathbf{k}}^{\text{src}}\|_2}{c}$ le temps d’arrivée de \mathbf{k} , g^{src} et g^{mic} les réponses temporelles de directivité de la source et du microphone directionnels, $\theta_{\mathbf{k}}^{\text{src}}$ et $\psi_{\mathbf{k}}^{\text{src}}$ les angles d’azimut et d’élévation de la source-image \mathbf{k} par rapport au microphone positionné en \mathbf{r} . Les fonctions $g^{\text{src}}(-\theta_{\mathbf{k}}^{\text{src}}, -\psi_{\mathbf{k}}^{\text{src}})$ et $g^{\text{mic}}(\theta_{\mathbf{k}}^{\text{src}}, \psi_{\mathbf{k}}^{\text{src}})$ désignent respectivement les réponses de la source et du microphone pour les directions d’émission et de réception associées à la source-image \mathbf{k} , déterminées à partir de la géométrie. Notons que, dans cette étude, l’atténuation atmosphérique est négligée ($\xi_{\text{air}} = \delta$), ce qui souvent le cas en acoustique des salles.

Il est à noter que la méthode des sources-images définit un système infini de sources-images. Pour une approche de résolution numérique du problème d’intérêt, il est indispensable de limiter le nombre de sources-images considérées dans le modèle afin que son coût computationnel soit raisonnable. Nous considérons donc un ordre de réflexion maximal pour la modélisation des sources-images, noté Q_{max} , dont il conviendra de choisir une valeur pertinente en veillant à inclure toutes les sources-images nécessaires à la résolution du problème. Ainsi, l’ensemble des sources-images modélisées \mathcal{K} inclut l’ensemble des sources-images d’ordre inférieur ou égal à Q_{max} . Le terme $e(t)$ est un terme résiduel capturant l’effet des sources-images d’ordre supérieur à Q_{max} non modélisées ainsi que le bruit de mesure et le bruit ambiant.

L’expression fréquentielle de x , notée alors \underline{x} , est obtenue par application de la Transformée de Fourier.

$$\underline{x}(\mathbf{r}, f) = \sum_{\mathbf{k} \in \mathcal{K}} \left(\frac{1}{4\pi c t_{\mathbf{k}}} \cdot e^{-j2\pi f t_{\mathbf{k}}} \cdot \underline{g}^{\text{mic}}(\theta_{\mathbf{k}}^{\text{src}}, \psi_{\mathbf{k}}^{\text{src}}) \cdot \underline{g}^{\text{src}}(-\theta_{\mathbf{k}}^{\text{src}}, -\psi_{\mathbf{k}}^{\text{src}}) \cdot \underline{v}_{\mathbf{k}} \right) (f) + \underline{e}(f). \quad (2.18)$$

2.5.2 Modèle discretisé

Nous définissons ici les signaux discrets échantillonnés à la fréquence d'échantillonnage \bar{f} basés sur les expressions précédentes. Nous écrivons $[\cdot]$ pour désigner l'argument discret du signal échantillonné à la fréquence \bar{f} . Dans ce cas, t et f désignent respectivement un indice temporel entier adimensionné et un indice fréquentiel discret adimensionné. La version discrète du modèle (2.17) est alors donnée par :

$$x_{\mathbf{r}}[t] = \sum_{\mathbf{k} \in \mathcal{K}} \left(\frac{\bar{f}}{4\pi c \tau_{\mathbf{k}}} \cdot \phi[\cdot - \tau_{\mathbf{k}}] * g^{\text{mic}}[\theta_{\mathbf{k}}^{\text{src}}, \psi_{\mathbf{k}}^{\text{src}}] * g^{\text{src}}[-\theta_{\mathbf{k}}^{\text{src}}, -\psi_{\mathbf{k}}^{\text{src}}] * v_{\mathbf{k}} \right) [t] + e[t]. \quad (2.19)$$

Le terme $\tau_{\mathbf{k}} = \bar{f} t_{\mathbf{k}}$ désigne le temps d'arrivée (adimensionné) de la source-image \mathbf{k} . Le filtre continu qu'est la fonction delta de Dirac $\delta(\cdot - \tau_{\mathbf{k}})$ dans l'équation (2.17) engendre un retard fractionnaire $\tau_{\mathbf{k}}$ sur le signal. Pour le cas discret, ce filtre doit logiquement être remplacé par un filtre discret, engendrant le même retard fractionnaire, tel que $\phi[t - \tau_{\mathbf{k}}] = \text{sinc}[\cdot - \tau_{\mathbf{k}}]$ où sinc est le sinus cardinal discrétisé.

Dans nos travaux, les paramètres géométriques des scènes acoustiques sont supposés connus même s'ils peuvent être entâchés d'une incertitude (erreur géométrique). Les temps d'arrivée des sources-images $\{\tau_{\mathbf{k}}\}_{\mathbf{k} \in \mathcal{K}}$ et leurs directions relativement au microphone $(\theta_{\mathbf{k}}^{\text{src}}, \psi_{\mathbf{k}}^{\text{src}})_{\mathbf{k} \in \mathcal{K}}$, calculables par la méthode des sources-images à partir des paramètres géométriques, sont donc des données du problème. Les réponses des sources et microphones g^{src} et g^{mic} sont considérées connues également. Le modèle (2.19) peut finalement être résumé comme suit :

$$x_{\mathbf{r}}[t] = \sum_{\mathbf{k} \in \mathcal{K}} \left(\underbrace{\eta_{\mathbf{k}}}_{\text{atténuation géométrique}} \cdot \underbrace{\phi_{\mathbf{k}}}_{\text{délai fractionnaire}} * \underbrace{g_{\mathbf{k}}}_{\text{directivités des appareils}} * \underbrace{v_{\mathbf{k}}}_{\text{réponses des murs}} \right) [t] + \underbrace{e[t]}_{\text{erreur du modèle}}, \quad (2.20)$$

ce qui se réduit en séparant les termes ne contenant pas l'information recherchée (absorption des parois) du terme $v_{\mathbf{k}}$ la contenant :

$$x_r[t] = \sum_{\mathbf{k} \in \mathcal{K}} (h_{\mathbf{k}} * v_{\mathbf{k}})[t] + e[t], \text{ avec } h_{\mathbf{k}} = \eta_{\mathbf{k}} \cdot \phi_{\mathbf{k}} * g_{\mathbf{k}}. \quad (2.21)$$

Le modèle fréquentiel est ensuite obtenu par application de la Transformée de Fourier Discrète (TFD) à l'équation (2.21), soit pour tout $f \in [-F, F]$,

$$\underline{x}_r[f] = \sum_{\mathbf{k} \in \mathcal{K}} (\underline{h}_{\mathbf{k}} \cdot \underline{v}_{\mathbf{k}})[f] + \underline{e}[f], \text{ avec } \underline{h}_{\mathbf{k}} = \eta_{\mathbf{k}} \cdot \underline{\phi}_{\mathbf{k}} \cdot \underline{g}_{\mathbf{k}}. \quad (2.22)$$

Considérons $\underline{x}_{\mathbf{k}}[f] = \underline{h}_{\mathbf{k}}[f] \cdot \underline{v}_{\mathbf{k}}[f]$ la contribution de la source-image \mathbf{k} dans \underline{x} à l'échantillon f . L'énergie associée est définie par :

$$\tilde{x}_{\mathbf{k}}[f] = |\underline{x}_{\mathbf{k}}[f]|^2 = |\underline{h}_{\mathbf{k}}[f]|^2 \cdot |\underline{v}_{\mathbf{k}}[f]|^2 = \tilde{h}_{\mathbf{k}}[f] \cdot \tilde{v}_{\mathbf{k}}[f], \quad (2.23)$$

avec $|\cdot|$ le module complexe. La quantité $\tilde{v}_{\mathbf{k}}[f]$ modélise donc l'effet cumulée des murs $\{k \in \mathbf{k}\}$ sur l'énergie du son qui se propage suivant le trajet associé à \mathbf{k} . Or lorsque le son se réfléchit sur un mur, une fraction de son énergie est absorbée à chaque fréquence par le mur suivant son profil d'absorption (voir equation (2.8)). En introduisant pour tout $k \in \mathbf{k}$, α_k le profil d'absorption du mur k et ω_k le profil de réflexion du mur k , la *densité spectrale d'énergie* (carré du module de la Transformée de Fourier Discrète) de la RIM cumulée associée à \mathbf{k} s'écrit à l'aide des équations (2.12) et (2.8) :

$$\tilde{v}_{\mathbf{k}}[f] = \prod_{k \in \mathbf{k}} |\underline{R}_k[f]|^2 = \prod_{k \in \mathbf{k}} \omega_k[f] = \prod_{k \in \mathbf{k}} (1 - \alpha_k[f]). \quad (2.24)$$

Par convention, $\tilde{v}_{()} = 1$ pour le son direct.

En résumé, à l'aide de la méthode des sources-images, nous avons établi un modèle de propagation (modèle direct) faisant intervenir les profils fréquentiels d'absorption des murs de la salle, la géométrie associée à la RIS et les réponses des appareils. complexe de séparation, de déconvolution et de localisation de sources-images dans les RISs qui rend très difficile l'estimation des paramètres d'intérêt. Les approches proposées dans cette thèse viseront à surpasser ces difficultés pour parvenir à une résolution du problème inverse considéré.

2.6 Les problèmes inverses

Comme nous l'avons explicité, notre objectif principal sera d'extraire les propriétés d'absorption des parois d'une RIS $x_{\mathcal{r}}[t]$ supposant que sa modélisation est proche de l'équation (2.19). Il s'agit donc bien d'un problème inverse à résoudre.

2.6.1 Définition

Les problèmes inverses sont une classe de problèmes consistant à déterminer des paramètres à l'origine d'un phénomène à partir d'observations qui en résultent. Il s'agit d'un sujet important dans une vaste gamme de domaines comme l'imagerie médicale ou encore la géologie. Dans cette thèse, il s'agit d'une application en acoustique des salles. Il est question d'estimer les propriétés acoustiques des murs d'une salle à partir de RISs.

Un problème inverse repose généralement sur une modélisation mathématique du processus de mesure d'une observation qui rend compte d'un phénomène physique. Bien que soumis à des hypothèses simplificatrices, le modèle défini a pour mission de décrire et généraliser au mieux le processus qui permet d'aboutir à des observations du phénomène à partir de paramètres initiaux.

Considérons un phénomène physique dont la modélisation admet un espace des *paramètres d'intérêt* (paramètres recherchés) noté \mathcal{U} , un espace des *paramètres connus* noté Θ et un espace des *observations* noté \mathcal{X} . Supposons que $\mathcal{U} \times \Theta$ constitue l'ensemble des paramètres inhérents au phénomène. La résolution du problème inverse correspondant consiste à estimer les paramètres d'intérêt $u \in \mathcal{U}$, pour toute mesure $(\theta, x_{u,\theta}) \in \Theta \times \mathcal{X}$ (paramètres connus et observation associée). Par la suite, nous noterons x au lieu de $x_{u,\theta}$ pour faciliter la lecture. Soit H la fonction qui associe aux paramètres $(u, \theta) \in \mathcal{U} \times \Theta$ l'observation $x \in \mathcal{X}$ correspondante. H désigne l'*opérateur direct* du problème et a pour rôle de modéliser plus proche le phénomène étudié, c'est-à-dire que pour toute mesure de (u, θ, x) , $x \approx H(u, \theta)$. Par la suite, la fonction $H(u, \theta)$ pourra être notée $H_{\theta}(u)$. Introduisons une variable aléatoire e représentant l'erreur de modélisation du phénomène par l'opérateur H ainsi que l'erreur de mesure. La modélisation complète du phénomène est alors donnée par :

$$x = H_{\theta}(u) + e. \tag{2.25}$$

2.6.2 Résolution par optimisation de fonction objectif

Lorsque qu'un modèle est défini pour le phénomène étudié, il est possible d'utiliser l'opérateur direct H_θ pour résoudre le problème inverse. Dans cette thèse, nous considérons des phénomènes physiques déterministes, c'est-à-dire que toutes conditions initiales donnent lieu à une unique observation, aux erreurs aléatoires près induites par le processus de mesure. Dans ce cas, le problème inverse est dit *bien posé* lorsque pour toute mesure $(\theta, x) \in \Theta \times \mathcal{X}$, x admet un unique antécédent $u \in \mathcal{U}$ par H_θ . En d'autres termes, il existe pour tout $(\theta, x) \in \Theta \times \mathcal{X}$ une unique solution $u \in \mathcal{U}$ au problème inverse. Le problème inverse peut alors s'écrire comme le problème de minimisation suivant :

$$\hat{u} = \underset{v \in \mathcal{U}}{\operatorname{argmin}} J_{x, H_\theta}(v), \quad (2.26)$$

où \hat{u} est une estimation du paramètre d'intérêt u et J_{x, H_θ} , appelée *critère* ou *fonction objectif*, est une fonction permettant l'estimation des paramètres d'intérêt par sa minimisation. Généralement, $J_{x, H_\theta} = C_x \circ H_\theta$, avec \circ l'opérateur de composition et C_x une fonction convexe différentiable. Par exemple, $J_{x, H_\theta}(v) = \|x - H_\theta(v)\|_2$, où $C_x(\cdot) = \|x - \cdot\|_2$ et $\|\cdot\|_2$ désigne la norme L^2 . Ce critère porte sur la reconstruction de l'observation à partir du modèle direct. Lorsque le problème est bien posé, J_{x, H_θ} admet un unique minimum.

La minimisation du critère repose la recherche des *zéros* ou *points critiques* de son gradient $\nabla J_{x, H_\theta}(v)$, c'est-à-dire des points qui vérifient :

$$\nabla(J_{x, H_\theta})(v) = 0. \quad (2.27)$$

Lorsque le problème est bien posé, la solution du problème inverse u est l'unique minimum du critère. Cependant, il n'est pas nécessairement son unique point critique du fait de l'existence de minima locaux.

La situation avantageuse se présente lorsque le critère admet un unique point critique étant la solution du problème inverse. Dans ce cas, l'estimation peut être réalisée en utilisant un algorithme de *descente de gradient*. Le principe de ces algorithmes itératifs est de diriger la recherche d'une solution dans des directions successives qui font décroître le gradient du critère dans l'objectif d'atteindre son minimum global. La figure 2.11 illustre un exemple de descente de gradient itérative d'un critère non convexe (présentant plusieurs points critiques). Pour des cas où le phénomène est trop complexe pour être modélisé ou que l'optimisation directe du critère ne permet pas d'obtenir des résultats satisfaisants mais qu'une large quantité de mesures est disponible, il est possible d'avoir recours à des réseaux de neurones pour tenter de résoudre le problème inverse.

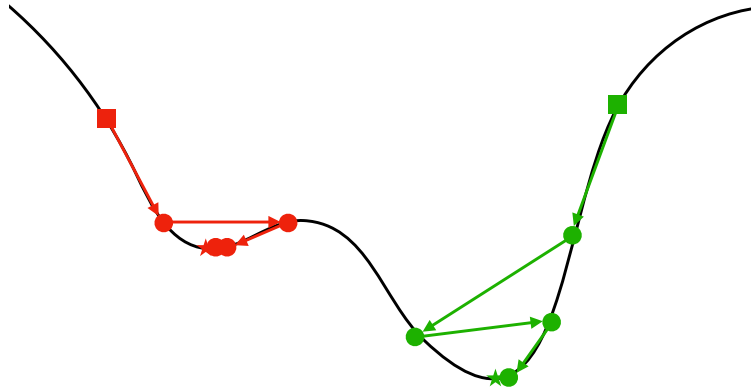


FIGURE 2.11 – Illustration de l’optimisation d’un critère non convexe par application d’un algorithme basique de descente de gradient. Quatre itérations de l’algorithme (symbolisées par des ronds) sont illustrées à partir de deux valeurs initiales distinctes (symbolisées par des carrés). La trajectoire verte montre une convergence vers le minimum global du critère (étoile verte) tandis que la trajectoire rouge converge vers un minimum local du critère (étoile rouge) .

2.6.3 Résolution par utilisation de réseaux de neurones

Lorsque l’opérateur direct problème inverse H est trop complexe pour permettre l’estimation par une approche d’optimisation directe ou bien qu’il est difficile de modéliser le phénomène en jeu, il est possible d’avoir recours à des modèles d’*apprentissage automatique* pour tenter de résoudre le problème inverse. L’apprentissage automatique consiste à optimiser des modèles paramétriques sur une vaste quantité de données pour la réalisation d’une tâche d’intérêt. Les modèles d’apprentissage les plus reconnus et utilisés aujourd’hui sont les réseaux de neurones profonds. Ces modèles sont constitués d’un empilement de couches de neurones artificiels (des fonctions non linéaire paramétriques élémentaires) par des liaisons pondérées. À partir d’un signal d’entrée, chaque couche du réseau de neurones opère une transformation non linéaire des neurones de la couche précédente, résultant en un signal de sortie. Il existe différentes architectures de réseaux de neurones qui possèdent des structures propres et qui sont privilégiées pour résoudre des tâches particulières. Parmi certaines des architectures les plus utilisées figurent les perceptrons multicouches (Rosenblatt, 1958) (Gardner et al., 1998), les réseaux convolutifs (LeCun et al., 1989), les réseaux récurrents (Hochreiter et al., 1997), les réseaux génératifs antagonistes (Goodfellow et al., 2014) ou encore les réseaux Transformers (Vaswani et al., 2017). La figure 2.12 schématise un perceptron multicouches. L’*entraînement (supervisé)* d’un réseau de neurones repose sur la constitution d’une base de données de signaux d’entrée et de signaux cibles associés. L’entraînement consiste alors à ajuster les paramètres du modèle de sorte à minimiser, pour tout signal d’entrée, une erreur entre le signal en sortie du réseau de neurones et le signal cible. Cette erreur est calculée par une *fonction de coût*. La mise à jour des paramètres du réseau de neurones est réalisée itérativement à partir des erreurs par *rétropropagation du gradient* (Hecht-Nielsen, 1992). Cette procédure consiste à actualiser les paramètres des couches du réseau depuis sa sortie jusqu’à son entrée, en les modifiant

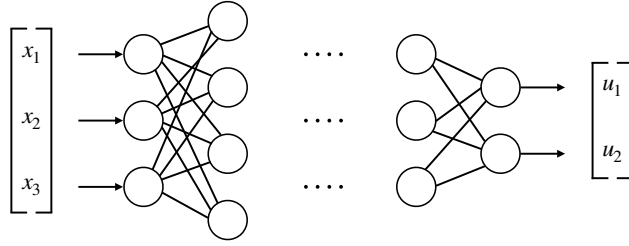


FIGURE 2.12 – Schéma d'un perceptron multicouche

davantage lorsqu'ils contribuent fortement à l'erreur commise.

Un réseau de neurones est évalué sur sa capacité à généraliser ses bonnes performances acquises sur un *jeu de données d'entraînement* à un *jeu de données de test* disjoint mais similaire. Typiquement, un *jeu de données de validation* disjoint mais similaire aux deux précédents est utilisé durant la phase d'entraînement pour contrôler l'évolution de la capacité de *généralisation* du modèle et éviter le *surapprentissage*, c'est-à-dire la spécialisation de ce dernier aux données d'entraînement uniquement.

L'emploi d'un réseau de neurones pour la résolution d'un problème inverse consiste à l'utiliser comme un opérateur inverse. Soit $\tilde{H}_{\theta, \psi}$ l'opérateur inverse modélisé par un réseau de neurones avec ψ les poids du modèle. $\tilde{H}_{\theta, \psi}$ a pour rôle de modéliser au mieux l'inverse du phénomène étudié, c'est à dire, s'assurer que pour toute mesure de (u, θ, x) , $u \approx \tilde{H}_{\theta, \psi}(x)$. En d'autres termes, le réseau de neurones est employé pour approximer les paramètres d'intérêt par une fonction non linéaire des paramètres connus et des observations.

Dans le cadre du problème étudié dans cette thèse, \mathcal{X} correspond aux réponses impulsionnelles, \mathcal{U} renvoi aux profils d'absorption des murs, Θ désigne aux paramètres géométriques ainsi que les directivités des sources et microphones.

2.7 Conclusion

Dans ce chapitre, nous avons introduit les concepts fondamentaux en lien avec le problème inverse d'estimation des propriétés d'absorption acoustique des surfaces d'une salle à partir de réponses impulsionnelles.

En résumé, les propriétés acoustiques d'une salle déterminent la manière dont celle-ci agit sur le son qui s'y propage. Elles sont très étroitement liées aux profils d'absorption de ses surfaces, qui décrivent précisément la manière dont l'énergie du son est absorbée par chacune d'entre-elles. Elles constituent donc une information nettement plus fine sur le comportement acoustique de la salle que les indicateurs globaux ou moyens communément utilisés. Tout signal sonore enregistré dans la salle est porteur de cette information acoustique. La RIS, indépendante de tout signal source, constitue une mesure privilégiée du champ sonore dans la salle pour tenter d'accéder à cette information. La méthode des sources-images définit un modèle direct de la RIS à partir notamment des réponses des murs pouvant être utilisé pour tenter de définir une approche numérique de résolution de ce problème inverse. Il s'agit d'un problème inverse ouvert et complexe qui aborde des thématiques diverses en lien avec la localisation de motifs, la séparation de sources, la déconvolution et la robustesse.

Le prochain chapitre dresse l'état de l'art concernant les avancées scientifiques récentes et connexes au problème inverse étudié dans cette thèse.

Préambule : Ce chapitre présente l'état de l'art en matière d'estimation des propriétés acoustiques des parois d'une salle, de bases de données de réponses impulsionnelles de salle, ainsi que de simulateurs d'acoustique de salle disponibles en libre accès.

3.1 Estimation de l'absorption acoustique d'un matériau

Pour comprendre le comportement acoustique d'une salle et pouvoir agir dessus de sorte à améliorer la perception du son pour une application visée ou bien le simuler, il est crucial de connaître précisément les propriétés acoustiques de ses parois. Cependant, l'estimation des propriétés acoustiques d'un matériau est par nature un problème inverse très complexe. Il s'agit d'ailleurs d'un domaine de recherche à part entière, dans lequel les questions du niveau requis pour la description de ces propriétés acoustiques ou des techniques à employer pour leur estimation ne font pas toujours consensus et restent ouvertes.

L'estimation de l'absorption acoustique d'un matériau a fait l'objet de nombreux travaux. Il existe une grande variété d'approches chacune ayant ses propres hypothèses et limites ainsi que sa propre complexité. En s'appuyant sur la classification proposée par (Brandão et al., 2015), nous proposons de regrouper les méthodes en trois groupes principaux que nous résumons par la suite :

- les méthodes en laboratoire (salle réverbérante, méthode du tube d'impédance) ;
- les méthodes en conditions de champ libre (extérieur, chambres anéchoïques, ...)
- les méthodes *in situ*, c'est-à-dire en salle réelle.

3.1.1 Les approches en laboratoire

Estimation de l'absorption via une salle réverbérante :

La méthode la plus connue est celle de la *salle réverbérante*, basée sur la courbe de décroissance énergétique du niveau sonore (intégration de Schroeder de la RIS, Schroeder, 1965). Elle repose sur la théorie de la réverbération qui lie l'absorption du matériau au *temps de réverbération*, durée

que met le niveau sonore à décroître de 60 dB dans une salle dès lors que la source sonore est interrompue.

La salle réverbérante est une salle conçue pour vérifier l'hypothèse de champ diffus, c'est-à-dire dans laquelle le champ de pression tend à devenir identique en tout point de l'espace (*uniforme*) et pour toutes les directions (*isotrope*) (Schroeder, 1965). Sous cette hypothèse, elle permet l'application de la loi de Sabine (Sabine, 1922), qui lie le temps de réverbération dans la salle à la moyenne des coefficients d'absorption des murs de la salle :

$$\text{TR}_{\text{salle}} \approx \frac{0.16V_{\text{tot}}}{\bar{\alpha}_{\text{salle}}S_{\text{tot}}}, \quad (3.1)$$

avec TR_{salle} le temps de réverbération dans la salle (supposé identique en tous points), $\bar{\alpha}_{\text{salle}}$ la moyenne pondérée des coefficients d'absorption des murs de la salle (pondérée par leurs aires respectives), S_{tot} l'aire totale de la salle et V_{tot} son volume total.

Le coefficient d'absorption α du matériau d'étude est estimé en effectuant deux mesurages du temps de réverbération, l'un dans la salle réverbérante vide (équation (3.1)) et l'autre en présence du matériau. En faisant l'hypothèse d'un champ diffus dans la salle et que l'aire du matériau est faible devant celle des murs de la salle, le temps de réverbération dans la salle en présence du matériau s'écrit :

$$\text{TR}_{(\text{salle} + \text{mat})} \approx \frac{0.16V_{\text{tot}}}{\bar{\alpha}_{\text{salle}}S_{\text{tot}} + \alpha_{\text{mat}}S_{\text{mat}}}, \quad (3.2)$$

avec $\text{TR}_{(\text{salle} + \text{mat})}$ le temps de réverbération dans la salle en présence du matériau (supposé identique en tous points), α_{mat} le coefficient d'absorption du matériau étudié et S_{mat} son aire totale. Par conséquent, le coefficient d'absorption du matériau étudié peut être déterminé à partir des équations (3.1) et (3.2) :

$$\alpha_{\text{mat}} \approx \frac{0.16V_{\text{tot}}}{S_{\text{mat}}} \left(\frac{1}{\text{TR}_{(\text{salle} + \text{mat})}} - \frac{1}{\text{TR}_{\text{salle}}} \right). \quad (3.3)$$

Malheureusement, cette méthode donne fréquemment lieu à des variations significatives des estimations effectuées par différents laboratoires, malgré les standards de conception et de mesure existants (ISO 354, 2003 ; ASTM C423, 2022). Vercammen, 2010 met en lumière la faible reproductibilité des résultats d'estimation entre les laboratoires avec des pistes d'amélioration éventuelles.

De plus, il démontre l'occurrence fréquente d'estimations aberrantes attribuées à l'effet de diffraction causé par les bords des échantillons de matériau (*edge effect*). Pour cette raison, les coefficients d'absorption estimés en chambre de réverbération sont communément appelés *coefficients de Sabine* et sont différenciés de ceux théoriques qui correspondent à l'hypothèse d'un champ réverbérant parfaitement diffus, appelés *coefficients d'absorption d'incidence aléatoire*. Différents travaux tels que [Jeong, 2013](#) proposent des méthodes pour estimer les coefficients d'absorption d'incidence aléatoire à partir de ces coefficients de Sabine.

Une première explication des disparités observées est que l'hypothèse de champ diffus est un cadre théorique idéal qui est plus ou moins vérifié selon les différentes chambres de réverbération ([Zhang et al., 2021](#)). En effet, [Nolan et al., 2020](#) et [Berzborn et al., 2019](#) démontrent le caractère anisotrope du champ sonore dans plusieurs chambres de réverbération standardisées et démontrent que ce phénomène est accentué par l'introduction d'une surface absorbante dans la chambre de réverbération. [Vorländer, 2013](#) affirme que les coefficients d'absorption mesurés en chambre de réverbération produisent des incertitudes telles qu'ils ne permettent pas de simulation haute précision car ces erreurs conduisent à des différences perceptibles par l'oreille humaine. [Balint et al., 2023](#) retracent les évolutions de la méthode et des questionnements sur sa validité depuis sa création jusqu'à aujourd'hui.

Le tube d'impédance :

En environnement maîtrisé (laboratoire), la méthode du *tube d'impédance* ou du *tube de Kundt*, permet d'estimer l'impédance acoustique d'un matériau. Elle est l'une des méthodes plus connues et a fait l'objet de plusieurs normes ([ISO 10534-1, 1994](#) ; [ISO 10534-2, 1994](#) ; [ASTM, 1998](#)). Cette méthode est fondée sur la génération d'une onde stationnaire unidimensionnelle dans un tube cylindrique fermé en plaçant une source à l'une des extrémités et le matériau d'intérêt sur la face opposée. Cette onde stationnaire résulte de l'interaction entre le son direct émis par la source et le son réfléchi sur le matériau. Le dispositif permet ainsi de déterminer le coefficient d'absorption en *incidence normale* du matériau à partir d'une fonction de transfert calculée pour deux mesures de la pression à des emplacements distincts dans le tube ([Chung et al., 1980a](#) ; [Chung et al., 1980b](#)).

Bien que relativement pratique et simple d'utilisation, cette méthode présente plusieurs limites. D'abord, elle permet uniquement la mesure du coefficient d'absorption pour une incidence normale ce qui constitue une description restreinte des propriétés acoustiques du matériau. En effet, il est reconnu que la direction d'incidence du son sur un matériau peut influencer significativement sur sa réponse et que la prise en compte de cette dépendance angulaire joue un rôle crucial dans la modélisation du phénomène de propagation du son dans la salle. Dans ce sens, [Jeong et al., 2014](#) déterminent dans un cadre simulé que la variation de la phase angulaire de l'impédance impacte significativement le champ de pression dans la salle ainsi que le temps de réverbération. De plus,

ils concluent à l'issue d'un test d'écoute qu'elle impacte la perception de la réverbération dans la salle. Aussi, malgré l'existence des normes de mesurage citées précédemment visant à uniformiser les pratiques, des variations significatives du coefficient d'absorption du matériau sont observées entre deux mesurages. En effet, ces estimations peuvent fluctuer significativement en fonction de la dimension et du montage de l'échantillon dans le tube d'impédance (Cummings, 1991), mais aussi du fait de l'influence propre du tube d'impédance et de son choix qui ne sont pas négligeable en général (Vigran et al., 1997). Stender et al., 2021 confirment avec une approche par apprentissage automatique que l'ensemble des facteurs cités participent à la faible reproductibilité observée de l'estimation des coefficients d'absorption d'incidence normal par ce type de méthodes. En plus de ces limitations, les données mesurées disponibles concernant l'impédance (ou de l'admittance) sont toutefois peu nombreuses.

Nous pouvons également souligner que, du fait notamment des normes associées à ces méthodes en laboratoire et de leurs contraintes propres (taille du tube d'impédance et de la salle réverbérante par exemple), l'absorption des matériaux est estimée pour des dimensions (surface totale et épaisseur) et des conditions de montage qui sont imposées. Malheureusement, celles-ci sont la plupart du temps différentes pour l'application *in situ*. Les propriétés acoustiques des parois mesurées en laboratoire et celles s'exerçant sur site sont donc susceptibles de différer significativement. Dans le cadre d'un diagnostic acoustique d'une salle existante, nous pouvons nous appuyer sur les bases de données d'absorption issues des mesures en laboratoire, mais cela ne peut être qu'une indication. Idéalement, il faudrait pouvoir estimer l'absorption directement sur site, à partir de méthodes développées à cet usage. Comme nous le verrons, ceci reste encore un objectif à atteindre.

3.1.2 Les approches en conditions de champ libre

La mesure en *conditions de champ libre* consiste à placer un matériau d'intérêt dans un environnement dit *pseudo-anéchoïque*, c'est-à-dire ne produisant pas de réflexions acoustiques parasites, afin d'effectuer des mesures acoustiques comportant uniquement le son direct et l'écho issu d'une réflexion du son sur ce matériau. Il peut s'agir d'une salle anéchoïque (ou semi-anéchoïque), d'un environnement extérieur ou bien de se placer suffisamment à distance de tout autre réflecteur acoustique indésirable lorsque cela est réalisable.

Ces mesures peuvent alors être exploitées pour extraire la réponse acoustique du matériau, par exemple la réponse impulsionnelle de mur (RIM, voir section 2.1.3). Sous cette appellation, nous regroupons également toutes les formulations analytiques du champ sonore au dessus (ou à proximité) du matériau à tester. Ces formulations explicitent uniquement le champ incident et le champ réfléchi par le matériau et ne considèrent pas le champ qui pourrait être issu d'autres matériaux comme cela est le cas d'une salle.

Les premières méthodes développées sont les méthodes d'écho-impulsion avec l'émission d'une signal qui se rapproche d'une impulsion. L'onde réfléchie p_r est alors extraite soit par un fenêtrage temporel, soit par annulation de l'onde incidente p_i (par exemple par soustraction) en supposant que cette dernière est connue ou en l'estimant dans des conditions pseudo-anéchoïque. Dans les deux cas, il existe plusieurs difficultés à résoudre. La première difficulté réside dans le fait de proposer des approches capables de bien différencier les deux ondes (incidente et réfléchie). Ceci peut être fait à partir d'un mesurage avec deux microphones (Cramond et al., 1984; Yuzawa, 1975), l'un proche et l'autre éloigné du matériau d'intérêt (Garai, 1993). D'autres approches reposent sur deux mesurages (J. Davies et al., 1979), l'un *in situ* (dans la salle) et l'autre en plaçant le matériau en extérieur ou en salle anéchoïque. Une autre approche consiste à réaliser ce deuxième mesurage au dessus d'un matériau de référence hautement absorbant pour déterminer le son direct (Hollin et al., 1977; Mommertz, 1995). Du fait du principe d'incertitude entre le temps et la fréquence, ces méthodes qui reposent sur une information temporelle très précise ont aussi intrinsèquement une limitation en basses fréquences. Il y a donc un compromis à trouver entre la taille des fenêtres temporelles et l'information basses fréquences que l'on accepte de perdre.

Une seconde difficulté consiste à créer un signal source soit le plus impulsionnel possible (ou neutre après post-traitement), ce qui est essentiel pour limiter l'étalement temporel des échos. Dans les première années, plusieurs auteurs ont cherché à y parvenir (Yuzawa, 1975; J. Davies et al., 1979) mais se sont très vite confrontés à des difficultés de répétabilité, de non-linéarité et de post-traitement. Idéalement, afin d'avoir une bonne séparation dans le domaine temporel, l'impulsion émise doit être étroite et de spectre en fréquence plat ce qui est difficile dans la pratique (Cramond et al., 1984). Pour outrepasser ces limitations, des méthodes basées sur des bruits stationnaires particuliers ont été proposées [10, 27], la RIS étant ensuite obtenue par déconvolution. Barry, 1974 et Hollin et al., 1977 utilisent un bruit blanc. Plus tard, Aoshima, 1981 puis, Suzuki et al., 1995 proposent un signal impulsionnel de spectre plat étiré dans le temps par filtrage (déconvolution passe-tout). D'autres signaux d'excitation particuliers ont ensuite été développés dans le but notamment d'offrir une meilleure immunité au bruit de fond lors de la mesure *in situ* : *MLS* (*Maximum Length Sequence*) (Schroeder, 1979; Rife et al., 1989; Stan et al., 2002) et *Sine Sweep* linéaire (Müller et al., 2001) ou exponentiel (Angelo Farina, 2000; Angelo Farina, 2007). À ce jour, les avantages et inconvénients de chacun de ces signaux font encore l'objet de plusieurs travaux (Torras-Rosell et al., 2010; Guidorzi et al., 2015). L'un des aspects positifs de ces méthodes en conditions de champ libre réside dans le fait qu'elles pourraient théoriquement permettre d'estimer le coefficient d'absorption du matériau en fonction de l'angle d'incidence de l'onde, soit $\alpha(\theta)$. Cependant, ceci nécessiterait des moyens considérables à mettre en oeuvre, comme par exemple de disposer d'une chambre anéchoïque (Richard et al., 2017), de bras robotiques pour automatiser la mesure précise de la réponse de la paroi en fonction de l'angle (Dupont et al., 2020) ou de capteurs très spécifiques et sophistiqués (Lanoye et al., 2006; Lanoye et al., 2004).

3.1.3 Les approches *in situ*

En raison des limites et des contraintes de ces approches développées précédemment, des efforts grandissants sont produits par la communauté scientifique afin de proposer des approches d'estimation *in situ* qui soient non destructives, les plus flexibles et accessibles possibles et qui permettent d'aboutir à une connaissance approfondie des propriétés acoustiques de chacune des parois de la salle. La difficulté principale de l'estimation *in situ* est qu'elle est sujette à un contexte réverbérant quelconque instauré par l'ensemble des parois de la salle.

Nous avons donc décidé de résumer les approches *in situ* existantes se rapprochant de notre objectif, à savoir améliorer le diagnostic acoustique de salle à partir de la seule connaissance du champ sonore réverbéré de la salle. Ces approches cherchent estimer des indicateurs acoustiques globaux de salle (temps de réverbération ou absorption moyenne) ou des descriptions plus précises des propriétés acoustiques des parois (impédance, coefficient d'absorption).

Plusieurs approches, motivées par le *ACE Challenge* (Eaton et al., 2016), concernent l'estimation du temps de réverbération par bande d'octave à partir de méthodes d'apprentissage automatique dites "aveugles". Cela signifie que l'estimation est opérée à partir de mesures du champs sonore uniquement, sans connaissances annexes telles que les dimensions des salles, les positions des sources et des microphones ou tout autre paramètre ayant une influence sur le champ sonore. Dans ce sens, nous pouvons citer les travaux de Bryan, 2020, Looney et al., 2020, Ciaburro et al., 2021 et de Duangpummet et al., 2022. Ces approches reposent sur l'hypothèse d'un champ réverbéré diffus, partant du principe que le temps de réverbération est identique en tout point de l'espace. En pratique, cette hypothèse n'est jamais strictement valide dans de vraies salles. Elle réduit donc le champ d'application de ces approches, nécessairement limitées à certaines formes de salle et de combinaisons d'absorption des parois. Perez et al., 2019 proposent une preuve de concept pour l'estimation du temps de réverbération pour des emplacements particuliers de la source et de microphone dans une salle, par apprentissage supervisé à partir de RISs, de la géométrie et d'une donnée approximative des profils d'absorption en bande d'octave des parois de la salle. Cette approche a été validée sur des données réelles uniquement pour une configuration géométrique et nécessite une quantité importante d'informations annexes. D'autres travaux s'intéressent à estimer conjointement le temps de réverbération ainsi que le volume de la salle (Ick et al., 2023) et sa surface totale (Srivastava et al., 2021). De même que pour les travaux précédents, le cadre d'application reste également limité. Afin de dépasser la limite imposée par l'hypothèse de champ diffus, Foy et al., 2021 proposent une approche d'apprentissage automatique sur des données synthétiques pour estimer directement à partir d'une RIS la moyenne des profils d'absorption des murs d'une salle. Cette approche a été validée sur des données réelles.

Récemment, plusieurs travaux ont cherché à transposer *in situ* les approches en conditions de champ libre présentées précédemment (section 3.1.2). La difficulté consiste à pouvoir estimer

l'absorption de chaque paroi à partir d'une mesure acoustique prise à proximité de celle-ci. L'objectif est d'éliminer ou de limiter l'influence des réflexions provenant des autres parois de la salle. Récemment, [Hald et al., 2019](#) proposent une méthode afin d'estimer l'impédance d'un matériau à partir de la mesure du champ de pression et de la vitesse normale particulière à son voisinage. Par suite, [Nolan, 2020](#) propose une méthode permettant de séparer spatialement l'énergie incidente et l'énergie réfléchie dans le champ de sonore avoisinant, dans le but d'estimer le coefficient d'absorption oblique $\alpha(\theta)$ d'un matériau. Ces approches présentent cependant un protocole de mesurage très est difficile à mettre en oeuvre. En effet, l'ensemble de ces techniques utilisent des appareils hautement spécialisés et devant être placés à des emplacements précis pour les besoins de l'approche. De plus, ces approches sont de nature chronophages et doivent être réitérées par chacune des parois présentes dans la salle. Par ailleurs, sous certaines conditions, ces approches deviennent mêmes inexploitable lorsqu'il est impossible de s'affranchir des contributions des autres parois avoisinantes pour effectuer des mesures acoustiques dédiées à l'étude d'une paroi d'intérêt. Face à ces limitations, une direction de recherche attrayante serait de pouvoir simplifier cette estimation *in situ* des propriétés acoustiques des parois d'une salle. L'approche visée devra être capable d'estimer conjointement l'absorption acoustique de l'ensemble des murs d'une salle et ce, à partir d'un nombre limité de mesures du champ sonore, idéalement prises à des emplacements libres dans la salle. Du meilleur de notre connaissance, ce problème inverse difficile a reçu relativement peu d'attention à ce jour, sans doute du fait de sa complexité.

Une autre catégorie de méthodes inverses proposée pour l'estimation des propriétés acoustiques d'une salle à partir de mesures du champ sonore reposent sur la résolution par éléments finis ou différences finies de l'équation des ondes. Parmi ces contributions peuvent être citées celles de [Nava et al., 2009](#), [Antonello et al., 2015](#), [Bertin et al., 2016](#), [Okawa et al., 2021](#) ou encore [Schmid et al., 2023](#). Une limite bien connue de ces approches ondulatoires réside dans leur coût computationnel, en terme d'espace mémoire requis et de temps de calcul, augmentant fortement avec la fréquence. En effet, plus la fréquence d'une onde est élevée, plus sa longueur d'onde est faible et donc plus le pas spatial du maillage requis doit être faible. En conséquence, [Bertin et al., 2016](#), [Okawa et al., 2021](#) et [Schmid et al., 2023](#) limitent leurs travaux à des salles à 2 dimensions uniquement. De leur côté, [Nava et al., 2009](#) et [Antonello et al., 2015](#) limitent leurs travaux aux fréquences inférieures à 125 Hz et 450 Hz respectivement, ce qui est insuffisant dès lors que l'on s'intéresse à l'acoustique des salles d'un point de vue réglementaire. Enfin, il est à noter que l'ensemble de ces approches nécessitent une connaissance exacte de la géométrie (positions des sources, des microphones et des murs). Or, dans la pratique, la mesure de ces paramètres géométriques sur le terrain présente des imprécisions. Afin d'apporter une souplesse, [Bertin et al., 2016](#) proposent une méthode basée sur la parcimonie et capable d'estimer conjointement la position de la source et l'impédance des parois, mais toujours dans le cadre 2D.

En parallèle de ces méthodes opérant l'estimation des propriétés d'absorption des parois d'une salle par inversion de la modélisation directe obtenue avec l'équation des ondes, des méthodes

récentes basées sur l'apprentissage automatique avec des réseaux de neurones se sont intéressées à ce problème. Celles-ci cherchent à estimer conjointement l'ensemble des profils d'absorption des parois à partir de RISs synthétiques par entraînement supervisé. Yu et al., 2020 utilisent ce paradigme pour estimer par valeurs croissantes les coefficients d'absorption de l'ensemble des parois d'une salle mais sans les attribuer aux parois. Une tentative d'attribution des coefficients estimés aux parois correspondantes a ensuite été réalisée mais sans aboutir à des résultats satisfaisants. De plus, les auteurs ont considéré des profils d'absorption constants par morceaux sur trois plages de fréquences ($[0 \text{ Hz}, 1000 \text{ Hz}]$, $[1000 \text{ Hz}, 2000 \text{ Hz}]$ et $[2000 \text{ Hz}, 4000 \text{ Hz}]$). Sur chacune de ces plages, un coefficient d'absorption est tiré aléatoirement suivant une loi uniforme dans l'intervalle $[0, 1]$. Il s'agit d'une description très restreinte (3 coefficients par paroi) et peu réaliste des profils fréquentiels d'absorption des matériaux existants. Pour finir, il est à noter que de telles méthodes sont reconnues pour être très sensibles au réalisme des données synthétiques d'entraînement. Cela pose la question de la capacité de généralisation de cette approche à des données réelles, non éprouvée dans l'article.

3.2 Bases de données de réponses impulsionnelles de salle

Un préalable nécessaire à la construction et l'évaluation de méthodes pour la résolution du problème inverse étudié dans cette thèse est de disposer de bases de données de RISs. Dans le cas de méthodes d'apprentissage automatique avec des réseaux de neurones, une quantité très importante de telles données peut être requise. L'application et le développement de méthodes de traitement du signal et d'apprentissage automatique pour la résolution de problèmes en lien avec l'acoustique des salles bénéficient d'un intérêt croissant. En conséquence, il existe désormais différentes bases de données libres d'accès (*open source*) concernant des RISs mesurées ou simulées. Cependant, nous avons besoin que ces bases de données soient annotées. Ceci signifie qu'elles doivent contenir les RISs mais aussi de l'information sur les conditions de leur acquisition qui sont utiles pour leur analyse telles que les profils d'absorption des parois, la géométrie, les réponses des sources et microphones associés. Nous présentons ici les bases de données annotées existantes et vérifions leur possible utilisation à notre cas.

3.2.1 Bases de données mesurées

L'utilisation de bases de données portant sur des RISs réelles offrirait l'avantage d'assurer une meilleure fiabilité des méthodes développées en vue de leur application à des situations concrètes.

Notons cependant que l'acquisition de bases de données mesurées annotées est complexe, il s'agit de procédures longues qui requièrent des moyens matériels, techniques, logiciels et humains importants. Par conséquent, il est difficile de constituer une base de données de RISs mesurées avec des annotations suffisamment précises et exhaustives et avec une large diversité de salles aussi bien

en termes de géométries que de types de matériau. Il existe néanmoins quelques bases de données de RISs annotées en libre accès. Les annotations disponibles dans ces bases de données de RISs répondent aux besoins d'applications ciblées.

[Dokmanić et al., 2011](#) ; [Crocco et al., 2016](#) et [Remaggi et al., 2016](#) publient des bases de données des RISs mesurées à des fins d'estimation de paramètres géométriques associés tels que la forme ou les dimensions de la salle. Malheureusement, ces bases de données comportent uniquement des annotations sur les géométries associées aux RISs et aucune concernant les propriétés acoustiques des matériaux aux parois. Il est en de même pour la base de données proposées par [Fernandez-Grande et al., 2021](#), incluant des RISs mesurées pour des approches de localisation de sources.

Les bases de données des RISs *RWCP3*, *AIR1*, *BUT ReverbDB2* et ([Nakamura et al., 2000](#) ; [Jeub et al., 2009](#) ; [Szöke et al., 2019](#)) sont destinées à des application relatives au rehaussement ou à la reconnaissance de la parole. Ces bases de données ne comportent pas d'annotations relatives aux géométries associées aux RISs mesurées. Il en est de même pour le corpus de RISs mesurées du *ACE challenge* ([Eaton et al., 2015](#)) annotées uniquement avec le temps de réverbération de la salle et le critère acoustique *DRR* (*Direct to Reverberant Ratio*), défini comme le rapport entre le champ direct et le champ réverbéré.

[Hadad et al., 2014](#) proposent une base de données composée de 1800 RISs mesurées dans une salle aux dimensions fixées mais à acoustique modulable. L'objectif de cette base de données est de fournir des RISs pour différentes positions de sources et de microphones et différentes configurations d'absorption aux parois faisant ainsi varier le temps de réverbération. L'utilisation initiale de cette base de données concerne des applications relatives à la séparation la parole, à son rehaussement ainsi qu'à la localisation de sources. Elle présente des annotations sur la géométrie et le temps de réverbération associés aux RISs mesurées, mais pas sur les profils d'absorption correspondant par exemple. Plus récemment, [Carlo et al., 2021](#) proposent la base de données *dEchorate* reposant sur la salle utilisée par [Hadad et al., 2014](#) et en utilisant un protocole similaire. Elle dispose d'annotations supplémentaires en lien avec les temps d'arrivée des échos dans les RISs. Elle a pour vocation de permettre l'exploitation de l'information fournie par les échos précoces afin d'améliorer les performances de diverses applications en traitement du signal audio.

Malheureusement, il semble que les bases données de RISs mesurées en libre accès ne contiennent pas les informations des profils d'absorption des matériaux des parois ou une annotation qui nous permettrait de pouvoir les retrouver facilement. Il en est de même concernant les mesures des réponses en direction et en fréquence des sources et microphones utilisés, qui serait des informations également utiles.

Afin d'apporter notre contribution aux bases de données disponibles de RISs mesurées et annotées ainsi que pour anticiper sur nos futurs travaux, nous avons participé à une campagne de mesure d'envergure à l'[Ircam](#) en collaboration avec l'équipe *Espaces acoustiques et cognitifs*. Au



(a)



(b)

FIGURE 3.1 – Photographies du dispositif expérimental de mesure installé dans l’Espace de projection de l’Ircam (figure 3.1a) et de la chambre anéchoïque de l’Ircam (figure 3.1b) lors de la campagne de mesure organisée en collaboration avec l’équipe Espaces acoustiques et cognitifs de l’Ircam (Novembre 2022).

cours de cette campagne, nous avons procédé à l’enregistrement synchronisé et calibré d’un grand nombre de RISs pour différentes géométries et configurations acoustiques au sein de l’*Espace de projection* (dit “*Espro*”) de l’Ircam. Il s’agit d’une salle aux propriétés acoustiques et aux dimensions modulables et personnalisables dont les différentes configurations sont connues précisément. Pour l’annotation de la géométrie, une grille mesurée au laser a été mise en place dans la salle afin de positionner précisément les différentes sources et différents microphones. La figure 3.1a présente une photographie du dispositif expérimental de mesure installé dans l’*Espro* lors de cette campagne de mesure. Pour l’annotation des directivités, les réponses en direction et en fréquence des sources et des microphones utilisés sont en cours de mesure et de traitement par l’équipe *Espaces acoustiques et cognitifs* au sein de la chambre anéchoïque de l’Ircam qui dispose de l’équipement adéquat. La figure 3.1b présente une photographie de la chambre anéchoïque de l’Ircam. La possibilité de collecter des données dans cette salle unique n’est malheureusement arrivée que tard au cours de la thèse, et l’acquisition complète des propriétés des sources, récepteurs et parois mis en oeuvre n’a pas été possible pour des raisons techniques. Ces données n’ont donc pas pu être exploitées pour ces travaux de thèse.

De ce fait, nous avons décidé dans le cadre de ses travaux de thèse de considérer des bases données annotées de RISs simulées ; la génération d’une base de données de RISs mesurées aurait été un travail trop important à réaliser durant ce doctorat. Par souci de se rapprocher de la vérité terrain, nous nous sommes intéressés aux simulateurs d’acoustique de salle offrant le plus de réalisme

possible tant en étant performant.

3.2.2 Bases de données simulées

Face à la difficulté de constituer des bases de données annotées de RISs mesurées, des efforts sont alloués au développement de logiciel d’acoustique des salles capables de modéliser des RISs simulées proches de RISs réelles. Pour nos travaux, l’objectif est de pouvoir utiliser un logiciel d’acoustique permettant de construire des bases de données des RISs synthétiques, annotées ainsi que suffisamment volumineuses, diversifiées et réalistes pour être exploitées notamment par nos approches basés sur des réseaux de neurones.

Ces logiciels d’acoustique prennent en entrée les paramètres de la géométrie, les propriétés acoustiques des murs et parfois des réponses des sources et des microphones. Ils reposent aussi sur un modèle d’acoustique des salles (Savioja et al., 2019; Prinn, 2023; Webb et al., 2011) (tir de rayons ou de particules, modèles des sources-images, BEM, FEM, FDTD...) qui offre plus ou moins de réalisme suivant la complexité de celui-ci. Nous distinguons deux grandes catégories de modèles acoustiques : les modèles ondulatoires cherchant à résoudre directement l’équation des ondes, et les méthodes géométriques décomposant le champ sonore en rayon porteur d’une énergie sonore à l’instar des rayons lumineux réfléchis par un miroir. Ces dernières peuvent être divisées en deux sous-groupes (figure 3.2) : les méthodes déterministes (le modèle des sources-images, la radiosité...) et les méthodes stochastiques (tir de rayons ou de particules sonores)

Les méthodes ondulatoires modélisent le champ sonore de la salle en résolvant le système d’équations basé sur l’équation des ondes (2.3) et sur une condition aux limites (condition de Robin). Cette résolution peut nécessiter de mailler spatialement le volume de la salle et ses frontières comme cela est le cas par exemple avec la méthode de résolution par éléments finis (FEM pour *Finite Element Method*) ou avec la méthodes des différences finies (FDTD pour *Finite Difference Time Domain*). Or la taille des mailles doit nécessairement être plus petite que la longueur d’onde étudiée $\lambda = c/f$. Ceci explique pourquoi les méthodes ondulatoires sont dites basses fréquences, leur coût computationnel serait vite prohibitif si l’on devait s’intéresser aux hautes fréquences. Ce coût serait encore accentué pour le cas des salles de grandes dimensions. Malgré cela, ces méthodes offre l’avantage de modéliser la pression complexe, renseignant à la fois sur l’amplitude et sur la phase du signal de pression. Dans la lignée des contributions de Röber et al., 2006 et Raghuvanshi et al., 2009, Hamilton, 2021 propose récemment une implémentation de simulateur d’acoustique de salle 3D par FDTD avec une exécution accélérée sur cartes graphiques (GPUs). Plus largement, Svensson et al., 2002 et Siltanen et al., 2010 mentionnent différents logiciels d’acoustique de salle basés sur des méthodes ondulatoires.

Les méthodes géométriques (déterministes et stochastiques) reposent sur une modélisation qui

considère la propagation de l'énergie sonore (information phase perdue) suivant des *rayons* (trajectoires rectilignes) réfléchis par les parois de la salle à l'instar de rayons lumineux sur des miroirs. Elles se décomposent en deux sous-catégories : les méthodes *déterministes* et les méthodes *stochastiques*. La référence en matière de méthodes déterministes est la méthode des sources-images (Allen et al., 1979), abordée en section 2.3. Pour rappel, elle repose sur la construction déterministe des trajets spéculaires (angle de réflexion égal à l'angle d'incidence) du son dans la salle. Cette méthode est largement utilisée pour la modélisation de la partie précoce des RISs en raison du faible nombre des sources-images, et pour laquelle il est crucial de modéliser l'amplitude et la phase de ces premiers échos si l'on veut rendre compte au mieux de l'acoustique de la salle. Cependant, étant donné que les seules bases de données d'absorption des matériaux à disposition sont classiquement celles issues de mesures en salle réverbérante (section 3.1.1), seuls les coefficients d'absorption de Sabine par bande d'octave issus de la pente de la décroissance énergétique (courbe de Schroeder) sont à notre disposition. De ce fait, les approches implémentées et basées sur le modèle des sources-images reposent souvent, non pas sur la pression acoustique, grandeur complexe possédant une amplitude et une phase, mais sur l'énergie acoustique, grandeur réelle correspondant au module au carré de la pression acoustique. Afin de pouvoir modéliser une RIS (signal de pression), il est alors nécessaire de reconstruire une phase par exemple, *via* une approche de minimum phase (Scheibler et al., 2018).

Pour la partie tardive des RISs, la méthode des sources-images devient inadéquate en raison du très grand nombre de sources-images nécessaires, proportionnel à t^3 , le coût computationnel augmentant alors considérablement. De toute façon, l'hypothèse de réflexions spéculaires inhérente à la méthode des sources-images tend à s'affaiblir lorsque l'ordre des réflexions augmente (Kuttruff, 1995).

En parallèle, les méthodes stochastiques sont reconnues pour la simulation de la partie tardive de la réponse avec un accent porté sur la cohérence et la validité de sa décroissance énergétique au cours du temps. Ce sont des méthodes énergétiques, n'intégrant pas la phase des ondes. Ceci revient à considérer que, en moyenne, la phase des ondes est proche de zéro sur la partie tardive des RISs. Ces méthodes stochastiques reposent sur un échantillonnage des composantes énergie-temps-fréquence des échos qui constituent le champ réverbéré (diffus) suivant une distribution de probabilité tirée de modèles statistiques (par exemple Badeau, 2019) ou construites en considérant un modèle de réflexions diffuses (voir section 2.3) des rayons sonores (Heinz, 1993; Schröder et al., 2007). Elles intègrent des méthodes de *Monte Carlo* qui permettent une simulation du champ diffus en adéquation avec la physique pour un coût computationnel raisonnable.

Savioja et al., 2015 et Brinkmann et al., 2021 présentent une revue approfondie des méthodes géométriques pour la simulation d'acoustique de salle. Leur limite principale sont qu'elles ne permettent de rendre compte de la nature ondulatoire du son (amplitude et phase), elles sont aussi incapables de modéliser de manière fiable les phénomènes basses fréquences tels que la diffraction ou les phénomènes de résonances (Bork, 2000; Brinkmann et al., 2019; Cox et al., 2016). Pour plus

de précision, le lecteur peut se référer à l'article de [Siltanen et al., 2010](#) présentant une analyse approfondie des avantages et inconvénients de ces différentes approches de simulation d'acoustique de salle.

La figure [3.2](#) illustre les zones de prédilection pour l'utilisation des différentes méthodes mentionnées dans le plan temps-fréquence ([Thomas, 2017](#); [Badeau, 2019](#)). Les frontières en temps (*temps de mélange* [Defrance et al., 2010](#)) et en fréquence (fréquence de Schroeder [Skålevik, 2011](#)) marquent respectivement une densification des réflexions et des *modes de résonance* (ondes stationnaires pour certaines fréquences particulières de la salle). Ces frontières suggèrent un changement de méthode de simulation pour un meilleur compromis entre réalisme et coût computationnel.

Ainsi, afin d'obtenir le meilleur compromis entre réalisme des simulations et coût computationnel, ces différentes méthodes sont souvent combinées entre elles pour tirer partie de leurs avantages respectifs et créer un logiciel de simulation hybride. Typiquement, les composantes basses fréquences et hautes fréquences des RISs peuvent être simulées respectivement avec une méthode ondulatoire et une méthode géométrique de l'autre. Aussi, les parties précoces et tardives des RISs peuvent être générées respectivement à l'aide d'une méthode géométrique déterministe et d'une autre méthode stochastique. Citons les logiciels tels que RAVEN ou encore Ramsete ([Schröder et al., 2011](#); [Farina et al., 2007](#)) qui permettent d'hybrider plusieurs méthodes entre elles afin de simuler efficacement des RISs réalistes en modélisant certains phénomènes ondulatoires comme la diffraction. [Tang et al., 2022](#) présentent *GWA (Geometric-Wave Acoustic dataset)*, une base de données annotées de RISs synthétiques hautement réalistes, simulées en utilisant des méthodes ondulatoires pour les basses fréquences et géométriques pour les hautes fréquences. Récemment, [Chen et al., 2022](#) introduisent *Soundscapes 2.0*, une plateforme permettant la simulation en temps réel d'environnements acoustiques et visuels hautement réalistes. Ils dévoilent également *SoundSpaces-PanoIR*, une vaste base de données annotées de RISs accompagnées d'images panoramiques correspondantes, faisant suite à la base de données *Soundscapes* [Chen et al., 2020](#).

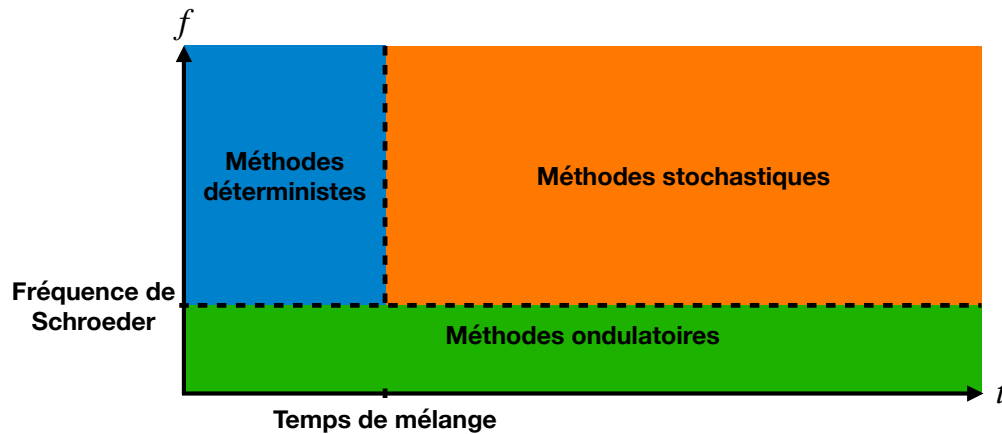


FIGURE 3.2 – Schéma illustrant dans le plan temps-fréquence les régions privilégiées pour la simulation des composantes correspondantes des RISs à l'aide des méthodes ondulatoires, déterministes et stochastiques (Thomas, 2017 ; Badeau, 2019).

3.2.3 Bibliothèques disponibles pour la simulation en MATLAB et Python

Nous résumons ici les bibliothèques disponibles en libre accès pour la simulation de RISs réalistes écrites en Matlab et Python, deux principaux langages de programmation scientifique utilisés dans les domaines de l'éducation, de la recherche et de l'industrie.

En Matlab, les implémentations de la méthode des sources-images pour des salles en forme de pavé droit développées par Campbell et al., 2005 et par Habets, 2006 font office de référence dans la communauté des acousticiens. En l'occurrence, Campbell et al., 2005 mettent à disposition un simulateur, nommé *Roomsim*, qui fait l'objet en permanence d'améliorations. Schimmel et al., 2009 ont proposé une version hybride de *Roomsim* modélisant les réflexions précoces d'une RIS par la méthode des sources-images et la partie tardive (champ réverbéré) par des méthodes géométriques stochastiques (Heinz, 1993 ; Lentz et al., 2007). Par suite, Wabnitz et al., 2010 ont fait évoluer *Roomsim* en une version nommée *MCRoomsim*, capable d'intégrer la directivité des antennes de microphones intégré dans *Roomsim*, ainsi que leurs structures à partir du formalisme des harmoniques sphériques. Sous python, Barumerli et al., 2021 implémentent *SofaMyRoom*, un module adapté du simulateur *Roomsim* de Schimmel et al., 2009 avec des fonctionnalités supplémentaires relatives à la simulation binaurale avec modélisation de la fonction de transfert relative à la tête caractérisant la déformation spectrale du son par la tête (visage et pavillon de l'oreille).

En raison de son plébiscite par les communautés rassemblées autour du domaine de l'intelligence artificielle, Python a connu une expansion concomitante à l'essor de ce domaine au cours des dernières années, devenant incontestablement un langage de programmation majeur dans la recherche et l'industrie. Par conséquent, un intérêt et une demande croissante sont observés pour le développement de bibliothèques Python pour la simulation de RISs réalistes. Scheibler et al., 2018

présentent le module *Pyroomacoustics* faisant office de référence pour la simulation de RISs en Python. De façon similaire à *Roomsim*, *Pyroomacoustics* s'appuie sur la méthodes des sources-images et les méthodes de tir de rayons. L'outil *Pyroomacoustics* offre l'avantage d'être modulaire et fonctionnel ainsi que d'être capable de simuler des RISs pour le cas de salles de forme polyédrique. En complément, dans la lignée des travaux contribution de [Fu et al., 2016](#), il est intéressant de citer la contribution [Diaz-Guerra et al., 2021](#), présentant récemment *gpuRIR*, une librairie Python permettant de paralléliser les calculs de la méthode des sources-images sur des GPUs pour une simulation accélérée de RISs.

Enfin, il est très récemment apparu de nouveaux simulateurs basés sur l'utilisation de *réseaux de neurones génératifs*. [Ratnarajah et al., 2020](#) présentent *IR-GAN* dont l'objectif est d'améliorer le réalisme des RISs simulées en extrayant tout d'abord de l'information de RISs mesurées puis, en utilisant celle-ci afin de générer des RISs simulées plus réalistes. [Ratnarajah et al., 2022](#) et [Luo et al., 2022](#) présentent respectivement *FAST-RIR* et *FRA-RIR* dans le but d'accélérer significativement le temps de simulation des méthodes géométriques de référence en apprenant à estimer directement les RISs à partir des paramètres d'entrée de ces méthodes. [Masztalski et al., 2020](#) proposent *StoRIR*, un générateur stochastique de RISs qui répond aux mêmes objectifs et aux mêmes constats.

Cependant, ces simulateurs sont évalués soit sur leur capacité à produire des RISs simulées en accord avec des indicateurs acoustiques globaux tels que le temps de réverbération ou le rapport champ direct sur champ réverbéré (DRR), soit sur leur capacité à améliorer les performances de modèles entraînés pour des applications en reconnaissance ou rehaussement de la parole. Ils ne visent pas à estimer précisément les amplitudes et temps d'arrivée des échos y compris ceux précoces, ce qui est primordiale pour le problème étudié dans cette thèse. Ils n'ont pas non plus vocation à rendre compte précisément des réponses des matériaux de la salle ou de celles des sources et des microphones utilisés.

Compte tenu de l'évolution de ces outils au cours de ces dernières années, et du fait que notre objectif est de pouvoir obtenir des RISs certes simulées, mais avec le plus de réalisme possible, nous avons finalement décidé de retenir les librairies (*Roomsim* et *Pyroomacoustics*) dans le but de construire nos bases de données annotées de RISs.

3.2.4 Les dernières avancées en matière de réalisme des simulations

Il existe différents travaux en récents, visant de renforcer le réalisme des RISs simulées. Certains de ces travaux ont pour objectif de pouvoir tenir compte de mesures réelles de réponses des appareils utilisés (sources et microphones) et des réponses des matériaux mise en oeuvre dans les salles d'étude (notamment les RIMs $R(t, \theta)$ ou $\underline{R}(f, \theta)$).

En l'occurrence, [Dokmanić et al., 2013](#) ont montré que la taille de la source sonore ainsi que

sa réponse ont une forte influence sur les échos au sein des RISs et *in fine* sur l'estimation de la géométrie de la salle à partir de ces RISs. Les auteurs pointent les limites des hypothèses de sources ponctuelles ou sphériques qui ne sont pas toujours viables en pratique. Allant dans ce sens, [Gómez Bolaños et al., 2014](#) proposent une procédure de mesure de RISs reposant sur la production d'une faible explosion obtenue en focalisant des lasers en un point de l'espace dans la salle, *LIB* (*Laser-Induced air Breakdown*). Ils démontrent l'aspect ponctuel et omnidirectionnel de la source ainsi produite, pouvant alors être assimilée à une impulsion, ainsi que son absence d'influence sur la RIS mesurée. Comparativement, ils soulignent alors l'impact que peut avoir une source omnidirectionnelle de taille non négligeable sur les RIS mesurées ainsi que la difficulté à compenser sa réponse par déconvolution, notamment en raison de l'imperfection de son caractère omnidirectionnel. Enfin, [Ribeiro et al., 2011](#) et [Thydal et al., 2021](#) évoquent respectivement l'importance de disposer précisément des réponses des sources et microphones pour l'estimation de la géométrie à partir de RISs ainsi que pour la simulation d'acoustique de salle.

Afin de considérer cet aspect, [Srivastava et al., 2022](#) ont implémenté récemment une extension du simulateur Pyroomacoustics permettant d'importer des réponses mesurées de sources et de microphones, dépendantes de la direction et de la fréquence, afin d'améliorer le réalisme des RISs simulées. Ils ont également introduit une contrainte de *phase minimale* (filtres causaux, stables et avec un temps de décroissance minimal) pour la réponse des parois produites par le simulateur afin de rendre compte de la très courte durée de leurs oscillations engendrées par le son. Avec cette extension, les auteurs ont pu démontrer que l'utilisation des réponses de sources et de microphones mesurées par [Brandner et al., 2018](#) pour la simulation permettait d'améliorer significativement la généralisation de réseaux de neurones entraînés sur des données simulées à des tests sur données réelles. La tâche visée portait sur l'estimation aveugle du volume, de la surface totale et du temps de réverbération en bandes d'octave d'une salle à partir d'enregistrements vocaux réverbérés. Bien que la mise à disposition en libre accès de cette nouvelle fonctionnalité soit réservée à une version future du simulateur Pyroomacoustics, nous avons eu l'opportunité de pouvoir en bénéficier pour nos travaux. Ceci a contribué à augmenter le réalisme de nos bases données de RISs simulées et annotées ainsi qu'à nous rapprocher de l'objectif de pouvoir, à terme, appliquer nos approches au cas de RISs réelles.

Une autre avancée concernent les travaux de [Mondet et al., 2020](#) qui proposent une approche d'optimisation sous contraintes pour la résolution du problème inverse d'estimation de l'impédance complexe d'un matériau à partir de ses coefficients d'absorption en bandes d'octave. Cette approche permet d'utiliser de l'information a priori sur la paroi, typiquement sa catégorie parmi celles de référence en acoustique des matériaux, afin d'obtenir une solution en adéquation avec les modèles traditionnels dans le domaine. D'une part, cette contribution ouvre la voie pour de nouvelles améliorations en matière de précision de description des propriétés acoustiques des parois et de réalisme en modélisation numérique et en simulation acoustique. Une telle avancée peut présenter des bénéfices pour un large éventail d'applications en traitement du signal audio. D'autre part, elle

pourrait être associée aux approches développées dans cette thèse pour étendre les résultats d’estimation des profils d’absorption des parois aux impédances complexes. Cette perspective renforce l’intérêt pour le problème inverse étudié dans cette thèse et permet de l’envisager comme une étape de la caractérisation exhaustive de l’acoustique d’une salle à partir de signaux audio.

Pour compléter, nous spécifions ici les travaux de [Rust et al., 2021](#) présentant une vaste base de données de mesures de réponses impulsionnelles matériaux (correspondant à des RIMs $R(t, \theta)$). Ces matériaux sont synthétisés en laboratoire par impression 3D et en visant des structures de matériaux réellement existants en acoustique des salles. Ces RIMs sont mesurées avec une méthode en conditions de champ libre (3.1.2) pour plusieurs positions de source et de microphone à l’aide de bras robotiques automatisés permettant ainsi de varier systématiquement l’angle de réflexion (incidence oblique). Cette base de données comportent environ un million de réponses mesurées pour 350 matériaux de référence. Elle est initialement destinée à des applications d’apprentissage automatique pour la prédiction d’indicateurs sur la diffusivité acoustique des matériaux à partir de données sur leurs structures ou bien de partitionnement des matériaux en fonction de l’ensemble de ces variables. Cette base de données est finalement publiée en tant que *GIR Dataset* par [Xydis et al., 2023](#), accompagnée de scripts facilitant l’utilisation et la visualisation des données. Elle constitue une précieuse opportunité pour l’amélioration du réalisme des simulations d’acoustique de salle. Elle donne accès à une très grande variété de matériaux avec une description riche de leurs propriétés acoustiques incluant la phase et la dépendance angulaire. Enfin, [Rust et al., 2021](#) présentent en annexe un exemple d’une RIM contenue dans la base de données GIR. Il est possible de noter que cette réponse est de très courte durée (moins de 1 ms). Cette observation est en accord avec la modélisation RIMs en tant que filtres à phase minimale, communément employée et considérée dans cette thèse.

3.3 Conclusion

Cet état de l’art a permis d’aboutir aux constatations suivantes. Dans un premier temps, il semble que la caractérisation acoustique de toutes les parois d’une salle, reste encore aujourd’hui difficile du fait de plusieurs verrous scientifiques à outrepasser. Les approches de séparation qui nécessitent de pouvoir différencier le champ réfléchi par chacune des parois présentes au sein de mesures du champ sonore semblent difficiles à appliquer dans un tel environnement fermé réverbérant. Plusieurs de ces approches étudiées précédemment nécessiteraient de prélever des échantillons de paroi afin de réaliser de mesures en conditions de champ libre par exemple. Cependant, de telles approches nécessairement “destructive” se trouvent être en général incompatibles avec notre application de diagnostic acoustique de l’existant. En parallèle, ces approches de séparation peuvent être considérées en association avec différentes formulations analytiques du champ acoustique dans le voisinage d’un matériau à caractériser. Malheureusement, il n’est pas connu à ce jour de méthodes permettant d’inverser en général ces formulations mathématiques très complexes. De plus, celles-ci

sont aussi exprimées à partir de l'impédance acoustique (ou de l'admittance), grandeur pour laquelle il n'existe pas de bases de données exploitables pour nos travaux. Face à ce constat, nous avons décidé de chercher à développer une approche qui, à partir de quelques mesures de RISs dans la salle d'étude, serait capable de remonter aux propriétés d'absorption de l'ensemble des matériaux d'une salle. Bien sûr, au regard de ce qui a été dit précédemment, il s'agit d'un challenge difficile à atteindre.

Le deuxième point que cet état de l'art a mis en lumière concerne le manque de données de RISs mesurées, accompagnées des annotations acoustiques et géométriques requises pour la conception et l'évaluation des approches de résolution du problème inverse d'intérêt. Cette constatation nous a incité à nous intéresser à la simulation de bases de données de RISs simulées. Il est alors apparu que les approches d'acoustiques ondulatoires, bien que très précises, sont limitées aux très basses fréquences compte tenu du coût computationnel important qu'elles requièrent pour la simulation haute fréquence. Cependant, pour rappel, la réglementation acoustique des salles et du bâtiment repose sur la gamme de fréquences s'étalant approximativement de 125 Hz à 8000 Hz. De ce fait, les approches acoustiques géométriques nous ont semblé être les plus adéquates pour la génération de nos bases de données. Les logiciels d'acoustique de référence semblent à ce jour reposer sur l'approche acoustique des sources-images. Afin de s'assurer que nos travaux seront, à terme, transposables au cas de RISs réelles, nous avons regardé les travaux existants, ainsi que ceux futurs, allant dans ce sens. En particulier, nous nous sommes intéressés aux logiciels de simulation basés sur cette méthode des sources-images et permettant d'augmenter le réalisme des RISs simulées en y incluant des aspects inhérents au mesurage sur site. Ces derniers concernent principalement les réponses des sources et des microphones utilisés qui impactent considérablement la mesure des RISs et dont il est donc important de tenir compte. Nous avons retenu les logiciels Roomsim et Pyroomacoustics afin de réaliser nos propres bases de données de RISs annotées en adéquation avec nos travaux de recherche, avec un souci constant du réalisme des simulations.

Estimation pour une géométrie spécifique par apprentissage supervisé virtuel

Préambule : Ce chapitre présente une méthode pour estimer les coefficients d'absorption dans six bandes d'octave des parois d'une salle donnée à partir d'une seule réponse impulsionnelle de salle et de la connaissance des paramètres géométriques associés, aux erreurs de mesure près. Cette estimation est réalisée avec un réseau de neurones profond entraîné sur une base de données simulées et annotées de réponses impulsionnelles de salle correspondant à cette même géométrie spécifique (*apprentissage supervisé virtuel*). Les travaux présentés dans ce chapitre ont fait l'objet d'un article (Dilungana et al., 2021) et d'une présentation à la conférence Inter-Noise 2021.

4.1 Introduction

Ce chapitre étudie la faisabilité de l'estimation des profils d'absorption de chacune des parois d'une salle donnée à partir de RISs. L'approche proposée est inspirée de celle développée dans Foy et al., 2021 reposant sur le paradigme d'*apprentissage supervisé virtuel*, appliqué à l'estimation de l'absorption moyenne par bande d'octaves d'une salle à partir de RISs. L'apprentissage supervisé virtuel consiste à entraîner un réseau de neurones à effectuer une tâche d'intérêt à partir d'une vaste base de données synthétiques générée en utilisant un simulateur. En principe, cette approche vise une application à des données réelles et requiert donc de disposer d'un simulateur paramétrique suffisamment réaliste pour fournir des données synthétiques labéllisées similaires à la donnée réelle disponible.

Cette approche présente l'avantage de tirer profit de la puissance reconnue des réseaux de neurones pour l'extraction d'information de données complexes dans un contexte supervisé, et ce même dans une situation où la quantité de données réelles disponibles est insuffisante pour l'entraînement. Cette situation se présente pour le problème considéré en raison de l'indisponibilité de vastes bases de données de RISs mesurées et annotées précisément des géométries et profils d'absorption ainsi que de la difficulté d'en constituer (section 3.2.1). À la différence de Foy et al., 2021 qui s'intéresse à l'estimation de l'absorption moyenne par bande d'octaves dans une salle à partir de RISs, notre étude porte sur l'estimation des profils d'absorption de chacune des parois d'une salle à partir de

RISs. La première nécessite une information géométrique globale tandis que la seconde repose sur connaissance précise de la géométrie (section 2.2). Ce travail adopte donc une approche différente qui s’appuie sur deux idées principales. D’une part, l’information des propriétés acoustiques des parois individuelles d’une salle est portée par les premiers échos des RISs (section 2.2). D’autre part, ce sont les positions relatives des parois, de la source et du récepteur qui déterminent les temps d’arrivée de ces premiers échos d’ordre faible dans les RISs (section 2.3). Pour deux RISs mesurées dans une même salle mais avec des positions distinctes de source et de microphone, l’ordre et la répartition des premiers échos peuvent être très différents. Il semble donc essentiel de tenir compte de la géométrie pour parvenir à associer les échos aux parois correspondantes pour l’estimation.

Dans le contexte de l’apprentissage supervisé avec des réseaux de neurones, il est non trivial de savoir comment renseigner les paramètres géométriques associés à chaque RIS en entrée du modèle. Par exemple, la concaténation simple des RISs et des paramètres géométriques semble être insuffisante compte tenu de la complexité de la relation qui lie la géométrie aux temps d’arrivée des échos dans la RIS (section 2.3). Cela est également incompatible avec l’emploi d’un réseau neuronal convolutif pour traiter l’ensemble de ces entrées “hétérogènes”. Idem pour toute tentative de renseigner uniquement les temps d’arrivée des échos, ce qui omet tout lien entre les échos et les parois. De plus, cela pose la question du nombre d’échos à renseigner pour l’estimation ou encore de l’invariance par permutation de la labéllisation des échos d’une RIS à la suivante.

Devant ce constat, la première approche proposée consiste à estimer les paramètres d’intérêt à partir d’une RIS donnée en utilisant un réseau de neurones entraîné à accomplir cette tâche sur un jeu de RISs simulées spécifiquement pour la même géométrie, mais avec des profils d’absorption diverses et réalistes. Cette approche revient à spécialiser le réseau de neurones à inverser le problème d’intérêt pour une géométrie spécifique. La géométrie est fixée sur la base de données d’apprentissage afin que le modèle puisse l’apprendre, c’est-à-dire qu’il puisse progressivement identifier les régions du signal utiles à l’estimation des profils d’absorption de chacune des parois pour cette géométrie.

Ce chapitre constitue une preuve de concept visant à établir la possibilité de cette estimation avec approche par apprentissage bout en bout n’incluant pas de modélisation physique, spécifique à une géométrie, et dans un cadre simulé considérant différents niveaux de réalisme pour l’entraînement et le test. L’objectif de cette étude est également d’identifier les paramètres de simulation qui impactent significativement l’estimation et d’obtenir des indices sur les composantes du signal qui portent l’information nécessaire à l’estimation. Plusieurs études empiriques approfondies sont menées sur des données simulées à l’aide du simulateur acoustique ROOMSIM (Schimmel et al., 2009). Celles-ci portent sur l’estimation des profils d’absorption des parois d’une salle en forme de pavé droit, décrits dans six bandes d’octave ($b \in \{0, .125, .25, .5, 1, 2, 4\}$ kHz), à partir d’une seule RIS et de sa géométrie connue aux erreurs de mesure près de l’ordre de quelques centimètres. Deux architectures classiques de réseaux de neurones sont considérées : le perceptron multicouche et le réseau neuronal convolutif. Plusieurs expériences préliminaires sont réalisées pour

évaluer et comparer les performances des modèles. La robustesse de ces modèles aux fluctuations de trois paramètres affectant le champ sonore est soigneusement étudiée, à savoir la présence d’erreurs géométriques, de bruit de mesure et de diffusion acoustique. Enfin, une analyse approfondie est menée pour qualifier l’importance des différentes composantes de la RIS pour l’estimation en comparant les performances de modèles entraînés respectivement à partir des parties précoces et tardives des RISs pour différents instants de coupures.

Le chapitre est organisé comme suit. La Section 4.2 détaille le processus de simulation et de traitement des jeux de RISs. La Section 4.3 décrit les architectures utilisées. La Section 4.4 présente les expériences menées et les résultats obtenus. Enfin, la Section 4.5 apporte une conclusion et met en avant quelques pistes d’amélioration possibles.

4.2 Simulation des bases de données de réponses impulsionnelles de salle

4.2.1 Paramétrage du simulateur

L’entraînement supervisé de réseaux de neurones profonds requiert une grande quantité de données annotées. En pratique, il est très difficile de mesurer des ensembles de données de RISs réelles suffisamment vastes, diversifiées et annotées par les profils d’absorption et par les géométries. Nous suivons donc une approche similaire à celle présentée dans Foy et al., 2021 et générons des bases de données de RISs annotées à l’aide du simulateur d’acoustique de salle en forme de pavé droit ROOMSIM (Schimmel et al., 2009). Il combine la méthode des sources-images pour modéliser les réflexions spéculaires et l’algorithme “diffuse-rain” (Heinz, 1993; Schröder et al., 2007) pour modéliser les effets aléatoires de diffusion acoustique.

Les bases de données considérées pour cette étude se composent de RISs simulées de durée $T = 500$ ms, échantillonnées à la fréquence $\bar{f} = 16$ kHz. Des réponses de sources et des microphones idéales, c’est-à-dire indépendantes de la direction et de la fréquence, sont considérées pour les simulations.

Chacune des bases de données simulées correspond soit à l’une des géométries fixes G1 ou G2 décrites dans le tableau 4.1 et représentées par la figure 4.1, soit à des géométries différentes tirées aléatoirement pour chaque RIS. Pour la seconde situation (désignée par \mathcal{G}), les dimensions des salles (en mètres) sont tirées uniformément dans $[3, 10] \times [3, 10] \times [2, 5]$ et les positions de la source et du microphone sont tirées uniformément dans la salle en respectant une distance de 1 mètre aux murs de la salle, ainsi qu’entre chaque source et au moins 1 microphone. Cette contrainte permet d’éviter les situations d’incidence rasantes des réflexions souvent mal modélisées par les logiciels d’acoustique existants. Pour certains des jeux de RISs simulés, l’erreur de mesure de la géométrie est modélisée en appliquant une perturbation de loi normale $\mathcal{N}(0, \sigma^2)$ où $\sigma = 0.02$ m aux positions

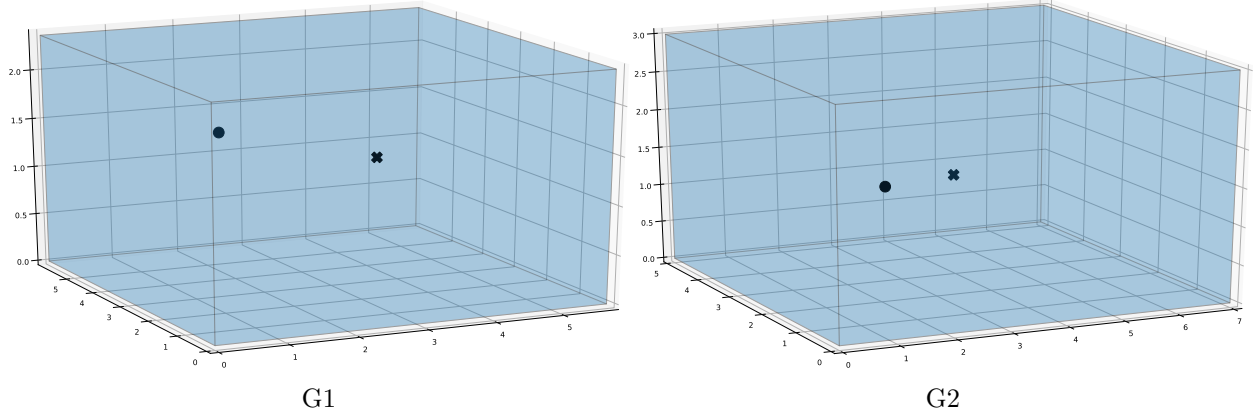


FIGURE 4.1 – Visualisation 3D des géométries G1 et G2 (détaillées dans le tableau 4.1). La source est représentée par une sphère et le microphone est représenté par une croix.

des parois, et à celles de la source et du récepteur avant la simulation de chaque RIS (voir tableau 4.3).

	Dimensions de la salle	Position de la source	Position du microphone
G1	[5.70, 5.97, 2.36]	[1.99, 4.50, 1.42]	[3.48, 2.57, 1.31]
G2	[7, 5, 3]	[1.7, 1.4, 1.6]	[3.9, 2.9, 1.3]

TABLEAU 4.1 – Géométries fixes G1 et G2 utilisées pour les expériences (en mètres).

Les profils d’absorption et de diffusion (voir 2.3) des parois sont définis par 6 coefficients (compris entre 0 et 1) associés à 6 bandes d’octaves centrées sur les fréquences $b \in \{.125, .25, .5, 1, 2, 4\}$ kHz. Ces bandes d’octaves sont celles couramment utilisées dans les réglementations en acoustique du bâtiment. Afin d’obtenir des bases de données diversifiées tout en préservant le réalisme acoustique, les profils d’absorption sont échantillonnés en utilisant une stratégie analogue à celle proposée dans Foy et al., 2021 favorisant les profils réfléchissant, comme résumé ci-après. Pour chaque salle simulée, le type de chaque paroi est tiré aléatoirement parmi les types “mur”, “sol”, “plafond” et “réfléchissant” suivant les probabilités $p_{\text{mur}} = \frac{1}{3}$, $p_{\text{sol}} = \frac{1}{12}$, $p_{\text{plafond}} = \frac{1}{12}$ et $p_{\text{refl}} = \frac{1}{2}$. En fonction du type attribué à chaque paroi, les coefficients d’absorption en bande d’octaves sont échantillonnées uniformément à l’intérieur des plages définies dans le tableau 4.2. Ces plages sont fondées sur des mesures en laboratoire de profils d’absorption de matériaux couramment rencontrés dans les bâtiments, compilées dans Foy et al., 2021. Dans chaque salle, un profil de diffusion unique décrit dans les six mêmes bandes d’octave est tiré pour toutes les parois uniformément à l’intérieur des plages définies dans la dernière colonne du tableau 4.2.

Bande d'octave	Abs. mur	Abs. sol	Abs. plafond	Abs. réfléchissant	Diff. (toutes)
125 Hz	[0.01, 0.50]	[0.01, 0.20]	[0.01, 0.70]	[0.01, 0.12]	[0.00, 0.30]
250 Hz	[0.01, 0.50]	[0.01, 0.30]	[0.15, 1.00]		[0.00, 0.30]
500 Hz	[0.01, 0.30]	[0.05, 0.50]	[0.40, 1.00]		[0.00, 0.30]
1 kHz	[0.01, 0.12]	[0.15, 0.60]	[0.40, 1.00]		[0.20, 1.00]
2 kHz	[0.01, 0.12]	[0.25, 0.75]	[0.40, 1.00]		[0.20, 1.00]
4 kHz	[0.01, 0.12]	[0.30, 0.80]	[0.40, 1.00]		[0.20, 1.00]

TABLEAU 4.2 – Plages des coefficients d’absorption (abs) et de diffusion (diff) par type de paroi et par bande d’octave.

4.2.2 Prétraitement des données

Par soucis de réalisme et afin de supprimer les artefacts à basse fréquence causés par le simulateur, un filtre ”passe-haut“ de fréquence de coupure $f_c = 50\text{Hz}$ est appliqué à chaque RIS simulée. En effet, les sources et microphones conventionnels n’opèrent pas à des fréquences aussi basses en général. Pour certains des jeux de RISs simulés, le bruit de mesure est modélisé en ajoutant un bruit blanc gaussien additif à chaque RIS de sorte à atteindre un rapport signal à bruit maximal (PSNR pour *Peak Signal to Noise Ratio* en anglais) échantillonné uniformément dans l’intervalle [40, 50] dB. Enfin, chaque RIS est mise à l’échelle par normalisation par son maximum c’est-à-dire, $\text{RIS}_{\text{normalisé}} = \text{RIS} / \max(\text{RIS})$.

4.2.3 Bases de données simulés

Huit bases de données annotées contenant chacun 15 000 RISs sont simulées. Les noms, les géométries et les niveaux de modélisation associés à chacune de ces bases de données de RISs sont présentées dans le tableau 4.3. Les bases de données correspondant à la géométrie G1 sont utilisées pour évaluer et comparer l’impact sur l’estimation des différents types d’aléas considérés.

La base de données G2-SDEB est utilisée pour comparer les performances de modèles entraînés et testés respectivement sur deux géométries différentes (avec G1-SDEB). Ces deux bases de données sont simulées avec le plus haut niveau de réalisme considéré dans cette étude mais pour deux géométries fixes distinctes. Pour finir, la base de données \mathcal{G} -SDEB, les géométries associées aux RISs sont tirées aléatoirement à chaque observation (4.2.1). Cette dernière est utilisée pour évaluer les performances d’un modèle entraîné et testé sur une base de données composée de RISs dont les géométries sont toutes différentes.

Base de données	Géométrie	Modélisation
G1-S	G1	réflexions spéculaires (S)
G1-SD	G1	réflexions spéculaires (S) + réflexions diffuses (D)
G1-SE	G1	réflexions spéculaires (S) + erreurs géométriques (E)
G1-SB	G1	réflexions spéculaires (S) + bruit blanc gaussien additif (B)
G1-SDE	G1	S + D + E
G1-SDEB	G1	S + D + E + B
G2-SDEB	G2	S + D + E + B
\mathcal{G} -SDEB	\mathcal{G} (aléatoires)	S + D + E + B

TABLEAU 4.3 – Description des huit des bases de données simulées utilisés pour les expériences.

4.3 Architectures et entraînement

4.3.1 Structure des modèles

Le premier réseau de neurones considéré dans cette étude est un perceptron multicouche (MLP pour “*Multilayer Perceptron*”) comportant 3 couches cachées de respectivement 8000×256 , 256×128 et 128×64 neurones. Chaque couche cachée est suivie d’une couche de normalisation de type “Batch Normalization” (normalisation par lot de données) et d’une fonction d’activation de type “Exponential Linear Unit” (s). Le second réseau de neurones considéré est un réseau neuronal convolutif (CNN pour “*Convolutional Neural Network*”) comportant 4 couches cachées consécutives de convolution 1D respectivement composés de 16, 32, 64 et 128 filtres de tailles 33, 17, 9 et 5, avec un décalage d’un pas (“stride”). Chaque couche de convolution est suivie d’une couche de type “Max Pooling” de largeur 4, d’une couche de normalisation de type “Batch Normalization” et d’une fonction d’activation de type “Rectified Linear Unit”. La sortie résultante de dimension 3968 est ensuite passée à travers une couche de perceptron à 64 neurones suivie par des ELUs. La couche de sortie de chaque réseau de neurones est constituée par l’association d’une couche de perceptron et d’une fonction sigmoïde pour produire le vecteur de sortie désiré. La fonction de coût utilisée pour le calcul de l’erreur et la mise à jour des poids est l’erreur quadratique moyenne (EQM). Les modèles sont optimisés par lots de 50 RISs (batches). L’optimiseur ADAM (Kingma et al., 2014) est utilisé avec un taux d’apprentissage égal à 0.01. La sortie du modèle est un vecteur de dimension 36 qui correspond aux estimations des coefficients d’absorption dans les 6 bandes d’octaves, pour les 6 parois de la salle en forme de pavé droit.

4.3.2 Méthodologie d’entraînement

Chaque bases de données de RISs simulées (voir tableau 4.3) est divisé en un jeu d’entraînement, un jeu de validation et un jeu de test correspondant respectivement à 70% (10500 observations),

15% (2250 observations) et 15% (2250 observations) des observations totales au sein de la base de données considérée. Pour les jeux de test présentant du bruit blanc gaussien, de l'*augmentation de données* est effectuée dans le but améliorer la robustesse du modèle au bruit de mesure. Ce procédé intervient en phase d'apprentissage et consiste à générer et appliquer le bruit blanc gaussien au niveau des batchs de RISs à chaque époque (*epoch* en anglais). Les paramètres du modèle d'apprentissage considéré produisant la plus faible erreur quadratique moyenne (EQM) sur le jeu de validation au cours de 300 époques sont conservés et considérés pour le jeu de test. Un arrêt anticipé de l'entraînement (*early-stopping* en anglais) est effectué lorsque les performances du modèle ne présentent plus d'évolution significatives durant 50 époques consécutives, en garantissant un minimum de 100 époques.

4.4 Expériences et résultats

4.4.1 Étude quantitative : performances, robustesse et généralisation des modèles

Une première étude est menée afin d'analyser l'impact des différents niveaux de modélisation sur les performances des modèles MLP et CNN définis dans 4.3.1 en utilisant les bases de données correspondant à la géométrie G1 (voir tableau 4.3). Notons que pour chaque expérience, chaque architecture est entraînée, validée et testée respectivement sur les jeux d'entraînement, de validation et de test issus de la même base de données conformément à 4.3.2, sans redondances entre ces différents ensembles de données. Un même modèle est entraîné et testé sur des données affectées par le même type de distorsions. La figure 4.2 décrit les erreurs absolues moyennes (EAM) d'estimation par les modèles MLP et le CNN des six coefficients d'absorption en bande d'octave des six parois de la salle pour l'ensemble des RISs contenues dans chacun des jeux de test. Aussi, la figure 4.2 présente à titre de référence l'EAM d'un estimateur qui retourne la moyenne au sein du jeu d'entraînement des coefficients d'absorption pour chaque bande d'octave. Cet estimateur de la moyenne est considéré être meilleur qu'une estimation au hasard mais correspond à une situation d'entraînement déplorable pour laquelle le modèle a seulement appris la distribution des annotations du jeu d'entraînement sans parvenir à extraire de l'information des RISs.

Une première analyse générale de la figure 4.2 permet de constater que les modèles MLP et CNN (4.3.1) entraînés *via* l'approche définie dans ce chapitre parviennent à extraire des RISs l'information nécessaire à l'estimation précise des coefficients d'absorption en bande d'octave des parois pour la géométrie G1. En effet, ils surpassent nettement les performances de l'estimateur de la moyenne, ce pour toutes les combinaisons de distorsions considérées, y compris pour la base de données G1-SDEB qui les combine toutes.

De plus, la comparaison des résultats associés aux bases de données G1-S, G1-SD, G1-SE et G1-SB, révèle que la hausse de l'EAM la plus importante est causée par la présence d'erreurs sur les

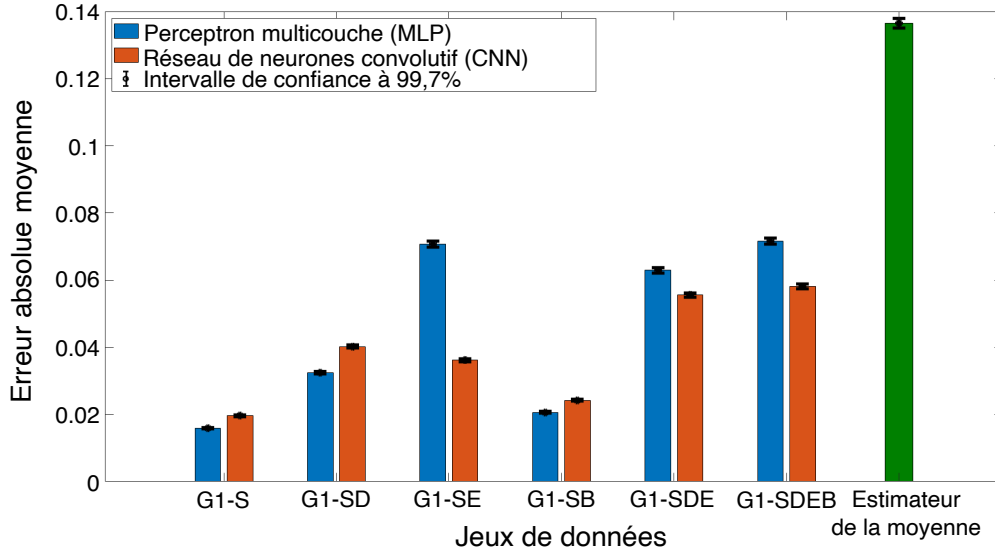


FIGURE 4.2 – Erreur absolue moyenne (EAM) sur l’estimation des coefficients d’absorption en bande d’octave de l’ensemble des parois d’une salle par des modèles de type MLP et CNN entraînés et testés respectivement sur les bases de données simulées associées à la géométrie fixe G1 présentés dans le tableau 4.3

paramètres géométriques. Cette observation confirme l’importance d’une connaissance précise de la géométrie ou bien d’une approche robuste à l’incertitude géométrique pour l’estimation des profils d’absorption des parois individuelles. En conséquence, une attention particulière sera portée à la robustesse aux erreurs géométriques des approches développées dans la suite de cette thèse. En se focalisant sur les résultats obtenus par les deux types d’architectures sur les bases de données présentant des erreurs géométriques (G1-SDE, et G1-SDEB et en particulier G1-SE), il est possible d’observer que les performances du CNN surpassant significativement celles du MLP dans de telles conditions. Ce constat met en évidence la robustesse accrue du modèle CNN face aux erreurs géométriques par rapport au modèle MLP. Il est possible d’intuiter que ce sont les couches de Max Pooling qui aident à localiser de manière plus robuste les premiers échos en sélectionnant successivement la valeur la plus élevée au sein des fenêtres glissantes de convolution.

Enfin, la comparaison des performances sur les bases de données G1-S et G1-SB permet d’observer que le MLP comme le CNN sont robustes au bruit de mesure modélisé par du bruit blanc gaussien additif. Des expériences préliminaires ont permis de mettre en lumière que l’augmentation de données pratiquée en phase d’entraînement contribue à atteindre cette robustesse.

En raison de sa robustesse aux erreurs géométriques et de ses meilleures performances sur la base de données simulée avec le niveau de réalisme le plus élevé (G1-SDEB), les prochaines expériences se focaliseront sur le CNN. De plus, notons que selon la figure 4.3, les erreurs d’estimation du modèle CNN sur la base de données (G1-SDEB) sont similaires dans toutes les bandes d’octave. Par conséquent, les résultats du CNN présentés par la suite seront moyennés sur tout l’ensemble

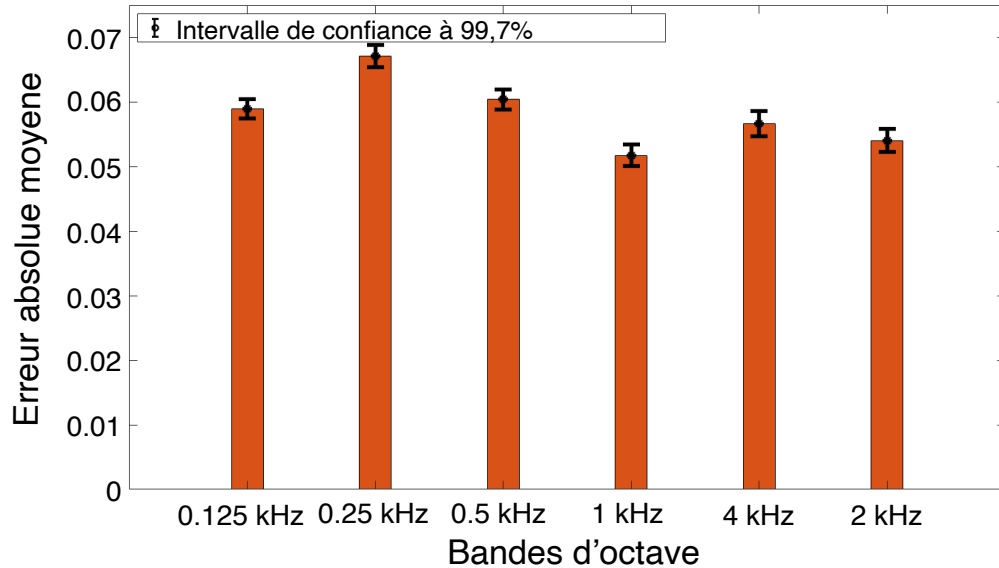


FIGURE 4.3 – Erreurs absolues moyennes (EAM) par bande d’octave du CNN pour la base de données G1-SDEB (voir tableau 4.3).

des bandes d’octave pour des questions de concision.

Une deuxième étude est conduite pour vérifier l’importance de fixer la géométrie du jeu d’entraînement pour l’estimation des profils d’absorption par cette approche. Le tableau 4.4 présente les EAM sur l’estimation des coefficients d’absorption obtenues pour les ensembles de données G1-SDEB, G2-SDEB et \mathcal{G} -SDEB. Les résultats obtenus par le CNN sur les bases de données G1-SDEB et G2-SDEB montrent qu’il parvient à estimer efficacement les paramètres cibles dans ces deux situations d’entraînement spécifique pour deux géométries fixes distinctes. Ce résultats suggèrent que l’approche proposée permet de considérer différentes géométries fixées et que celle-ci peut être étendue à d’autres géométries, au-delà du cas de salles en forme de pavé droit. En revanche, les performances du CNN sur \mathcal{G} -SDEB sont nettement dégradées et comparables à celles de l’estimateur de la moyenne. Cela confirme l’importance de fixer la géométrie de la scène acoustique dans les bases de données considérées par les modèles appris. Au fil de l’apprentissage, le modèle apprend à localiser et identifier les échos pertinents pour l’estimation des profils d’absorption des parois de la géométrie spécifique. Néanmoins, le fait que le CNN soit légèrement meilleur que l’estimateur de la moyenne dans cette situation suggère qu’il pourrait parvenir à extraire de l’information globale sur l’acoustique de la salle qui pourrait servir à l’estimation. Typiquement, cette information pourrait être liée à l’absorption moyenne de la salle par bande d’octaves, estimée par apprentissage supervisé virtuel dans [Foy et al., 2021](#), qui est plus informative que l’absorption moyenne par bande d’octave sur le jeu d’entraînement.

	G1-SDEB	G2-SDEB	\mathcal{G} -SDEB	Estimateur de la moyenne
Erreur absolue moyenne	0.0582	0.0682	0.1282	0.1364

TABLEAU 4.4 – Erreurs absolues moyennes (EAM) du CNN pour l’estimation des coefficients d’absorption en bande d’octave pour les bases de données G1-SDEB, G2-SDEB et \mathcal{G} -SDEB.

4.4.2 Étude qualitative : informativité des différentes composantes de la réponse impulsionnelle de salle

La troisième et dernière expérience menée dans cette étude vise à identifier les parties des RISs qui sont les plus utiles au modèle appris pour l’estimation des paramètres ciblés. À cette fin, chacune des RISs des jeux d’entraînement, de validation et de test issus de G1-SDEB (15 000 RISs au total) est divisée en une partie *précoce* et une partie *tardive*, et ce pour dix instants de coupure différents. Ces instants de coupure sont pris entre 5 ms et 500 ms sur une échelle temporelle logarithmique. Pour chaque instant de coupure, un modèle CNN est entraîné et testé sur le jeu des parties précoces d’une part et un autre modèle CNN est entraîné et testé sur le jeu des parties tardives d’autre part. L’instant de coupure égal à 500ms (durée des RISs) ne donne lieu qu’à un jeu de parties précoces équivalent aux RISs complètes. Il en résulte alors 19 modèles appris de CNN dont les performances sont présentées dans la figure 4.4.

La figure 4.4 montre donc l’évolution de l’EAM du CNN pour l’estimation des profils d’absorption en bande d’octave des parois d’une salle dans les conditions de G1-SDEB, en fonction de l’instant de coupure. Chaque graphique est associé à une des parois de la salle et fait apparaître le temps d’arrivée théorique du son direct et celui des échos spéculaires du premier ordre qui correspondent à la paroi. De plus, chacun d’entre eux renseigne les résultats obtenus en considérant soit les parties précoces, soit les parties tardives des RISs pour l’entraînement et le test.

L’examen de l’ensemble de ces graphiques révèle des points communs entre les parois pour l’évolution de l’erreur absolue moyenne en fonction du contenu de la RIS. D’une part, l’analyse des courbes associées aux parties tardives des RISs (courbes bleues) montre que l’erreur absolue moyenne augmente systématiquement et significativement dès lors que le son direct disparaît des RISs. Cette erreur est alors du même ordre de grandeur que celle obtenue pour le cas du CNN entraînés avec la base de données \mathcal{G} -SDEB pour laquelle les géométries sont distinctes et tirées aléatoirement. Elle reste inférieure à celle de l’estimateur de la moyenne (courbe verte) avant de la dépasser dans certains cas à partir de 50ms. Ce résultat suggère que le son direct est porteur d’une information géométrique forte et qu’en son absence, le modèle appris extrait des RISs de l’information acoustique et géométrique plus globales des parties tardives.

D’autre part, l’analyse des courbes associées aux parties précoces (courbes rouges) montre que

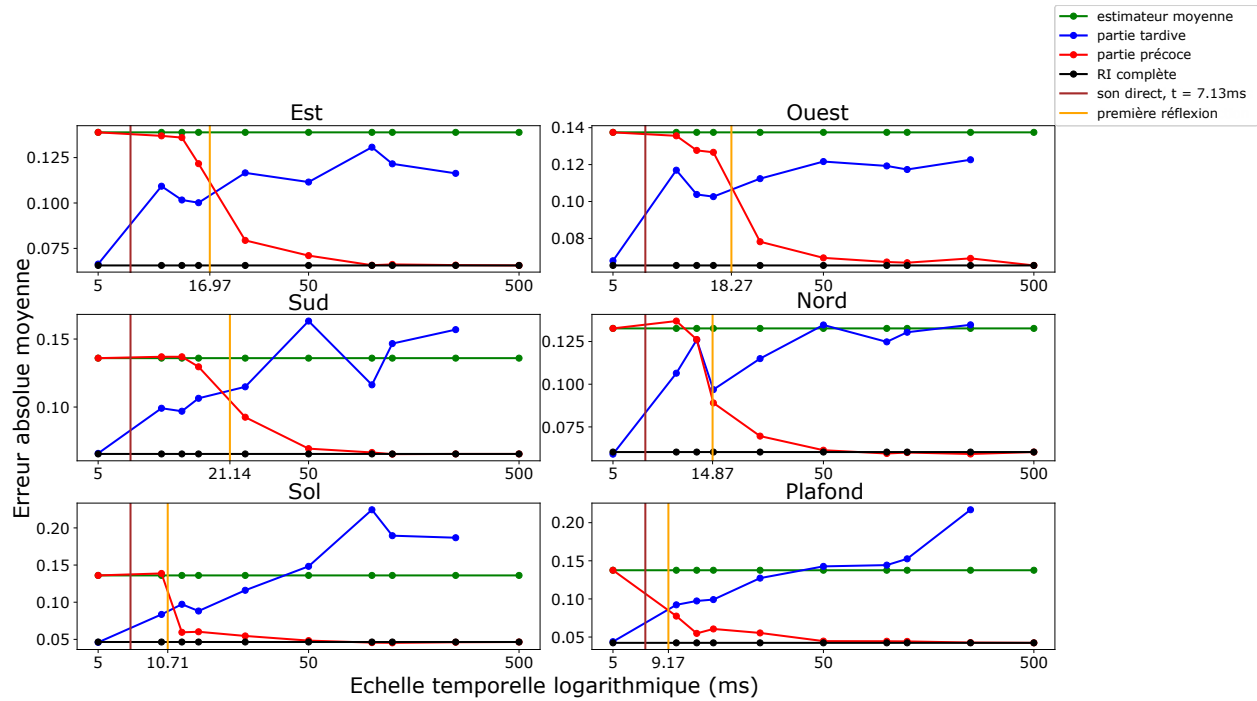


FIGURE 4.4 – Erreurs absolues moyennes du CNN sur l’estimation profils d’absorption en bande d’octave des 6 parois d’un salle pour la base de données G1-SDEB. Les erreurs obtenues en utilisant uniquement la partie précoce ou uniquement la partie tardive des RISs sont comparées pour différents instants de coupure. Pour chaque paroi, les temps d’arrivée du son direct et de l’écho du premier ordre associé sont indiqués (calcul par la méthode des sources-images appliquée à la géométrie du tableau 4.1).

pour chaque paroi, la plus forte décroissance de l’erreur absolue moyenne est systématiquement observée dès lors que l’écho d’ordre 1 associé à la paroi apparaît dans la RIS. Ce résultat confirme l’intuition suivant laquelle ce sont les échos du premier ordre qui portent l’information cruciale pour l’estimation des coefficients d’absorption des parois. Pour toutes les parois, les erreurs d’estimation obtenues à partir de la partie précoce des RISs deviennent proches de celles obtenues à partir de la RIS totale (courbes noires) une fois que le son direct et l’écho du premier ordre ont été observés. Cependant cette erreur continue de diminuer progressivement à mesure que la partie tardive de la RIS est incluse. À noter que plus le temps d’arrivée de du premier écho d’ordre 1 augmente, plus ce constat semble être vérifié et plus l’EAM est grande. En raison de l’augmentation de la densité des échos dans la RIS au cours du temps, l’augmentation du temps d’arrivée correspond à une augmentation du niveau de mélange avec d’autres échos. Dans cette situation, les modèles appris semblent être capables d’extraire de l’information additionnelle des échos d’ordre supérieur bénéfique pour améliorer les résultats d’estimation. Cependant, les parties diffuses plus tardives des RISs, après le temps de mélange qui marque la séparation entre les échos précoces et ceux tardifs (environ 50ms pour les géométries considérées), ne semblent pas porter d’information utile pour l’estimation des profils d’absorption individuels. Ce constat confirme que passé le temps de mixage,

les échos ne permettent plus d’apporter d’information précise sur la géométrie ce qui ne permet pas l’estimation d’améliorer des profils d’absorption individuels.

4.5 Conclusion

Dans ce chapitre, nous avons abordé le problème inverse de l’estimation des profils d’absorption des parois d’une salle décrits dans six bandes d’octaves à partir d’une seule RIS et de la connaissance approximative de la géométrie associée. Cette estimation est réalisée en utilisant des réseaux de neurones profonds entraînés sur des jeux simulés et annotés de RISs (entraînement virtuellement supervisé) correspondant à cette même géométrie.

Les expériences menées ont montré que les modèles entraînés par cette approche permettent d’obtenir une estimation fiable des coefficients d’absorption en bande d’octave. Les modèles appris surpassent significativement l’estimateur de la moyenne, quelle que soit la combinaison de distorsions appliquée aux données parmi celles considérées, et ce jusqu’au plus haut niveau de réalisme disponible. De plus, il a été observé que l’incertitude géométrique est l’élément de modélisation le plus susceptible d’impacter les résultats d’estimation et qu’il ne faut par conséquent pas négliger en vue d’applications réelles. Aussi, l’architecture CNN s’est montrée être d’une plus grande efficacité pour l’estimation que l’architecture MLP, principalement en raison de sa robustesse notable vis à vis des erreurs géométriques. Enfin, les études ont révélé que le son direct et les échos du premier ordre sont les composantes des RISs les plus importantes pour l’estimation des profils d’absorption des parois par cette approche. Il est aussi montré que les modèles appris parviennent à exploiter les échos d’ordre supérieur pour en extraire de l’information supplémentaire utile à l’estimation. Notons pas ailleurs que bien que présentée pour des salles en forme de pavé droit, l’approche proposée est en principe applicable à tout type de géométrie pouvant être modélisé par un simulateur acoustique.

Toutefois, cette approche présente plusieurs limites. Premièrement, cette approche ne se généralise pas à des géométries autres que celle rencontrée en phase d’entraînement. Ce constat était attendu car cette approche vise volontairement à spécialiser le modèle pour la géométrie spécifiquement considérée lors de son entraînement. Cette méthodologie est choisie en raison de la difficulté de renseigner l’information géométrique en entrée des réseaux de neurones. Il en résulte que cette approche est difficile à mettre en oeuvre : il est nécessaire simuler un large jeu de RISs et d’entraîner complètement un réseau de neurones pour chaque nouvelle géométrie considérée. Une application de cette approche à des données simulées avec un niveau de réalisme encore plus élevé incluant des mesures réelles de réponses de sources et de microphones aurait pu être envisagé grâce à une extension récente du simulateur Pyroomacoustics (Scheibler et al., 2018) présentée dans Srivastava et al., 2022. Cependant, la durée seule de l’étape de simulation d’une base de données de 15000 RISs avec cette extension du simulateur Pyroomacoustics est estimée à 39h38min. À cette durée s’ajoute

la durée cumulée de l'ensemble des étapes relatives à la sélection et l'entraînement d'un réseau de neurones, qui dépend des ressources de calcul à disposition mais est en général également chronophage. Il s'agit de la limite principale et objective de cette approche qui motive le développement d'approches "informées" par la géométrie dans la suite de cette thèse. Une approche informée par la géométrie pour la résolution du problème inverse pour des simulation incluant des mesures réelles de réponses de sources et de microphones à dépendance spatiale et fréquentielle sera présenté dans la suite de cette thèse.

Enfin, de par la nature de cette approche, des questions communes se posent quant à la généralisation de l'apprentissage sur des données synthétiques à des tests sur des données réelles. En effet, il est possible qu'une telle approche soit fortement influencée par la modélisation des données (par exemple la modélisation des réponses des murs) ou bien par la diversité au sein du jeu d'entraînement. Un travail de modélisation et d'entraînement additionnel approfondi pourrait être envisagé afin de garantir cette généralisation. De plus, les réseaux de neurones sont ici utilisés tels des boîtes noires. Une piste d'amélioration pourrait être de faire intervenir des modèles physiques ce qui pourrait présenter l'avantage d'accroître la robustesse et la généralisation des modèles.

Le chapitre suivant présente donc une méthode d'estimation *informée par la géométrie* par optimisation sur les spectrogrammes des réponses impulsionnelles. Il est important de noter que les résultats obtenus dans ce chapitre servent de guide pour le développement des approches suivantes mais ne sont pas comparables directement à ceux présentés dans les chapitres suivant. En effet, la suite de cette thèse s'intéresse à l'estimation des coefficients d'absorption sur une échelle linéaire des fréquences, avec une résolution fréquentielle nettement supérieure aux six bandes d'octaves considérées dans cette étude.

Estimation informée par la géométrie par optimisation sur les spectrogrammes des réponses impulsionnelles de salle

Préambule : Ce chapitre présente une méthode d’optimisation pour estimer conjointement les profils d’absorption des parois d’une salle, décrits sur une échelle linéaire des fréquences, à partir des spectrogrammes de plusieurs réponses impulsionnelles de salle. Cette méthode repose sur l’identification au sein des spectrogrammes des fenêtres temporelles de taille fixe qui incluent entièrement un unique écho acoustique isolé afin d’éviter les difficultés causées par leur superposition dans le domaine énergétique. Ces fenêtres sont identifiées en se basant sur la géométrie associée aux réponses impulsionnelles de salle, connue approximativement aux erreurs de mesure près. Une procédure probabiliste inspirée de l’algorithme *Random Sample Consensus* (RANSAC) (Bolles et al., 1981) est mise en place afin de construire un sous-ensemble de fenêtres pertinentes formant un modèle acoustique cohérent. Une fois sélectionnées, les fenêtres temporelles pertinentes sont utilisées pour estimer les paramètres d’intérêt à chaque fréquence en minimisant sous contraintes une fonction objectif définie dans le domaine des spectrogrammes. Les travaux présentés dans ce chapitre ont fait l’objet d’une publication (Dilungana et al., 2022) et d’une présentation à la conférence EUSIPCO 2022.

5.1 Introduction

Dans le chapitre précédent, une méthode d’estimation des profils d’absorption des murs d’une salle par apprentissage supervisé virtuel pour une géométrie spécifique était proposée. Bien qu’efficace pour la résolution du problème étudié, cette méthode nécessite le renouvellement d’une procédure complète de simulation de données et d’entraînement de réseau de neurones pour chaque nouvelle géométrie considérée, avec un coût computationnel important. Dans ce chapitre, une nouvelle approche informée par la géométrie et applicable directement pour toute géométrie compatible avec la méthode des sources-images est présentée.

Cette approche est purement basée sur la formulation d’un problème d’optimisation issu d’une

modélisation signal des RISs. Elle repose sur l'utilisation des *spectrogrammes* des RISs. Le spectrogramme d'une RIS correspond au carré du module de sa *Transformée de Fourier à Court Terme* (TFCT). Cette représentation énergétique temps-fréquence présente plusieurs avantages vis à vis du problème étudié. Premièrement, le spectrogramme fournit une description énergétique de la RIS permettant de faire intervenir directement les coefficients d'absorption comme indiqué par (2.24). De plus, il présente l'avantage de conserver une résolution temporelle du signal permettant d'identifier les zones d'apparition des échos et murs associés grâce à la méthode des sources-images. Aussi, il conserve une résolution fréquentielle permettant une analyse fine de l'effet des murs sur le son dans différentes plages de fréquences. Enfin, il permet de s'affranchir de la difficulté de modéliser les phases spectrales des échos acoustiques, due à la fois aux erreurs géométriques et aux réponses en phase inconnues des murs.

Cependant, plusieurs difficultés se présentent lorsqu'il s'agit d'accéder aux coefficients d'absorption à partir des spectrogrammes des RISs. Tout d'abord, la présence partielle d'un écho au sein d'une fenêtre temporelle du spectrogramme entraîne une invalidité du théorème de convolution : il n'est alors plus possible d'exprimer son énergie en fonction du produit des coefficients d'absorption des murs associées. De plus, la présence de plusieurs échos au sein d'une même fenêtre temporelle du spectrogramme produit des *interférences* en conséquence de la suppression de la phase du signal, rendant alors l'information énergétique inexploitable. Pour pallier ces obstacles, l'approche proposée vise donc à exploiter l'information sur la géométrie pour identifier grâce à la méthode des sources-images les fenêtres temporelles des spectrogrammes qui incluent entièrement un unique écho isolé. Néanmoins, la présence d'erreurs sur la mesure des paramètres géométriques se répercute sur le calcul des temps d'arrivée des sources-images par la méthode des sources-images et donc sur l'identification des fenêtres valides.

L'approche proposée consiste donc à utiliser les spectrogrammes des RISs, la connaissance approximative des paramètres géométriques et l'algorithme probabiliste RANSAC (Bolles et al., 1981) pour sélectionner de manière robuste les fenêtres temporelles des RISs qui incluent entièrement un unique écho. L'estimation est ensuite formulée comme un problème d'optimisation sous contraintes. Le sous-ensemble des fenêtres finalement sélectionnées est utilisé pour optimiser une fonction objectif des paramètres d'intérêt dans le domaine des amplitudes carrées tenant compte des relations multiplicatives entre les coefficients de réflexion pour les réflexions multiples. Une étude approfondie permet de montrer l'efficacité et la robustesse aux erreurs géométriques et au bruit de mesure de cette approche dans le cas de RISs simulées en négligeant l'effet des sources et des microphones. Des expériences simulées sur 500 salles aléatoires en forme de pavé droit révèlent que l'approche permet de retrouver conjointement 80% des coefficients d'absorption des 6 murs dans 16 bandes de fréquences espacées linéairement, avec une erreur absolue moyenne inférieure à 0.1 en utilisant seulement 4 RISs par salle. De plus, il est démontré que l'augmentation du nombre de RISs utilisés pour l'estimation améliore les résultats, tout en renforçant la robustesse face au bruit de mesure et à l'incertitude géométrique.

Le chapitre est organisé comme suit. La Section 5.2 présente le principe de l’approche. La Section 5.3, porte sur les expériences menées et les résultats obtenus. Enfin, la Section 5.4 apporte une conclusion et met en avant des axes d’amélioration possibles.

5.2 Présentation de l’approche

5.2.1 Modélisation du signal

Considérons une salle de géométrie connue délimitée par S murs dans laquelle sont mesurées M RISs. Pour $m \in \llbracket 1, M \rrbracket$, soit x_m une RIS. D’après (2.20), la modélisation de x_m issue de la méthode des sources-images peut s’écrire :

$$x_m[t] = \sum_{\mathbf{k} \in \mathcal{K}} (\eta_{m,\mathbf{k}} \cdot \phi_{m,\mathbf{k}} * g_{m,\mathbf{k}} * v_{\mathbf{k}})[t] + e_m[t]. \quad (5.1)$$

avec $\eta_{m,\mathbf{k}}$ le facteur d’atténuation géométrique, $\phi_{m,\mathbf{k}}$ le filtre de délai fractionnaire, $g_{m,\mathbf{k}}$ la réponse source-microphone et $v_{\mathbf{k}}$ la réponse impulsionnelle cumulée (RIM), tous correspondant à la source-image avec $\mathbf{k} \in \mathcal{K}$ et la RIS x_m . Le terme $e_m[t]$ désigne l’erreur du modèle.

Dans cette étude, les sources et microphones utilisés sont considérés identiques et omnidirectionnels. De plus, le dispositif expérimental considéré est *calibré*, c’est-à-dire que les gains d’émission et de réception des appareils sont harmonisés. Il en résulte que les réponses source-microphone appliquées à chacune des RISs sont identiques et ne dépendent pas de la direction. Cette réponse unique est modélisée par un filtre à réponse impulsionnelle finie (RIF) $d[t]$ considéré inconnu :

$$\forall m \in \llbracket 1, M \rrbracket, \forall \mathbf{k} \in \mathcal{K}, \forall t \in \llbracket 1, T \rrbracket, g_{m,\mathbf{k}}[t] = d[t].$$

Soit $F \in \mathbb{N}^*$. Soit $\tilde{x}_{m,t}[f] = \tilde{x}_m[t : t + 2F - 1][f]$ le carré du module de la TFD de la RIS fenêtrée $x_m[t : t + 2F - 1] \in \mathbb{R}^{2F}$ à la fréquence f . L’ensemble $\{\tilde{x}_{m,t}[f], (t, f) \in \llbracket 1, T \rrbracket \times \llbracket 0, F \rrbracket\}$ constitue le *spectrogramme* de x avec une fenêtre temporelle glissante de taille $2F$ et un pas égal à 1. Il peut être représenté par la matrice $\tilde{X}_m \in (\mathbb{R}^+)^{F+1, T}$ de terme général $\tilde{X}_m[f, t] = \tilde{x}_{m,t}[f]$.

Définissons L_v la longueur maximale supposée des échos, en nombre d’échantillons. Nous supposons ici que la taille de la fenêtre glissante du spectrogramme est supérieure à la longueur maximale des échos, c’est-à-dire $2F > L_v$. Comme évoqué dans 2.2, les réponses temporelles des murs sont supposées être de courte durée. La modélisation d’un écho fait intervenir des convolutions multiples entre la réponse source-microphone et les réponses impulsionnelles des murs associées (RIMs). L’hypothèse $2F > L_v$ est donc d’autant plus plausible que la réponse source-microphone

$d[t]$ est de courte durée. Cette considération sera discutée par la suite.

Suivant cette hypothèse, l'ensemble des *fenêtres pertinentes* associées à $\mathbf{k} \in \mathcal{K}$, c'est-à-dire qui incluent entièrement et isolent \mathbf{k} est défini par :

$$\mathcal{J}_{\mathbf{k}}^* = \{(m, t) | \tau_{m, \mathbf{k}} \in \mathcal{W}_t^-, \tau_{m, \mathbf{k}' \neq \mathbf{k}} \notin \mathcal{W}_t^+\}, \quad (5.2)$$

où $\mathcal{W}_t^- = [t : t + 2F - L_v]$ et $\mathcal{W}_t^+ = [t - L_v : t + 2F - 1]$. En d'autres termes, $\mathcal{J}_{\mathbf{k}}^*$ est l'ensemble des couples (m, t) pour lesquels \mathbf{k} est la seule source-image visible dans la fenêtre $[t : t + 2F - 1]$ de x_m . La figure 5.1 illustre deux situations pour laquelle le couple (m, t) vérifie ou non ce critère pour \mathbf{k} . Par application de la TFD et du théorème de la convolution discrète (approximation), pour tout $\mathbf{k} \in \mathcal{K}$, l'énergie du spectrogramme $\tilde{x}_{m,t}$ pour $(m, t) \in \mathcal{J}_{\mathbf{k}}^*$ vérifie l'expression simple suivante :

$$\forall \mathbf{k} \in \mathcal{K}, \forall (m, t) \in \mathcal{J}_{\mathbf{k}}^*, \forall f \in \llbracket 0, F \rrbracket, \quad \tilde{x}_{m,t}[f] = (\eta_{m, \mathbf{k}}^2 \cdot \tilde{d} \cdot \tilde{v}_{\mathbf{k}})[f] + \tilde{e}_{m,t}[f], \quad (5.3)$$

où $\tilde{d} = |\text{TFD}\{g_{m, \mathbf{k}}\}|^2 = |\text{TFD}\{d\}|^2$, $\tilde{v}_{\mathbf{k}} = |\text{TFD}\{v_{\mathbf{k}}\}|^2$ et $\tilde{e}_{m,t}[f]$ est un terme d'erreur.

Notons que les filtres à délai fractionnaire sont supposés n'avoir d'effet que sur les phases des signaux et non sur les amplitudes. Il en résulte que $\tilde{\phi}_{m, \mathbf{k}} = |\text{TFD}\{\phi_{m, \mathbf{k}}\}|^2 = 1$. Dans cette étude, le filtre de délai fractionnaire considéré est le sinus cardinal :

$$\forall m \in \llbracket 1, M \rrbracket, \forall \mathbf{k} \in \mathcal{K}, \phi_{m, \mathbf{k}}[\cdot] = \text{sinc}[\cdot - \tau_{\mathbf{k}}] = \frac{\sin[\cdot - \tau_{\mathbf{k}}]}{[\cdot - \tau_{\mathbf{k}}]}. \quad (5.4)$$

L'équation (5.3) fait intervenir $\tilde{v}_{\mathbf{k}}$ qui s'exprime en fonction des profils de réflexion $\{\omega_k\}_{k \in \mathbf{k}}$ conformément à (2.24) :

$$\begin{aligned} \forall \mathbf{k} \in \mathcal{K} \setminus (\cdot), \forall f \in \llbracket 0, F \rrbracket, \tilde{v}_{\mathbf{k}}[f] &= \prod_{k \in \mathbf{k}} \tilde{R}_k[f] \\ &= \prod_{k \in \mathbf{k}} |\text{TFD}\{R_k\}|^2[f] \\ &= \prod_{k \in \mathbf{k}} \omega_k[f], \end{aligned}$$

avec R_k la réponse impulsionnelle du mur (RIM) numérotée k , et $\tilde{v}_{(\cdot)} = 1$ pour le son direct.

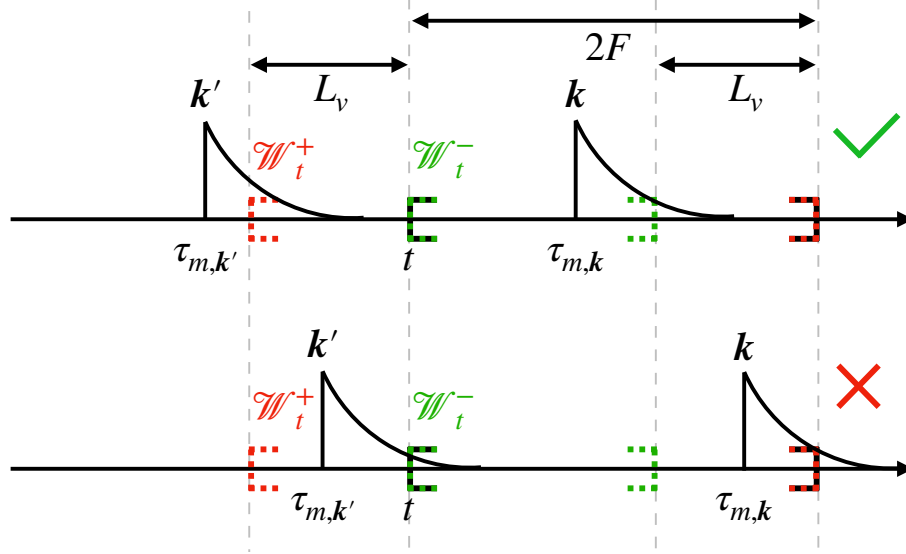


FIGURE 5.1 – Illustration d’une situation de validation et d’une situation de non validation de la condition définie par \mathcal{J}_k^* par la fenêtre temporelle $[t : t + 2F - 1]$ d’une RIS x_m délimitée par des crochets noirs. Les intervalles \mathcal{W}_t^- et \mathcal{W}_t^+ correspondant sont respectivement délimités par des crochets verts et des crochets rouges. Le premier schéma présente une situation valide où la fenêtre $[t : t + 2F - 1]$ permet d’isoler et d’inclure complètement l’écho associé à k . Le second schéma présente une situation non valide où la fenêtre $[t : t + 2F - 1]$ ne permet ni d’isoler l’écho k , ni de l’inclure complètement.

Notons que ω_k correspond au *profil de réflexion* du mur k , tandis que $\alpha_k = 1 - \omega_k$ correspond à son *profil d’absorption*, tous deux prenant des valeurs dans $[0, 1]^{F+1}$ (seules les fréquences non négatives sont considérées). Classiquement, en acoustique du bâtiment, les profils d’absorption sont généralement définis sur 6 bandes d’octave espacées logarithmiquement centrées sur les fréquences .125, .250, \dots , 4 kHz. Dans cette étude, ces derniers sont décrits sur une l’échelle linéaire des fréquences positives dans le domaine de Fourier, cela étant plus naturel d’un point de vue du traitement du signal et de la théorie de l’information. Néanmoins, une échelle peut être convertie en l’autre en utilisant des schémas d’interpolation appropriés.

5.2.2 Fonction objectif

L’expression (5.3) vérifiée par les spectrogrammes des RISs sur l’ensemble des fenêtres pertinentes défini pour tout $k \in \mathcal{K}$ par (5.2), permet de définir un critère pour l’estimation des paramètres d’intérêt par une approche d’optimisation. Pour chaque f , l’estimation des coefficients de réflexion $\omega_1[f], \dots, \omega_S[f]$ ainsi que du gain source-microphone $\tilde{d}[f]$ étant donné $\{\tilde{x}_{m,t}[f]\}_{m,t}$ et les temps

d'arrivée $\{\tau_{m,\mathbf{k}}\}_{m,\mathbf{k}}$ peut maintenant être formulé comme le problème de minimisation suivant :

$$\underset{\substack{\tilde{d}[f] \geq 0, \\ \omega_{1:S}[f] \in [0,1]}}{\operatorname{argmin}} \sum_{\substack{\mathbf{k} \in \mathcal{K}, \\ q(\mathbf{k}) \leq Q}} \sum_{(m,t) \in \mathcal{J}_{\mathbf{k}}^*} |\tilde{x}_{m,t}[f] - (\eta_{m,\mathbf{k}}^2 \cdot \tilde{d} \cdot \tilde{v}_{\mathbf{k}})[f]|^2, \quad (5.5)$$

avec les expressions pour $\tilde{v}_{\mathbf{k}}[f]$ données par (2.24) et $q(\mathbf{k}) \in \llbracket 0, Q_{\max} \rrbracket$ indiquant l'ordre de la source-image \mathbf{k} . Notons que Q , l'ordre maximal des échos dont les fenêtres valides sont utilisées pour l'estimation, peut être inférieur à Q_{\max} , l'ordre maximal de modélisation des échos par la méthode des sources-images, utilisé ici pour identifier les fenêtres pertinentes de (5.3). Notons également que la construction des $\{\mathcal{J}_{\mathbf{k}}^*\}_{\mathbf{k}}$ constitue une méthode de localisation et de séparation des échos dans la RIS. L'optimisation du critère (5.5) à chaque fréquence à partir de ces sous-ensembles de fenêtres pertinentes permet d'opérer la déconvolution aveugle de la réponse source-microphone $d[t]$ ainsi que l'estimation des profils de réflexion $\omega_1, \dots, \omega_S$.

Il s'agit d'un problème d'optimisation non convexe, non linéaire et contraint pouvant être vu comme un problème des moindres carrés linéaire sous contraintes pour chaque variable individuelle. Il peut être résolu à l'aide d'un solveur non linéaire correctement initialisé. Dans cette étude, nous avons utilisé le solveur `fmincon` de Matlab. En pratique, une estimation initiale de $\tilde{d}[f]$ peut être obtenue en utilisant $Q = 0$. La solution obtenue correspond à la moyenne des valeurs prises par les spectrogrammes sur les fenêtres pertinentes associées au son direct ($\mathcal{J}_{\emptyset}^*$) après compensation de l'atténuation géométrique. Une estimation initiale pour chaque $\omega_k[f]$ peut ensuite être obtenue de manière similaire en fixant $\tilde{d}[f]$ et en utilisant $Q = 1$.

5.2.3 Analogie avec la factorisation en matrices non-négatives

Il est intéressant de noter que le critère (5.5) est inspiré de la *Factorisation en Matrices Non-négatives* (Paatero et al., 1994) (NMF pour *Non-negative Matrix Factorization* en anglais). En effet, le problème sous-jacent peut être formulé en tant que produit de deux matrices aux coefficients positifs, la première étant inconnue et fonction des paramètres acoustiques recherchés et la seconde étant connue et construite à partir de l'information géométrique. Considérons dans un premier temps la matrice $\tilde{X}_m^* \in (\mathbb{R}^+)^{F+1,T}$, construite à partir du spectrogramme \tilde{X}_m en annulant à toutes les fréquences les fenêtres temporelles qui ne sont pertinentes pour aucune des sources-images de \mathcal{K} au sens de (5.2). Cette matrice \tilde{X}_m^* peut être vue comme la multiplication terme à terme de \tilde{X}_m

et d'un masque binaire \mathcal{M}_m associé, de terme général : $\mathcal{M}_m[f, t] = \begin{cases} 1 & \text{si } (m, t) \in \cup_{\mathbf{k}} \mathcal{J}_{\mathbf{k}}^* \\ 0 & \text{sinon} \end{cases}$. Notons

$\tilde{\mathbf{X}}^* = [\tilde{X}_1^* \quad \tilde{X}_2^* \quad \dots \quad \tilde{X}_M^*]$ la concaténation des M spectrogrammes masqués. Dans un second

temps, considérons la matrice $H_m^* \in (\mathbb{R}^+)^{K,T}$ telle que $H_m^*[\mathbf{k}, t] = \begin{cases} \eta_{m,\mathbf{k}}^2 & \text{si } (m, t) \in \mathcal{J}_{\mathbf{k}}^* \\ 0 & \text{sinon} \end{cases}$, appelée *matrice d'activation*. Cette matrice d'activation pondérée par l'atténuation géométrique renseigne pour chaque source-image les fenêtres temporelles pertinentes dans \tilde{X}_m^* . En conséquence de (5.2), chaque colonne de H_m^* comporte au plus un coefficient non nul. Notons $\mathbf{H}^* = [H_1^* \ H_2^* \ \dots \ H_M^*]$ la concaténation des M matrices d'activation. Les matrices $\tilde{\mathbf{X}}^*$ et \mathbf{H}^* ainsi construites permettent de définir le problème de NMF suivant :

$$\underset{W \geq 0}{\operatorname{argmin}} \|\tilde{\mathbf{X}}^* - W\mathbf{H}^*\|_2. \quad (5.6)$$

La matrice estimée $W \in (\mathbb{R}^+)^{F+1,K}$ est appelée *dictionnaire*. Chacune de ses colonnes, appelée *atome*, représente l'énergie d'une source-image de \mathcal{K} en fonction de la fréquence. Dans une situation idéale où les ensembles $\mathcal{J}_{\mathbf{k}}^*$ sont parfaitement construits et qu'une quantité suffisante de mesures est disponible pour assurer qu'ils soient tous non vides, il est attendu que le dictionnaire de terme général $W[f, \mathbf{k}] = \tilde{d} \cdot \tilde{v}_{\mathbf{k}}[f]$ soit solution de (5.6). En effet, par construction de $\tilde{\mathbf{X}}^*$ et \mathbf{H}^* , le critère (5.6) permet d'estimer W en combinant plusieurs mesures de RISs et uniquement à partir des fenêtres temporelles qui vérifient l'expression (5.3). Par ailleurs, les atomes étant estimés indépendamment, l'expression (2.24) indique que l'estimation seule des atomes associés aux sources-images d'ordre 0 et 1 permet d'accéder aux coefficients de réflexion $\omega_1[f], \dots, \omega_S[f]$ à chaque fréquence. De plus, la contrainte $W \geq 0$ peut être omise grâce à la positivité des coefficients de $\tilde{\mathbf{X}}^*$ et \mathbf{H}^* ainsi que par l'absence de mélange entre les échos. À noter qu'il est envisageable de lier l'estimation des atomes grâce à des versions contraintes de la NMF (Bertin et al., 2010).

Néanmoins, des expériences préliminaires ont montré que cette approche permet d'obtenir des résultats satisfaisant lorsque la géométrie est parfaitement connue, mais qu'elle s'effondre en présence d'erreurs sur la mesure des paramètres géométriques. En effet, une connaissance approximative de la géométrie peut conduire à considérer comme pertinentes certaines fenêtres comportant plusieurs échos avec des interférences qui impactent et dégradent l'estimation. C'est ce que constat qui motive la construction du critère (5.5) inspiré de (5.6), permettant d'implémenter explicitement les contraintes multiplicatives entre les coefficients de réflexion pour plus de robustesse dans l'estimation et d'utiliser une procédure de sélection des fenêtres pertinentes avancée pour tenter de s'affranchir des interférences en présence d'incertitude géométrique.

5.2.4 Modélisation probabiliste de l'incertitude géométrique

En pratique, les positions des sources, des microphones et des murs de la salle ne sont jamais connues avec exactitude. Ces paramètres géométriques sont connus avec une certaine incertitude

liée aux imprécisions des appareils de mesure et au processus de mesure lui-même. Cependant, des erreurs de quelques centimètres suffisent pour impacter significativement l'estimation des temps d'arrivée et donc le calcul de $\mathcal{J}_{\mathbf{k}}^*$. En effet, des fenêtres erronées ne comportant pas l'écho attendu ou présentant plusieurs échos peuvent alors être incluses par erreur dans cet ensemble. À l'inverse, certaines fenêtres correctes peuvent être injustement exclues de $\mathcal{J}_{\mathbf{k}}^*$. Notons qu'un choix inapproprié du paramètre L_v peut contribuer à renforcer ces effets, impactant alors significativement les résultats obtenus par la minimisation de (5.5).

Afin de tenir compte de cette incertitude, nous faisons ici l'hypothèse que l'erreur sur la mesure de chaque paramètre géométrique (positions des sources, des microphones et des murs) suit une loi normale centrée d'écart-type σ_{geo} , en centimètres. Un modèle probabiliste gaussien est alors posé sur les temps d'arrivée calculés par la méthode des sources-images à partir de la mesure erronée des paramètres géométriques, notés $\{\tau_{m,\mathbf{k}}^*\}_{m,\mathbf{k}}$:

$$\begin{aligned} \forall m \in \llbracket 1, M \rrbracket, \forall \mathbf{k} \in \mathcal{K}, \quad \tau_{m,\mathbf{k}}^* &= \tau_{m,\mathbf{k}} + \Delta_{m,\mathbf{k}} \text{ avec } \Delta_{m,\mathbf{k}} \sim \mathcal{N}(0, \sigma_{q(\mathbf{k})}^2), \\ \implies \tau_{m,\mathbf{k}} &\sim \mathcal{N}(\tau_{m,\mathbf{k}}^*, \sigma_{q(\mathbf{k})}^2), \end{aligned} \quad (5.7)$$

où $\mathcal{N}(\mu, \sigma^2)$ désigne la loi normale de moyenne μ et d'écart-type σ , $\tau_{m,\mathbf{k}}^*$ est le temps d'arrivée *erroné* de \mathbf{k} dans x_m , $\tau_{m,\mathbf{k}}$ est le temps d'arrivée *exact* de \mathbf{k} dans x_m , $\Delta_{m,\mathbf{k}}$ est le délai relatif entre $\tau_{m,\mathbf{k}}^*$ et $\tau_{m,\mathbf{k}}$ (résultant de l'incertitude géométrique) et $\sigma_{q(\mathbf{k})}$ est l'écart-type de l'erreur commise lors du calcul des temps d'arrivée des sources-images d'ordre $q(\mathbf{k})$.

Par la suite, le symbole “*” sera utilisé pour désigner des scalaires, vecteurs ou matrices dont le calcul est affecté par les erreurs géométriques.

Nous faisons l'hypothèse que les erreurs commises sur la mesure des paramètres géométriques sont indépendantes et identiquement distribuées suivant $\mathcal{N}(0, \sigma_{\text{geo}}^2)$. L'écart-type $\sigma_{q(\mathbf{k})}$ lié au cumul des erreurs sur les positions d'une source, d'un microphone et de $q(\mathbf{k})$ murs est alors exprimé par :

$$\sigma_{q(\mathbf{k})} = (q(\mathbf{k}) + 2)\sigma_{\text{geo}}\bar{f}/c. \quad (5.8)$$

Les $\{\tau_{m,\mathbf{k}}^*\}_{m,\mathbf{k}}$ sont des données du problème tandis que les $\{\tau_{m,\mathbf{k}}\}_{m,\mathbf{k}}$ et les $\{\Delta_{m,\mathbf{k}}\}_{m,\mathbf{k}}$ sont des inconnues du problème. En conséquence, nous définissons $\mathcal{J}_{\mathbf{k}}^*$, l'ensemble des fenêtres pertinentes associées à $\mathbf{k} \in \mathcal{K}$ calculé à partir des $\{\tau_{m,\mathbf{k}}^*\}_{m,\mathbf{k}}$, pouvant être utilisé à la place de $\mathcal{J}_{\mathbf{k}}^*$ dans (5.5) :

$$\mathcal{J}_{\mathbf{k}}^* = \left\{ (m, t) \mid \tau_{m,\mathbf{k}}^* \in \mathcal{W}_t^-, \tau_{m,\mathbf{k}' \neq \mathbf{k}}^* \notin \mathcal{W}_t^+ \right\}. \quad (5.9)$$

Par ailleurs, nous définissons à partir de (5.7) les probabilités $\{\pi_{m,t,\mathbf{k}}\}_{m,t,\mathbf{k}}$ telles que :

$$\begin{aligned}\pi_{m,t,\mathbf{k}} &= p((m,t) \in \mathcal{J}_{\mathbf{k}}^*) \\ &= p(\tau_{m,\mathbf{k}} \in \mathcal{W}_t^- \cap \bigcap_{\substack{\mathbf{k}' \in \mathcal{K} \\ \mathbf{k}' \neq \mathbf{k}}} \tau_{m,\mathbf{k}' \neq \mathbf{k}} \notin \mathcal{W}_t^+).\end{aligned}\quad (5.10)$$

À partir de (5.7) et en faisant l'hypothèse que les variables aléatoires $\{\Delta_{m,\mathbf{k}}\}_{m,\mathbf{k}}$ sont indépendantes, il est possible de calculer les $\{\pi_{m,t,\mathbf{k}}\}_{m,t,\mathbf{k}}$ de la manière suivante :

$$\begin{aligned}\pi_{m,t,\mathbf{k}} &= p(\tau_{m,\mathbf{k}} \in \mathcal{W}_t^-) \prod_{\substack{\mathbf{k}' \in \mathcal{K} \\ \mathbf{k}' \neq \mathbf{k}}} p(\tau_{m,\mathbf{k}'} \notin \mathcal{W}_t^+) \\ &= p(\tau_{m,\mathbf{k}} \in \mathcal{W}_t^-) \prod_{\substack{\mathbf{k}' \in \mathcal{K} \\ \mathbf{k}' \neq \mathbf{k}}} (1 - p(\tau_{m,\mathbf{k}'} \in \mathcal{W}_t^+)) \\ &= \frac{1}{2} \left(\operatorname{erf}\left(\frac{t + 2F - L_v - \tau_{m,\mathbf{k}}^*}{\sqrt{2}\sigma_q(\mathbf{k})}\right) - \operatorname{erf}\left(\frac{t - \tau_{m,\mathbf{k}}^*}{\sqrt{2}\sigma_q(\mathbf{k})}\right) \right) \\ &\quad \prod_{\substack{\mathbf{k}' \in \mathcal{K} \\ \mathbf{k}' \neq \mathbf{k}}} \left(1 - \frac{1}{2} \left(\operatorname{erf}\left(\frac{t + 2F - \tau_{m,\mathbf{k}'}^*}{\sqrt{2}\sigma_q(\mathbf{k}')}\right) - \operatorname{erf}\left(\frac{t - L_v - \tau_{m,\mathbf{k}'}^*}{\sqrt{2}\sigma_q(\mathbf{k}')}\right) \right) \right),\end{aligned}$$

où erf désigne la *fonction d'erreur* ($\{\operatorname{erf} : t \mapsto \frac{2}{\sqrt{\pi}} \int_0^t e^{-u^2} du\}$), telle que pour $Y \sim \mathcal{N}(\mu, \sigma^2)$, pour $(a, b) \in \mathbb{R}^2$, $p(Y \in [a, b]) = \frac{1}{2} \left(\operatorname{erf}\left(\frac{b-\mu}{\sqrt{2}\sigma}\right) - \operatorname{erf}\left(\frac{a-\mu}{\sqrt{2}\sigma}\right) \right)$. Notons que en pratique, les fenêtres avec des probabilités inférieures à 0.1% sont écartées.

Nous définissons alors $\mathcal{J}_{\mathbf{k}}^{\text{top}}$, l'ensemble des fenêtres associées à $\mathbf{k} \in \mathcal{K}$ dont la probabilité de pertinence est maximale :

$$\mathcal{J}_{\mathbf{k}}^{\text{top}} = \{\operatorname{argmax}_{(m,t)} \pi_{m,t,\mathbf{k}}\}. \quad (5.11)$$

Ce nouvel ensemble peut également être utilisé à la place de $\mathcal{J}_{\mathbf{k}}^*$ dans (5.5). Cependant, se fier uniquement à $\mathcal{J}_{\mathbf{k}}^{\text{top}}$ présente des limites car cet ensemble d'observations est bien plus restreint que $\mathcal{J}_{\mathbf{k}}^*$. De plus, indépendamment de l'incertitude géométrique, certaines des observations indexées par $\mathcal{J}_{\mathbf{k}}^*$ ou $\mathcal{J}_{\mathbf{k}}^{\text{top}}$ peuvent contenir des informations incorrectes en raison d'interférences avec des échos d'ordre supérieur à Q_{\max} ou en raison du bruit de mesure. Afin de pallier ce problème, une approche combinant le modèle géométrique probabiliste (5.7) avec une procédure robuste de sélection de fenêtres basée sur l'algorithme RANSAC (Bolles et al., 1981) est proposée.

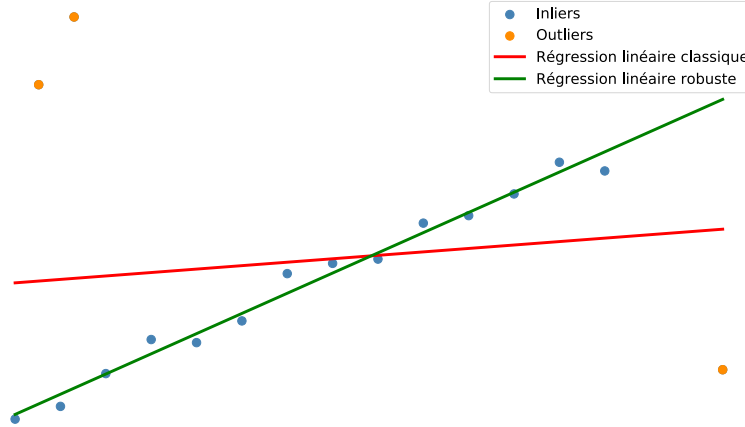


FIGURE 5.2 – Illustration de la robustesse dans le cas d’une régression linéaire en deux dimensions. Le modèle linéaire associé à la droite rouge est estimé par le critère des moindres carrés sur l’ensemble des observations. Le modèle linéaire associé à la droite verte est lui estimé par le critère des moindres carrés sur un ensemble d’inliers, constitué typiquement à l’issue de l’algorithme RANSAC. Ce modèle permet de généraliser convenablement la tendance linéaire suivie par la majorité des observations sans influence des outliers, c’est-à-dire de manière robuste aux outliers.

5.2.5 Algorithme de sélection robuste

L’algorithme **RANSAC** (pour Random Sample Consensus) est une méthode itérative d’estimation des paramètres d’un modèle donné à partir d’un ensemble d’observations. C’est un algorithme dit *robuste* : il est capable de fournir un modèle qui généralise correctement la distribution des observations sans être affecté par celles qui présentent des valeurs extrêmes ou aberrantes, appelées *outliers*. À chaque itération, une estimation des paramètres du modèle est effectuée à partir d’un sous-ensemble d’observations *candidates* échantillonné aléatoirement. Ce modèle estimé est évalué grâce à un score calculé à partir d’un sous-ensemble d’observations en adéquation avec celui-ci (par rapport à un seuil d’erreur fixé) appelées *inliers*. Classiquement, ce score correspond à une erreur moyenne d’estimation par le modèle ou encore au nombre d’inliers. Finalement, le modèle retenu est celui qui maximise le score au cours des itérations. La figure 5.2 donne un exemple de résultat d’estimation robuste des paramètres d’un modèle linéaire en deux dimension pouvant typiquement être obtenu grâce à l’algorithme RANSAC. L’algorithme RANSAC peut aussi être utilisé pour identifier et exclure les outliers d’un ensemble d’observations. C’est l’utilisation qui en sera faite dans cette étude.

Pour une valeur fixée de f , le modèle paramétrique à considérer est celui défini par (5.3). Les paramètres libres du modèle sont $(\tilde{d}[f], \omega_1[f], \dots, \omega_S[f])$ conformément à (2.24). À chaque itération, pour $\mathbf{k} = (), (1), \dots, (S)$, un couple d’indice (m, t) est tiré suivant des probabilités proportionnelles aux $\{\pi_{m,t,\mathbf{k}}\}_{m,t,\mathbf{k}}$ (une fois écartées celles inférieures à 0.1%). Ainsi, un ensemble minimal composé de S fenêtres candidates et permettant d’estimer un modèle acoustique préliminaire de la salle est constitué. Cet ensemble est utilisé pour calculer une estimation des paramètres du modèle

Algorithm 5.1 Approche inspirée de RANSAC

Entrée : $f, \mathcal{K}, N_{\text{iter}}, \{\tilde{x}_{m,t}[f], \eta_{m,\mathbf{k}}^*, \pi_{m,t,\mathbf{k}}\}_{m,t,\mathbf{k}}$

Sortie : $\{\mathcal{J}_{\mathbf{k},f}^{\text{ransac}}\}_{\mathbf{k} \in \mathcal{K}}$

```
1:  $\psi := -\infty$ ; // Score initial
2: for  $n = 1, 2, \dots, N_{\text{iter}}$  do
3:   Tirer  $(m, t)$  avec une probabilité proportionnelle à  $\pi_{m,t,()}$ ;
4:    $\hat{d}[f] := \frac{\tilde{x}_{m,t}[f]}{(\eta_{m,\mathbf{k}}^*)^2}$ ; // Son direct
5:   for  $k = 1, 2, \dots, S$  do
6:     Tirer  $(m, t)$  avec probabilité proportionnelle à  $\pi_{m,t,(k)}$ ;
7:      $\hat{\omega}_k[f] := \frac{\tilde{x}_{m,t}[f]}{\hat{d}[f](\eta_{m,\mathbf{k}}^*)^2}$ ; // Coefficient de réflexion
8:   end for
9:   for  $\mathbf{k} \in \mathcal{K}$  do
10:     $\hat{v}_{\mathbf{k}}[f] := \prod_{k \in \mathbf{k}} \hat{\omega}_k[f]$ ;
11:     $\hat{\mathcal{J}}_{\mathbf{k},f}^{\text{ransac}} := \left\{ (m, t) \mid \left| \frac{\tilde{x}_{m,t}[f] - (\eta_{m,\mathbf{k}}^*)^2 \hat{d} \cdot \hat{v}_{\mathbf{k}}[f]}{(\eta_{m,\mathbf{k}}^*)^2 \hat{d} \cdot \hat{v}_{\mathbf{k}}[f]} \right| < 0.1 \right\}$ ;
12:   end for
13:    $\hat{\psi} := \sum_{\mathbf{k} \in \mathcal{K}} \sum_{(m,t) \in \hat{\mathcal{J}}_{\mathbf{k},f}^{\text{ransac}}} (\eta_{m,\mathbf{k}}^*)^2$ ; // Calcul du score
14:   if  $\hat{\psi} > \psi$  then
15:      $\psi := \hat{\psi}$ ;
16:      $\forall \mathbf{k}, \mathcal{J}_{\mathbf{k},f}^{\text{ransac}} := \hat{\mathcal{J}}_{\mathbf{k},f}^{\text{ransac}}$ ;
17:   end if
18: end for
```

$(\hat{d}[f], \hat{\omega}_1[f], \dots, \hat{\omega}_S[f])$ conformément à (5.5). Ensuite, pour chaque $\mathbf{k} \in \mathcal{K}$, le sous-ensemble d'inliers $\hat{\mathcal{J}}_{\mathbf{k},f}^{\text{ransac}}$ est constitué à partir des couples (m, t) pour lesquels l'erreur relative entre $\tilde{x}_{m,t}[f]$ et le modèle acoustique estimé est inférieure à 10%. Le modèle estimé est évalué grâce à un score proportionnel au nombre d'inliers. Finalement, le modèle qui maximise ce score sur toutes les itérations de l'algorithme est sélectionné. Les sous-ensembles d'observations finaux $\{\mathcal{J}_{\mathbf{k},f}^{\text{ransac}}\}_{\mathbf{k} \in \mathcal{K}}$ sont constitués à partir des inliers du modèle retenu. Cette procédure est résumée sous forme la forme d'un pseudo-code par l'Algorithme 5.1.

Le score utilisé à la ligne 13 favorise les modèles qui fournissent une grande quantité d'inliers pour temps d'arrivée courts, grâce aux poids $(\eta_{m,\mathbf{k}}^*)^2 = (\bar{f}/c\tau_{m,\mathbf{k}}^*)^2$. Des études préliminaires ont montré que favoriser de la sorte des observations avec des temps d'arrivée courts permettait d'améliorer les résultats. Ceci est principalement dû au fait que la densité des échos est plus faible dans les parties précoces des RISs (Abel et al., 2006), ce qui limite le risque d'interférences entre les échos. Cela contribue également à atténuer l'impact du bruit de mesure, car le rapport signal/bruit (SNR) des RISs tend à diminuer au cours du temps. L'ensemble de fenêtres $\mathcal{J}_{\mathbf{k},f}^{\text{ransac}}$ ainsi obtenu peut alors être utilisé comme remplacement de $\mathcal{J}_{\mathbf{k}}^*$ dans (5.5). Notons que $\mathcal{J}_{\mathbf{k},f}^{\text{ransac}}$ peut différer en fonction de la fréquence considérée.

5.3 Expériences et Résultats

5.3.1 Cadre expérimental

Les expériences présentées dans cette section ont pour objectif d'évaluer les performances de la méthode d'estimation des profils d'absorption par minimisation du critère (5.5) sur les sous-ensemble de fenêtres de spectrogrammes définis par \mathcal{J}_k^* , $\mathcal{J}_k^{\text{top}}$ et $\mathcal{J}_k^{\text{ransac}}$.

Ces expériences sont menées sur un ensemble de RISs synthétiques simulées pour des salles en forme de pavé droit ($S = 6$ murs) en utilisant le simulateur d'acoustique de salle ROOMSIM (Schimmel et al., 2009). Il est important de noter que la méthode proposée peut être étendue à des formes de salles arbitraires tant qu'il est possible d'appliquer la méthode des sources-images et de considérer les occlusions acoustiques. Un total de 500 salles sont simulées, avec des dimensions échantillonnées uniformément dans l'intervalle $[3, 10] \times [3, 10] \times [2, 5]$ mètres (longueur, largeur et hauteur). Chaque salle simulée comporte 5 sources et 5 microphones donnant lieu à un maximum de $M = 25$ RISs exploitables. Les sources et microphones sont positionnés aléatoirement dans la salle en respectant une distance minimale aux murs de 1 mètre et une distance minimale source-microphone de 1 mètre. Les profils d'absorption des murs, définis par 6 coefficients associés aux bandes d'octave espacées logarithmiquement centrées aux fréquences .125, .250, ..., 4 kHz, sont générés tirés aléatoirement de sorte à reproduire des acoustiques de salle réalistes et variées.

Les RISs simulées sont tronquées à l'instant $T = 250$ ms, échantillonnées à la fréquence $\bar{f} = 16$ kHz, et incluent les échos jusqu'à l'ordre de réflexion $Q_{\text{sim}} = 20$. Des réponses idéales de sources et de microphones sont considérées pour la simulation des RISs. Cette hypothèse revient à considérer un système parfaitement calibré où les réponses des sources et des microphones ont été parfaitement compensées pour toutes les directions d'arrivée. La réponse source-microphone d correspondante est donc une distribution Dirac. Bien que son estimation soit cruciale pour estimer correctement les profils de réflexion $\omega_1, \dots, \omega_6$ par (5.5), les métriques d'évaluation utilisées portent uniquement sur l'estimation des profils d'absorption, paramètres cibles dans cette étude. Pour émuler des scénarios du monde réel, un bruit blanc gaussien est ajouté aux RISs de sorte à atteindre un rapport signal à bruit maximal fixe (PSNR pour Peak Signal to Noise Ratio en anglais). Afin de modéliser l'incertitude géométrique, les paramètres géométriques exacts (positions des sources, des microphones et des murs) sont affectés d'un bruit gaussien de moyenne nulle et d'écart-type σ_{geo} avant d'être utilisés pour calculer les temps d'arrivée erronés $\{\tau_{m,k}^*\}_{m,k}$ par la méthode des sources-images. Dans toutes les expériences, la valeur de σ_{geo} est supposée connue et est utilisée pour calculer les ensembles $\mathcal{J}_k^{\text{top}}$ et $\mathcal{J}_k^{\text{ransac}}$. Cela revient à considérer que la précision de l'appareil utilisé pour mesurer la géométrie est connue.

Les spectrogrammes des RISs sont calculés avec une fenêtre temporelle glissante de taille $2F = 32$ échantillons (2 ms) et un pas de 1 échantillon. La longueur supposées des échos est $L_v = 8$

échantillons (0.5 ms). L'ordre maximal des échos pris en compte pour le calcul des sous-ensembles de fenêtres de spectrogrammes est $Q_{\max} = 2$. Il a été constaté lors d'expériences préliminaires que modéliser explicitement des échos d'ordre supérieur a tendance à dégrader les résultats. Pour correspondre à la résolution fréquentielle des spectrogrammes, les profils d'absorption $\alpha[f] = 1 - \omega[f]$ pour $f = 0, \dots, F$ sont obtenus par interpolation linéaire des profils d'absorption en bande d'octave à la même manière du simulateur ROOMSIM. Les bandes de fréquences correspondantes espacées linéairement sont approximativement centrées autour des fréquences 0 Hz, 470 Hz, \dots , 8 kHz. Des expériences préliminaires ont révélé qu'aucune des approches proposées n'était capable d'estimer correctement les valeurs d'absorption dans la bande de fréquence la plus basse (DC), produisant des estimations proches du hasard. Par conséquent, cette bande de fréquences est omise dans les résultats suivants. Cette limitation peut s'expliquer par la taille relativement courte des fenêtres de Fourier utilisées pour le calcul des spectrogrammes. D'autre part, il a été observé que l'augmentation de la taille de ces fenêtres entraîne une diminution du nombre de fenêtres contenant des échos isolés provoquant alors une dégradation des résultats.

Deux métriques sont utilisées pour évaluer l'estimation des coefficients d'absorption sur toutes les fréquences, les murs et les pièces : *l'erreur absolue moyenne* (EAM) et le *pourcentage total de paramètres correctement estimés* (CE%), c'est-à-dire avec une erreur inférieure à 0.1 (rappelons que les coefficients d'absorption prennent des valeurs entre 0 et 1). Ce seuil d'acceptabilité est arbitraire mais vise à refléter des exigences viables à des fins de diagnostic acoustique.

5.3.2 Performances de la méthode

La première étude compare les résultats obtenus par la minimisation du critère (5.5) sur chacun des ensembles $\mathcal{J}_{\mathbf{k}}^*$, $\mathcal{J}_{\mathbf{k}}^{\text{top}}$ et $\mathcal{J}_{\mathbf{k}}^{\text{ransac}}$. Ces résultats sont présentés dans le tableau 5.1 pour $\sigma_{\text{geo}} \in \{0, 2\}$ cm (écart-type de l'erreur sur la mesure des paramètres géométriques) et $Q \in \{1, 2\}$ (ordre maximal des sources-images dont les fenêtres sont considérées pour l'estimation par (5.5)). Pour ces expériences, le bruit de fond est tel que le PSNR des RISs est égal à 50 dB et l'ordre maximal de modélisation des sources-images est fixé à $Q_{\max} = 2$. Les métriques EAM et CE% sont calculées sur les 500 salles simulées, avec l'estimation des profils d'absorption effectuée en utilisant 3 sources et 3 microphones par salle (M=9 RISs). Notons que dans certains cas rares (moins de 1% des cas pour $\mathcal{J}_{\mathbf{k}}^{\text{ransac}}$ calculé avec 4 RISs ou plus), un sous-ensemble $\mathcal{J}_{(k)}$ peut être vide, rendant alors impossible ou du moins mauvaise l'estimation de α_k . Dans de tels situations, son estimation $\hat{\alpha}_k[f]$ est arbitrairement fixée à la valeur médiane 0.5.

La première ligne du tableau 5.1 montre les résultats obtenus en utilisant $\mathcal{J}_{\mathbf{k}}^*$ et en supposant que la géométrie est parfaitement connue, c'est-à-dire $\sigma_{\text{geo}} = 0$ cm. Dans cette situation, $\mathcal{J}_{\mathbf{k}}^*$ n'est affecté par aucune erreur géométrique : $\mathcal{J}_{\mathbf{k}}^* = \mathcal{J}_{\mathbf{k}}^*$ (voir équation (5.2)). En utilisant uniquement les échos d'ordre inférieur à 1 pour la minimisation de (5.5) (c'est-à-dire $Q = 1$), nous obtenons

Sous-ensemble	σ_{geo}	$Q = 1$		$Q = 2$	
		EAM	CE (%)	EAM	CE (%)
\mathcal{J}_k^*	0 cm	0.095	79.7	0.128	71.1
	2 cm	0.108	74.7	0.132	69.6
$\mathcal{J}_k^{\text{top}}$	2 cm	0.128	71.7	0.119	69.7
$\mathcal{J}_k^{\text{ransac}}$	2 cm	0.088	82.9	0.082	84.6

TABLEAU 5.1 – Erreur absolue moyenne (EAM) et pourcentage des estimations correctes (CE%) obtenus en minimisant (5.5) sur différents sous-ensembles de fenêtres de RISs. L'écart-type de l'erreur sur la mesure des paramètres géométriques est $\sigma_{\text{geo}} \in \{0, 2\}$ cm. L'ordre maximal des échos dont les fenêtres sont considérées pour l'estimation est $Q \in \{1, 2\}$. L'ordre maximal de modélisation des sources-images est $Q_{\text{max}} = 2$. Le PSNR des RISs est fixé à 50dB.

une EAM légèrement inférieure à 0.1 et près de 80% des coefficients d'absorption correctement estimés. Ce résultat démontre la viabilité du critère d'optimisation proposé en (5.5). Bien que ces résultats soient satisfaisant, l'erreur d'estimation s'explique notamment par le fait qu'il survient (peu fréquemment) des cas où des fenêtres pertinentes associées à des sources-images d'ordre inférieur à $Q = 1$ sont affectées par des échos non modélisés car d'ordre strictement supérieur à $Q_{\text{max}} = 2$ (ordre 3 ou plus). Cependant, et comme attendu, les deux métriques se détériorent considérablement lorsque des perturbations géométriques ($\sigma_{\text{geo}} = 2$ cm) sont ajoutées. De plus, dans les deux cas, la méthode n'est pas en mesure d'exploiter les échos d'ordre 2 ($Q = 2$) qui contribuent à détériorer davantage les résultats. Cela est dû au fait que des échos d'ordre supérieur ont plus de chances d'interférer avec ceux d'ordre 2, ce qui ne peut pas être pris en compte par cette approche directe et déterministe. En utilisant $\mathcal{J}_k^{\text{top}}$, c'est-à-dire uniquement les fenêtres de probabilité de pertinence maximale pour chaque écho (5.11), les résultats s'améliorent légèrement lorsque $Q = 2$, mais se détériorent pour $Q = 1$ en raison de la réduction des observations disponibles. Enfin, la solution proposée basée sur l'algorithme RANSAC et prenant en compte l'incertitude sur la géométrie surpasse les deux autres dans toutes les configurations, réduisant ainsi l'EAM à près de 0.08 et augmentant le pourcentage d'estimations correctes à près de 85% en utilisant des fenêtres jusqu'à $Q = 2$. Ces résultats démontrent la capacité de l'approche de sélection robuste proposée à exclure avec succès les observations affectées par des interférences avec des échos non modélisés ou par des valeurs extrêmes du bruit de mesure. La figure 5.4 montre un exemple de résultats obtenus avec cette méthode.

5.3.3 Influence du nombre d'observations

La seconde étude se concentre sur la configuration ayant obtenu les meilleures performances pour la première étude, à savoir l'utilisation des ensembles $\mathcal{J}_k^{\text{ransac}}$ pour les échos d'ordre inférieur ou égal à $Q = 2$. Cette étude évalue conjointement l'influence sur les performances du nombre d'observations utilisées ainsi que la robustesse au bruit de mesure et aux erreurs géométriques. La

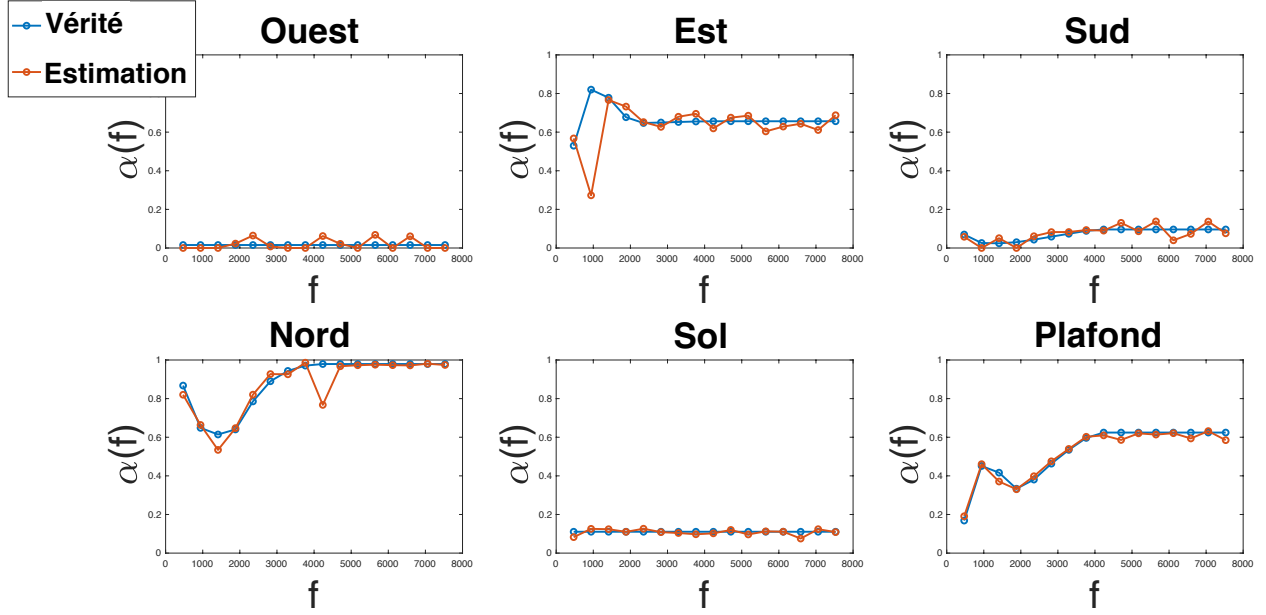


FIGURE 5.3 – Exemple d’estimation des profils d’absorption des 6 murs d’une salle en forme de pavé droit ($S = 6$ murs) avec la méthode de sélection robuste basée sur RANSAC. L’estimation est effectuée en minimisant (5.5) sur les ensembles $\mathcal{J}_k^{\text{ransac}}$ constitués à partir de $M = 9$ RISs (3 sources et 3 microphones). Les RISs présentent un PSNR égal à 50dB et l’incertitude géométrique un écart-type σ_{geo} égale à 2cm. L’ordre maximal de modélisation Q_{max} , c’est-à-dire considérée pour identifier les fenêtres pertinentes est égal à 2. L’ordre Q des sources-images associées aux fenêtres pertinentes utilisées pour l’estimation est égal à 2.

figure 5.4 présente l’EAM et le CE% obtenus sur l’ensemble des 500 salles simulées pour l’utilisation d’un nombre variable de sources et de microphones (et donc de RISs), pour $\text{PSNR} \in \{60, 50\}$ dB et $\sigma_{\text{geo}} \in \{2, 5\}$ cm. Comme prévu, la figure 5.4 révèle que la réduction du PSNR et l’augmentation de σ_{geo} dégradent systématiquement les performances. En revanche, l’augmentation du nombre de RISs disponibles par salle améliore systématiquement les métriques EAM et CE%, finissant même par de compenser la dégradation.

Ces résultats suggèrent que l’approche proposée réussit à sélectionner les fenêtres de RISs les plus pertinentes malgré leur nombre croissant. Notons qu’avec un PSNR de 50 dB et $\sigma_{\text{geo}} = 2$ cm, l’utilisation de seulement 4 RISs (2 sources et 2 microphones) suffit pour atteindre un EAM satisfaisant proche de 0.1, avec 78% d’estimations correctes. De manière intéressante, des résultats similaires sont obtenus pour une configuration légèrement plus pratique, composée de 1 source et de 5 microphones.

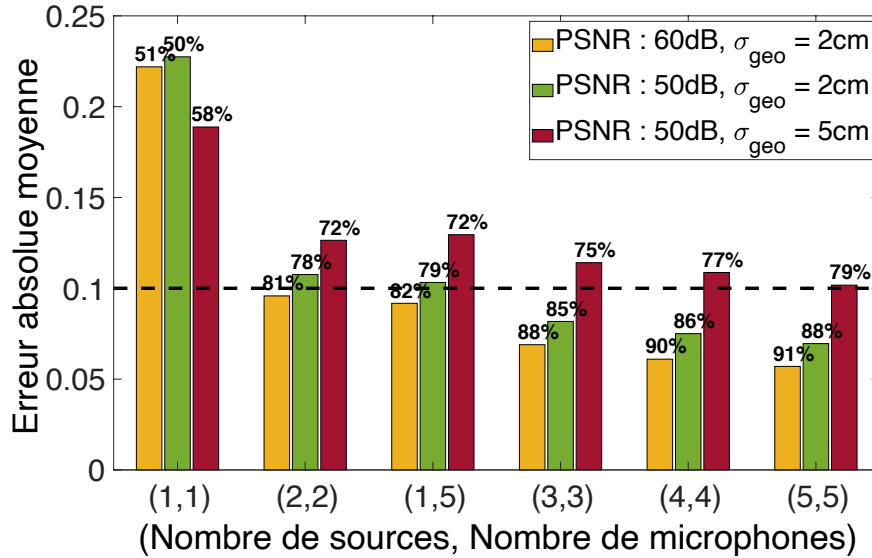


FIGURE 5.4 – Erreur absolue moyenne (EAM) et pourcentage des estimations correctes (CE) obtenus avec la procédure de sélection robuste basée sur RANSAC et $Q = 2$ pour différents nombres de sources et microphones et différents niveau de bruit de mesure et d’incertitude géométrique

5.4 Conclusion

Dans ce chapitre, le problème inverse d’estimation des profils d’absorption des murs d’une salle à partir de plusieurs réponses impulsionnelles a été étudié. Cette estimation a été réalisée à partir des spectrogrammes des réponses impulsionnelles et de la géométrie. À chaque fréquence, une fonction objectif définie dans le domaine des amplitudes carrées est optimisée sur un sous-ensemble de fenêtres temporelles des spectrogrammes dites pertinentes. Ces fenêtres pertinentes, supposées inclure entièrement et isoler une écho acoustique unique, sont sélectionnées grâce à une approche guidée par la géométrie. Une étude simulée a montré que cette approche effectue des estimations précises des coefficients d’absorption au-dessus de 400 Hz, ce en présence d’erreurs géométriques et de bruit de mesure. Il a été démontré que la version probabiliste et robuste de cette approche basée sur l’algorithme RANSAC permet d’améliorer l’estimation en exploitant de l’information des échos jusqu’au second ordre de réflexion. Une expérience étudiant l’influence du nombre de sources et de microphones sur les erreurs commises a démontré que l’augmentation du nombre de réponses impulsionnelles utilisées réduit significativement les erreurs d’estimation et améliore la robustesse de l’approche. À noter que dans l’étude, les coefficients d’absorption ont été supposés indépendants de l’angle d’incidence des réflexions et uniformes sur l’ensemble des murs et que seuls des dispositifs omnidirectionnels parfaitement calibrés ont été considérés. Néanmoins, il semble raisonnable d’envisager des extensions de l’approche présentée pour relâcher ces hypothèses ou exploiter des connaissances externes sur les dispositifs et les murs.

Le code permettant de reproduire les résultats présenté dans ce chapitre est disponible à l’adresse

<https://github.com/stedlg/rir2abs>.

La limite principale de cette approche réside dans le fait qu'elle repose sur l'absence de superposition entre les échos acoustiques au sein des réponses impulsionnelles. En effet, dans des applications réalistes, les échos sont d'une durée telle qu'il n'est pas envisageable d'inclure entièrement et d'isoler dans une fenêtre temporelle ne serait-ce qu'un seul écho. Cela est principalement lié à l'effet des sources et des microphones dont les réponses temporelles sont longues car dépendent de la fréquence. La figure 2.10c montre un exemple du niveau de superposition entre les échos lorsque de telles réponses d'appareils sont utilisées pour la simulation des RISs. De plus, le critère d'optimisation (5.5) considère une réponse source-microphone identique pour toutes les directions de l'espace ce qui ne permet pas de l'appliquer directement au cas général de sources et microphones directionnels. Enfin, une autre limite de cette approche est qu'elle considère indépendamment les fréquences pour l'estimation des coefficients d'absorption.

Le chapitre suivant s'attaque donc à la résolution du problème inverse pour un niveau de réalisme élevé. Des mesures réelles de réponses de sources et de microphones, qui dépendent de la direction et de la fréquence, sont utilisées pour la simulation des RISs. La méthode qui sera présentée considère l'estimation des réponses temporelles des murs par une approche d'optimisation dans le domaine temporel.

Approche par optimisation dans le domaine temporel

Préambule : Ce chapitre présente une approche d’optimisation pour l’estimation des réponses impulsionnelles des parois d’une salle, à partir de réponses impulsionnelles de salle, des paramètres de la géométrie connus aux erreurs de mesure près et des mesures des réponses des sources et des microphones utilisés. L’estimation est réalisée dans le domaine temporel en définissant deux sous-problèmes d’optimisation complémentaires. Le premier porte sur l’estimation des erreurs induites par l’incertitude géométrique sur les temps d’arrivée des échos calculés avec la méthode des sources-images en optimisant sous contraintes un critère défini dans le domaine temporel continu avec l’algorithme ADAM (Kingma et al., 2014). Ce critère est défini à partir d’une extension du modèle de réponse impulsionnelle de salle issu de la méthode des sources-images permettant de rendre compte de ces délais sur les temps d’arrivée des échos et suppose initialement des parois parfaitement rigides. Le second sous-problème s’intéresse à l’estimation des réponses impulsionnelles des parois de la salle, formulée comme un problème d’optimisation linéaire et multi-observation dans le domaine temporel, avec la méthode du gradient conjugué (Hestenes et al., 1952).

6.1 Introduction

Dans le chapitre précédent, l’estimation des profils d’absorption des murs d’une salle était accomplie à l’aide d’une approche par optimisation sur les spectrogrammes de RISs, informée par la géométrie. L’approche alors présentée reposait sur la sélection des fenêtres temporelles des spectrogrammes incluant entièrement une unique écho. Cette approche a permis d’estimer efficacement les paramètres d’intérêt lors d’expériences conduites sur des données simulées en considérant des réponses de sources et de microphones idéales et en modélisant l’incertitude sur la mesure des paramètres géométriques. Les échos au sein des RISs synthétiques correspondantes sont de très courtes durées et certains d’entre eux sont isolés ce qui permet d’envisager leur extraction par un fenêtrage temporel approprié. Cependant, comme évoqué en section 2.4 et illustré en figure 2.10c, dans le cas général de sources et de microphones réalistes aux réponses à dépendance directionnelle et fréquentielle, les échos peuvent être de durées telles qu’il devient vain d’appliquer ce type d’approche. Suite aux conclusions tirées à l’issue des chapitres 4 et 5, il semble donc intéressant d’opter

pour une méthode d’optimisation, robuste aux erreurs géométriques et permettant de répondre à l’ensemble des tâches inhérentes au problème posé, c’est-à-dire la localisation, la séparation et la déconvolution des échos mélangés dans les RISs.

Pour répondre à cette problématique, nous proposons dans ce chapitre une nouvelle approche pour l’estimation des réponses impulsionnelles des murs (RIMs) d’une salle à partir de RISs, de mesures des réponses des sources et des microphones ainsi que des paramètres géométriques aux erreurs près de l’ordre de quelques centimètres. La RIM constitue une description précise de ses propriétés acoustiques rendant compte de son effet sur l’énergie et sur la phase du son incident. Elle donne accès à une richesse d’information supérieure au profil de réflexion du mur (information équivalente au profil d’absorption) qui décrit son effet sur l’énergie du son incident uniquement.

L’approche proposée est basée sur une extension du modèle temporel de RIS donné par la méthode des sources-images. Ce modèle étendu permet de rendre compte des délais entre les temps d’arrivée erronés et exacts des échos induits par les erreurs sur la mesure de la géométrie. Ce nouveau modèle permet de considérer deux étapes consécutives pour la résolution du problème inverse d’intérêt. La première étape consiste à estimer conjointement les délais fractionnaires en optimisant sous contraintes un critère dans le domaine temporel continu défini à partir du modèle étendu avec l’algorithme de descente de gradient *ADAM* (Kingma et al., 2014). Les délais estimés sont alors utilisés pour rectifier les instants d’arrivée et les facteurs d’atténuation géométrique des échos (*délais correctifs*). La seconde étape porte sur l’estimation conjointe des RIMs, formulée comme un problème d’optimisation linéaire et multi-observation dans le domaine temporel, avec la méthode du gradient conjugué (Hestenes et al., 1952). La matrice du système linéaire renseigne les localisations et les réponses source-microphone associées aux échos dans la RIS. Lorsque l’information de localisation contenue dans cette matrice est incorrecte, l’information extraite du signal ne correspond pas à celle recherchée. La première étape vise donc à corriger les erreurs sur l’information de localisation des échos portée par la matrice du système linéaire. La seconde étape correspond à l’inversion du système linéaire une fois la matrice modifiée.

Cette approche par optimisation définie entièrement dans le domaine temporel présente l’atout de conserver sa validité sur toute restriction temporelle des signaux considérés. Il est alors possible de considérer uniquement les premiers instants des RISs pour la résolution du problème d’intérêt. D’une part, cela présente l’avantage de focaliser l’approche sur une partie des RISs pour laquelle le modèle issu de la méthode des sources-images est reconnu comme étant une bonne approximation. D’autre part, cela permet de réduire de manière drastique la taille du système en le limitant aux premières millisecondes des RISs ainsi qu’aux sources-images qui y sont visibles. L’approche est implémentée exclusivement avec *Pytorch* (Paszke et al., 2019), la librairie open source de référence pour l’apprentissage automatique en Python. Elle permet une exécution accélérée des calculs grâce à un large éventail de fonctions et algorithmes optimisés pour la parallélisation sur plusieurs cartes graphiques.

Une étude détaillée des performances de l’approche est effectuée à partir d’expériences conduites dans un cadre simulé réaliste utilisant des données générées à partir de mesures réelles de réponses en direction et en fréquence de sources et de microphones (*hybrides*), et modélisant les incertitudes géométriques. Les superpositions entre les échos sont omniprésentes au sein des réponses impulsionnelles de salle générées. Les sources et microphones utilisés sont positionnés arbitrairement dans la salle.

Les expériences conduites sur les RISs simulées révèlent que la méthode du gradient conjugué permet d’estimer les RIMs avec une excellente précision dans le cas où la géométrie est parfaitement connue. Cependant, celle-ci s’effondre en cas d’erreurs géométriques mêmes faibles. L’étape préalable d’estimation et de compensation des erreurs commises sur le calcul des temps d’arrivées des échos permet de pallier à cet écueil. L’approche par optimisation proposée permet ainsi d’obtenir des résultats d’estimation des RIMs satisfaisants, même en cas d’erreurs géométriques de l’ordre de 2cm.

Le chapitre est organisé comme suit. La Section 6.2 décrit l’extension du modèle temporel de RIS issu de la méthode des sources-images et la formulation du problème d’optimisation en deux étapes. La Section 6.3 présente une étude par *ablation* afin d’obtenir une compréhension générale et détaillée du comportement de la méthode proposée ainsi que des résultats obtenus. La Section 6.4 apporte une conclusion et met en avant des perspectives de travaux possibles. La Section 6.5 discute de certains aspects de la méthode et de tentatives préliminaires.

6.2 Formulation du problème

6.2.1 Modélisation du signal

Considérons une salle de géométrie connue et compatible avec l’utilisation de la méthode des sources-images (présentée en section 2.3) délimitée par S murs ainsi que M RISs mesurées dans cette salle. Pour $m \in \llbracket 1, M \rrbracket$, soit x_m une RIS. Considérons son modèle issu de la méthode des sources-images (2.19) s’écrivant :

$$x_m[t] = \sum_{\mathbf{k} \in \mathcal{K}} \left(\frac{\bar{f}}{4\pi c \tau_{m,\mathbf{k}}} \cdot \phi[\cdot - \tau_{m,\mathbf{k}}] * g_m^{\text{mic}}[\theta_{m,\mathbf{k}}, \psi_{m,\mathbf{k}}] * g_m^{\text{src}}[-\theta_{m,\mathbf{k}}, -\psi_{m,\mathbf{k}}] * v_{\mathbf{k}} \right) [t]. \quad (6.1)$$

La constante \bar{f} correspond à la fréquence d’échantillonnage du signal. La constante c désigne la vitesse du son. Les paramètres $\{\tau_{m,\mathbf{k}}\}$ désignent les temps d’arrivée des sources images de \mathcal{K} , c’est-à-dire incluses dans le modèle de la RIS x_m . Les paramètres $\{\theta_{m,\mathbf{k}}\}$ et $\{\psi_{m,\mathbf{k}}\}$ sont les angles

d'azimut et d'élévation des sources-images de \mathcal{K} par rapport au microphone. Les fonctions g_m^{src} et g_m^{mic} représentent respectivement les réponses de la source et du microphone utilisés pour obtenir la RIS x_m .

Comme évoqué en section 2.3, la méthode des sources-images utilise les paramètres géométriques associées à une RIS (emplacements des murs, de la source et du microphone) pour déterminer les positions correspondantes des sources-images. Ces positions permettent de calculer les temps d'arrivée $\{\tau_{m,\mathbf{k}}\}$ ainsi que les angles d'azimut $\{\theta_{m,\mathbf{k}}\}$ et d'élévation $\{\phi_{m,\mathbf{k}}\}$ des sources-images figurant dans le modèle (6.1). Cependant, dans la réalité, des erreurs sur la mesure des paramètres géométriques surviennent inévitablement et se répercutent sur le calcul de ces paramètres associés aux sources-images. Notons alors $\{\tau_{m,\mathbf{k}}^*\}$, $\{\theta_{m,\mathbf{k}}^*\}$ et $\{\phi_{m,\mathbf{k}}^*\}$ respectivement les temps d'arrivée, angles d'azimut et angles d'élévation des sources-images calculés à partir des paramètres géométriques erronés car affectés par les erreurs de mesure. Il s'agit des données du problème tandis que les paramètres $\{\tau_{m,\mathbf{k}}\}$, $\{\theta_{m,\mathbf{k}}\}$ et $\{\phi_{m,\mathbf{k}}\}$ sont des inconnues du problème.

Dans cette étude, nous faisons l'hypothèse que le niveau d'incertitude géométrique considéré entraîne une variation négligeable des réponses source-microphone qui s'appliquent à chaque source-image, c'est-à-dire :

$$\begin{aligned} g_m^{\text{mic}}[\theta_{m,\mathbf{k}}^*, \psi_{m,\mathbf{k}}^*] &\simeq g_m^{\text{mic}}[\theta_{m,\mathbf{k}}, \psi_{m,\mathbf{k}}] \\ \text{et } g_m^{\text{src}}[-\theta_{m,\mathbf{k}}^*, -\psi_{m,\mathbf{k}}^*] &\simeq g_m^{\text{src}}[-\theta_{m,\mathbf{k}}, -\psi_{m,\mathbf{k}}]. \end{aligned} \quad (6.2)$$

Cette hypothèse semble réaliste dans le cas général. Cependant, comme constaté avec les chapitres 4 et 5 et illustré par la figure 2.10b, des erreurs géométriques de l'ordre de quelques centimètres produisent un effet significatif sur le calcul des temps d'arrivée des échos. Pour tout $\mathbf{k} \in \mathcal{K}$, notons $\Delta_{m,\mathbf{k}}$ l'erreur commise sur le temps d'arrivée exact $\tau_{m,\mathbf{k}}$ en raison des erreurs géométriques :

$$\forall m \in [1, M], \forall \mathbf{k} \in \mathcal{K}, \quad \tau_{m,\mathbf{k}}^* = \tau_{m,\mathbf{k}} + \Delta_{m,\mathbf{k}}, \quad \Delta_{m,\mathbf{k}} \in \mathbb{R}. \quad (6.3)$$

Nous proposons donc une nouvelle extension du modèle de RIS issu de la méthode des sources-images permettant de rendre compte de l'effet des incertitudes géométriques. Ce nouveau modèle intègre pour chaque source-image $\mathbf{k} \in \mathcal{K}$ une fonction du temps notée $c_{m,\mathbf{k}}(\Delta_{\mathbf{k}})$ qui permet compenser l'impact de l'erreur géométrique sur le temps d'arrivée et le facteur d'atténuation géométrique grâce au délai correctif $\Delta_{\mathbf{k}}$:

$$x_m[t] = \sum_{\mathbf{k} \in \mathcal{K}} \left(\underbrace{\frac{\bar{f}}{4\pi c \tau_{m,\mathbf{k}}^*}}_{\eta_{m,\mathbf{k}}^*} \cdot \underbrace{\phi[\cdot - \tau_{m,\mathbf{k}}^*]}_{\phi_{m,\mathbf{k}}^*} * \underbrace{g_m^{\text{mic}}[\theta_{m,\mathbf{k}}, \psi_{m,\mathbf{k}}] * g_m^{\text{src}}[-\theta_{m,\mathbf{k}}, -\psi_{m,\mathbf{k}}]}_{g_{m,\mathbf{k}}} * \underbrace{\left(\frac{\tau_{m,\mathbf{k}}^*}{\tau_{m,\mathbf{k}}^* - \Delta_{m,\mathbf{k}}} \phi[\cdot + \Delta_{m,\mathbf{k}}] \right)}_{c_{m,\mathbf{k}}(\Delta_{m,\mathbf{k}}) : \text{correction}} * v_{\mathbf{k}} \right) [t]. \quad (6.4)$$

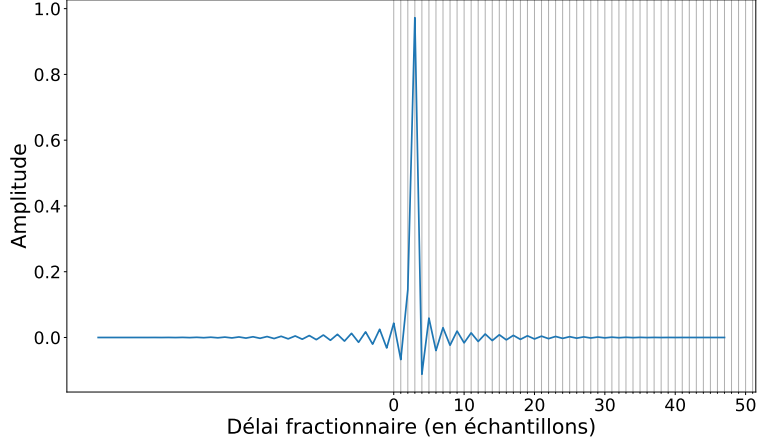


FIGURE 6.1 – Sinus cardinal fenêtré $\phi[\cdot - \Delta_{m,\mathbf{k}}]$ pour un délai fractionnaire échantillonné à la fréquence $\bar{f} = 16\text{kHz}$ et un délai fractionnaire $\Delta_{m,\mathbf{k}} = 0.18\text{ ms}$, équivalent à 2.87 échantillons.

Les parenthèses dans l'écriture de $c_{m,\mathbf{k}}(\Delta_{m,\mathbf{k}})$ précisent qu'il s'agit d'une fonction continue du délai $\Delta_{m,\mathbf{k}} \in \mathbb{R}$.

Dans cette étude, le filtre de délai fractionnaire ϕ considéré est le sinus cardinal fenêtré suivant :

$$\phi[t] = \begin{cases} \frac{1}{2} \left(1 + \cos \left[\frac{2\pi t}{T_w} \right] \right) \text{sinc}[t] & \text{si } -\frac{T_w}{2} \leq t \leq \frac{T_w}{2} \\ 0 & \text{sinon.} \end{cases} \quad (6.5)$$

Son support est de taille $2 * T_w + 1$. La figure 6.1 représente $\phi[t - \Delta_{m,\mathbf{k}}]$ pour $\bar{f} = 16\text{kHz}$ et $\Delta_{m,\mathbf{k}} = 0.18\text{ ms}$ et $T_w = 40$ échantillons.

6.2.2 Écriture matricielle du modèle

L'écriture matricielle du modèle permet d'aider à la définition et à la compréhension des approches présentées dans ce chapitre. Cependant, elle ne correspond pas à l'implémentation faite des algorithmes utilisés. En effet, elle fait apparaître des matrices parcimonieuses aux très grandes dimensions pour lesquelles le calcul matriciel direct nécessiterait des ressources computationnelles démesurées. En pratique, les algorithmes sont implémentés avec des fonctions Pytorch optimisées et compatibles pour le calcul parallèle sur cartes graphiques, notamment avec les fonctions de convolution du module `torch.nn.functional`. Ils sont utilisés sur des matrices denses et de tailles modérées.

L'écriture matricielle du modèle (6.4) est obtenue par :

$$x_m[t] = \sum_{\mathbf{k} \in \mathcal{K}} \left(\underbrace{\eta_{m,\mathbf{k}}^* \cdot \phi_{m,\mathbf{k}}^* * g_{m,\mathbf{k}} * c_{m,\mathbf{k}}(\Delta_{m,\mathbf{k}})}_{h_{m,\mathbf{k}}^*} * v_{\mathbf{k}} \right) [t]$$

$$\Rightarrow \mathbf{x}_m = \sum_{\mathbf{k} \in \mathcal{K}} \mathbf{H}_{m,\mathbf{k}}^* \mathbf{C}_{m,\mathbf{k}}(\Delta_{m,\mathbf{k}}) \mathbf{v}_{\mathbf{k}} = \begin{bmatrix} x_{m,1} \\ x_{m,2} \\ \vdots \\ x_{m,T} \end{bmatrix} \in \mathbb{R}^T,$$

avec :

$$\mathbf{H}_{m,\mathbf{k}}^* = \begin{bmatrix} h_{m,\mathbf{k},1}^* & \cdots & 0 \\ \vdots & \ddots & \vdots \\ h_{m,\mathbf{k},T'}^* & \cdots & h_{m,\mathbf{k},1}^* \\ \vdots & \ddots & \vdots \\ h_{m,\mathbf{k},T}^* & \cdots & h_{m,\mathbf{k},T-T'+1}^* \end{bmatrix} \in \mathbb{R}^{T,T'} \text{ matrice de Toeplitz,}$$

$$\mathbf{C}_{m,\mathbf{k}}(\Delta_{m,\mathbf{k}}) = \begin{bmatrix} c_{m,\mathbf{k}}(\Delta_{m,\mathbf{k}})_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ c_{m,\mathbf{k}}(\Delta_{m,\mathbf{k}})_L & \cdots & c_{m,\mathbf{k},1}(\Delta_{m,\mathbf{k}})_1 \\ \vdots & \ddots & \vdots \\ c_{m,\mathbf{k}}(\Delta_{m,\mathbf{k}})_{T'} & \cdots & c_{m,\mathbf{k}}(\Delta_{m,\mathbf{k}})_{T'-L+1} \end{bmatrix} \in \mathbb{R}^{T',L} \text{ matrice de Toeplitz,}$$

$$\mathbf{v}_{\mathbf{k}} = \begin{bmatrix} v_{\mathbf{k},1} \\ v_{\mathbf{k},2} \\ \vdots \\ v_{\mathbf{k},L} \end{bmatrix} \in \mathbb{R}^L,$$

ainsi que T et L les durées (en échantillons) de la RIS x_m et RIMs cumulées $\{\mathbf{v}_{\mathbf{k}}\}$, et T' la taille du support des fonctions de correction ($T' = 2 * T_w + 1 + \lfloor \bar{f} * \max(\{\Delta_{m,\mathbf{k}}\}) \rfloor$).

L'écriture matricielle du modèle (6.4) pour la RIS x_m peut donc s'écrire :

$$\mathbf{x}_m = \mathbf{H}_m^* \mathbf{C}_m(\Delta_m) \mathbf{v}, \quad (6.6)$$

avec $\Delta_m = \{\Delta_{m,\mathbf{k}}\}_{\mathbf{k} \in \mathcal{K}}$,

$$\begin{aligned}
\mathbf{H}_m^* &= \left[\mathbf{H}_{m,\mathbf{k}}^* \right]_{\mathbf{k} \in \mathcal{K}} \in \mathbb{R}^{T \times KT'}, \\
\mathbf{C}_m(\boldsymbol{\Delta}_m) &= \text{diag}_{KT',KL} \left(\left[\mathbf{C}_{m,\mathbf{k}}(\Delta_{m,\mathbf{k}}) \right]_{\mathbf{k} \in \mathcal{K}} \right) \in \mathbb{R}^{KT' \times KL}, \\
\mathbf{v} &= \left[\mathbf{v}_{\mathbf{k}} \right]_{\mathbf{k} \in \mathcal{K}} \in \mathbb{R}^{KL},
\end{aligned}$$

et $K = |\mathcal{K}|$, le nombre de sources-images incluses dans le modèle.

Enfin, le modèle matriciel *multi-observation* associant les M RISs est donné par :

$$\mathbf{x} = \mathbf{H}^* \mathbf{C}(\boldsymbol{\Delta}) \mathbf{v}, \quad (6.7)$$

avec $\boldsymbol{\Delta} = \{\Delta_{m,\mathbf{k}}\}_{m \in [1,M], \mathbf{k} \in \mathcal{K}}$,

$$\begin{aligned}
\mathbf{x} &= \begin{bmatrix} \mathbf{x}_1 \\ \mathbf{x}_2 \\ \vdots \\ \mathbf{x}_M \end{bmatrix} \in \mathbb{R}^{MT}, \\
\mathbf{H}^* &= \text{diag}_{MT, MKT'} \left(\left[\mathbf{H}_1^* \quad \cdots \quad \mathbf{H}_m^* \quad \cdots \quad \mathbf{H}_M^* \right] \right) \in \mathbb{R}^{MT, MKT'}, \\
\text{et } \mathbf{C}(\boldsymbol{\Delta}) &= \begin{bmatrix} \mathbf{C}_1(\Delta_1) \\ \vdots \\ \mathbf{C}_m(\Delta_m) \\ \vdots \\ \mathbf{C}_M(\Delta_m) \end{bmatrix} \in \mathbb{R}^{MKT', KL}.
\end{aligned}$$

6.2.3 Problème d'optimisation général

L'objectif est de définir à partir du modèle étendu présenté précédemment une approche d'optimisation permettant l'estimation des S RIMs de la salle. Pour ce faire, cette approche doit également permettre d'estimer les délais entre les temps d'arrivée erronés et exacts des échos dans les RISs induits par les erreurs sur la mesure des paramètres géométriques.

Rappelons que pour tout $\mathbf{k} \in \mathcal{K}$, $v_{\mathbf{k}}$ s'exprime en tant que convolutions multiples des RIMs $\{v_{(k)}\}_{k \in \mathbf{k}}$:

$$\begin{aligned}
\forall \mathbf{k} \in \mathcal{K}, v_{\mathbf{k}}[t] &= \left(\bigotimes_{k \in \mathbf{k}} v_{(k)} \right) [t] \\
&= \left(\bigotimes_{k \in \mathbf{k}} R_k \right) [t],
\end{aligned} \tag{6.8}$$

avec R_k la RIM numéro $k \in \llbracket 1, S \rrbracket$.

Il est donc possible de définir à partir du modèle multi-observation (6.7) le critère suivant pour la résolution du problème complet de correction des délais et d'estimation des RIMs de la salle :

$$\begin{aligned}
&\inf_{\Delta, \mathbf{v}_{(1)}, \dots, \mathbf{v}_{(S)}} \|\mathbf{x} - \mathbf{H}^* \mathbf{C}(\Delta) \mathbf{v}\|, \\
&\text{où } \mathbf{v} = \left[\mathbf{v}_{\mathbf{k}} \right]_{\mathbf{k} \in \mathcal{K}} \in \mathbb{R}^{KL}.
\end{aligned} \tag{6.9}$$

Il s'agit d'un critère non linéaire et non convexe, difficile à implémenter. Afin de simplifier le problème, nous considérons séparément l'optimisation conjointe des délais et l'optimisation conjointe des RIMs.

D'une part, lors de l'étape d'optimisation des délais correctifs Δ , v qui représente les RIMs cumulées est une inconnue du problème. Nous choisissons de fixer la variable v avec la valeur qui correspond à une salle dont les murs sont parfaitement rigides (c'est-à-dire parfaitement réfléchissants) :

$$\forall \mathbf{k} \in \mathcal{K}, v_{\mathbf{k}} = \boldsymbol{\delta}_0, \text{ avec } \boldsymbol{\delta}_0 = \begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix} \in \mathbb{R}^L. \tag{6.10}$$

Cette valeur par défaut présente plusieurs avantages. Premièrement, elle définit des RIMs qui vérifient l'hypothèse de phase minimale (voir 3.2.4). De plus, elle permet de modéliser exactement le son direct dont l'influence sur la partie précoce des RISs est prédominante, avec la réponse v_{\emptyset} associée à la source réelle.

Enfin, toujours à des fins de simplification, nous faisons l'hypothèse que les erreurs qui s'appliquent aux temps d'arrivée des échos sont indépendantes. Bien que ces erreurs soient vraisemblablement corrélées, cette hypothèse paraît viable devant la grande complexité des relations qui semblent les lier. L'estimation des délais est donc effectuée indépendamment pour chaque RIS en optimisant le critère mono-observation suivant :

$$\forall m \in \llbracket 1, M \rrbracket, \quad \inf_{\Delta_m, \mathbf{v}_k = \delta_0} \|\mathbf{x}_m - \mathbf{H}_m^* \mathbf{C}_m(\Delta_m) \mathbf{v}\|_2. \quad (6.11)$$

D'autre part, il est possible de constater que lorsque Δ est fixé, le critère (6.9) correspond à un problème d'optimisation linéaire multi-observation par rapport à la variable v . Ce système permet d'utiliser plusieurs RISs d'une même salle pour estimer conjointement les RIMs cumulées associées à chaque source-image, mais sans prendre en compte les relations non-linéaires entre RIMs comme dans (6.8). En particulier, il permet d'estimer les RIMs $\{v_{(k)}\}_{k \in \llbracket 1, S \rrbracket}$ par l'intermédiaire des sources-images du premier ordre. Le nombre de paramètres correspondant augmente considérablement, passant du nombre de murs S au nombre de sources-images modélisées K . Néanmoins, l'optimisation dans le domaine temporel permet de considérer uniquement les fenêtres temporelles des RISs suffisamment longues pour inclure les échos d'ordre 1, réduisant ainsi drastiquement la taille du modèle par rapport aux sources-images. Généralement, il est nécessaire de considérer seulement les quelques premières dizaines de millisecondes des RISs. De plus, ce critère donne accès au très large panel d'algorithmes d'optimisation de référence pour l'inversion des systèmes linéaires. L'estimation des RIMs est donc effectuée en minimisant un critère pouvant s'écrire sous la forme :

$$\inf_{\mathbf{v}} \|\mathbf{x} - \mathbf{H}^* \mathbf{C}(\Delta) \mathbf{v}\|. \quad (6.12)$$

Il est naturel de penser à alterner l'estimation des délais correctifs Δ avec le critère (6.11) d'une part et l'estimation des RIMs cumulées associées aux sources-images v avec le critère (6.12) d'autre part, en utilisant à chaque fois les derniers paramètres estimés, et ce jusqu'à la convergence de l'un des deux critères. Cependant, des expériences préliminaires menées en ce sens ont montré que l'alternance directe des ces deux problèmes d'optimisation conduit à dégrader significativement les résultats par rapport à l'enchaînement simple de (6.11) puis de (6.12). Ce constat est lié au fait que ces deux étapes permettent d'obtenir des estimations des S RIMs nettement meilleures que celles des RIMs cumulées pour les sources-images dont l'ordre est supérieur à 2. L'estimation de v correspondante est telle qu'elle conduit à des résultats pour l'estimation des délais avec (6.11) inférieurs à ceux obtenus en utilisant la valeur par défaut $\mathbf{v}_k = \delta_0$. En conséquence, nous considérons dans cette étude uniquement l'enchaînement de ces deux étapes pour la résolution du problème inverse d'intérêt. Des travaux futures pourront porter sur le développement d'une approche alternée utilisant uniquement les S RIMs estimées à l'issue de (6.12) pour construire la nouvelle valeur de \mathbf{v} utilisée par (6.11), comme, par exemple, à partir de convolutions multiples des S RIMs estimées, à la manière de (6.8).

6.2.4 Estimation des délais entre les temps d'arrivée erronés et exacts

Pour $m \in \llbracket 1, M \rrbracket$, l'estimation des délais entre les temps d'arrivée erronés et exacts des échos dans la RIS x_m est effectuée en considérant le critère (6.11) avec la norme L^2 :

$$\inf_{\Delta_m, \mathbf{v}_k = \delta_0} \|\mathbf{x}_m - \mathbf{H}_m^* \mathbf{C}_m(\Delta_m) \mathbf{v}\|_2. \quad (6.13)$$

En l'état, le critère (6.13) permet d'estimer pour chaque source-image des délais correctifs quelconques dans \mathbb{R} . Ceci ne reflète pas l'hypothèse d'erreurs géométriques de l'ordre de quelques centimètres, amplifie le risque de permutation temporelle des sources-images et pose problème pour la convergence de l'estimation. Par conséquent, nous introduisons $\Delta_{\max} > 0$, le délai correctif absolu maximal utilisé pour contraindre l'estimation de chaque délai dans l'intervalle $[-\Delta_{\max}, \Delta_{\max}]$. Nous définissons alors le problème d'optimisation sous contraintes suivant :

$$\begin{cases} \inf_{\Delta_m, \mathbf{v}_k = \delta_0} \|\mathbf{x}_m - \mathbf{H}_m^* \mathbf{C}_m(\Delta_m) \mathbf{v}\|_2 \\ \text{tel que : } \forall \mathbf{k} \in \mathcal{K}, \Delta_{m, \mathbf{k}} \in [-\Delta_{\max}, \Delta_{\max}]. \end{cases} \quad (6.14)$$

En pratique, cette contrainte est imposée aux délais correctifs estimés grâce au changement de variable :

$$\forall \mathbf{k} \in \mathcal{K}, \Delta_{m, \mathbf{k}} = \Delta_{\max} \cdot \tanh(u_{m, \mathbf{k}}), \text{ avec } \Delta_{\max} \in \mathbb{R}^+, u_{m, \mathbf{k}} \in \mathbb{R}. \quad (6.15)$$

L'optimisation des délais correctifs est ensuite effectuée par l'intermédiaire des variables $\mathbf{u}_m = \{u_{m, \mathbf{k}}\}_{m \in \llbracket 1, M \rrbracket, \mathbf{k} \in \mathcal{K}}$ suivant le critère :

$$\inf_{\mathbf{u}_m, \mathbf{v}_k = \delta_0} \|\mathbf{x}_m - \mathbf{H}_m^* \mathbf{C}_m(\Delta_{\max} \cdot \tanh(\mathbf{u}_m)) \mathbf{v}\|_2. \quad (6.16)$$

Les paramètres du critère (6.16) sont estimés itérativement en utilisant l'algorithme de descente de gradient ADAM, largement reconnu et employé en apprentissage automatique pour l'entraînement des réseaux de neurones profonds disposant d'une très grande quantité de paramètres. Au cours d'expériences préliminaires, il s'est montré être plus performant que l'algorithme du gradient stochastique pour le problème considéré, avec une difficulté moindre pour le réglage du taux d'apprentissage.

De plus, il a été constaté au cours d'expériences préliminaires que l'optimisation du critère (6.16)

ne permettait d'estimer correctement les délais dus aux erreurs géométriques que lorsqu'ils étaient très faibles, de l'ordre de moins d'un échantillon. Ce phénomène est observé dans le cas considérant des réponses idéales pour les sources et microphones comme pour celui utilisant des mesures réelles de réponses à dépendance directionnelle et fréquentielle. Comme le montrent les exemples des figures 6.2 et 6.3, l'examen de l'évolution de l'erreur quadratique moyenne pour un écho isolé en fonction du délai fractionnaire restreint à l'intervalle $[-\Delta_{max}, \Delta_{max}]$ (pour $\bar{f}\Delta_{max} = 3.73$ échantillons) révèle l'existence de minima locaux dans ces deux situations. L'observation des signaux correspondant montrent que ces minima locaux sont provoqués par une chute brutale suivie d'une hausse de leur amplitude en l'espace d'un nombre très faible d'échantillons, systématiquement observée sur les cas sélectionnés.

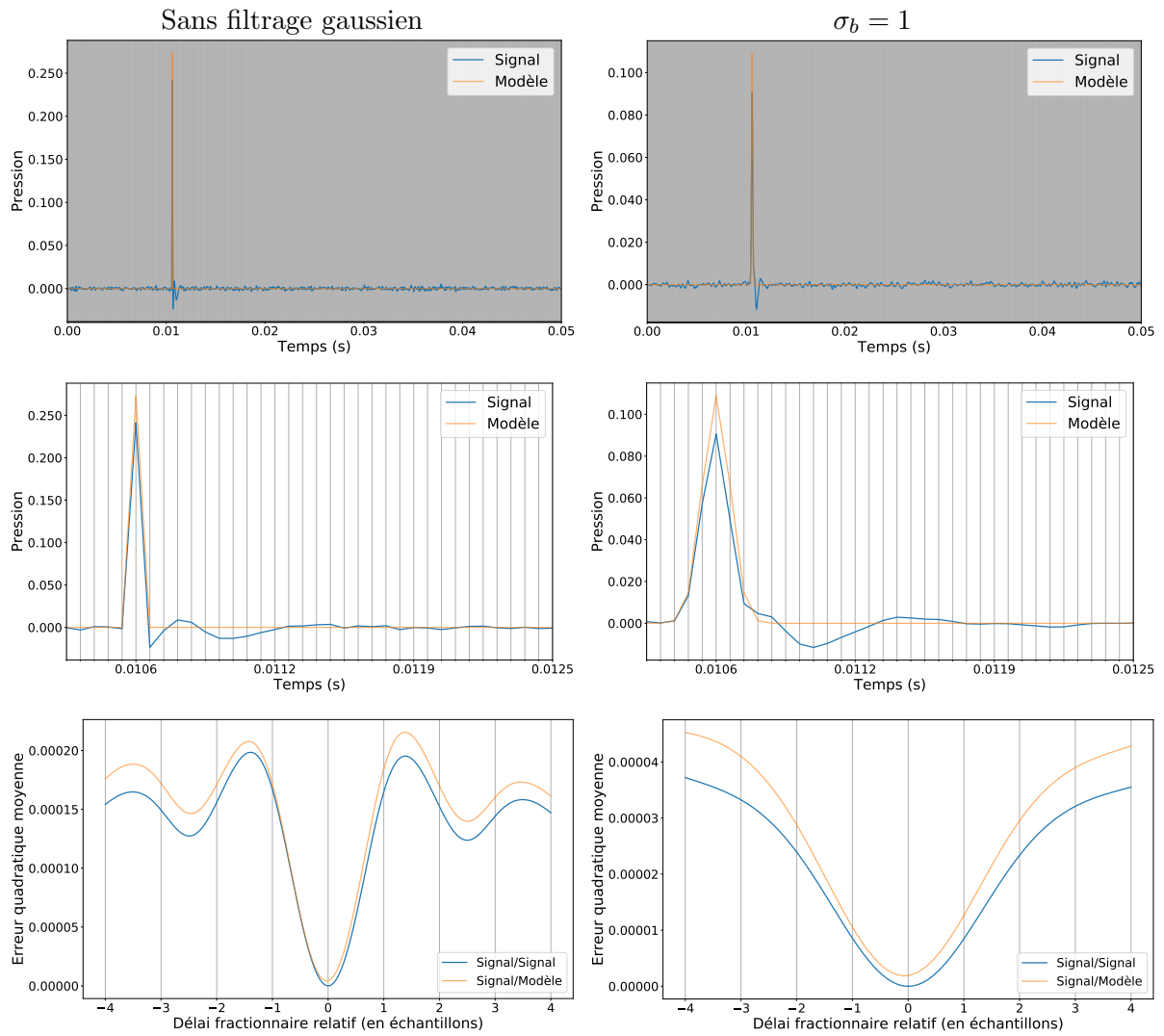


FIGURE 6.2 – Évolution de l'erreur quadratique moyenne en fonction du délai fractionnaire correspondant respectivement au critère sans filtrage gaussien (6.16) (colonne de gauche) et avec filtrage gaussien (6.18) (colonne de droite) pour un écho isolé dans le cas de sources et microphones aux réponses idéales.

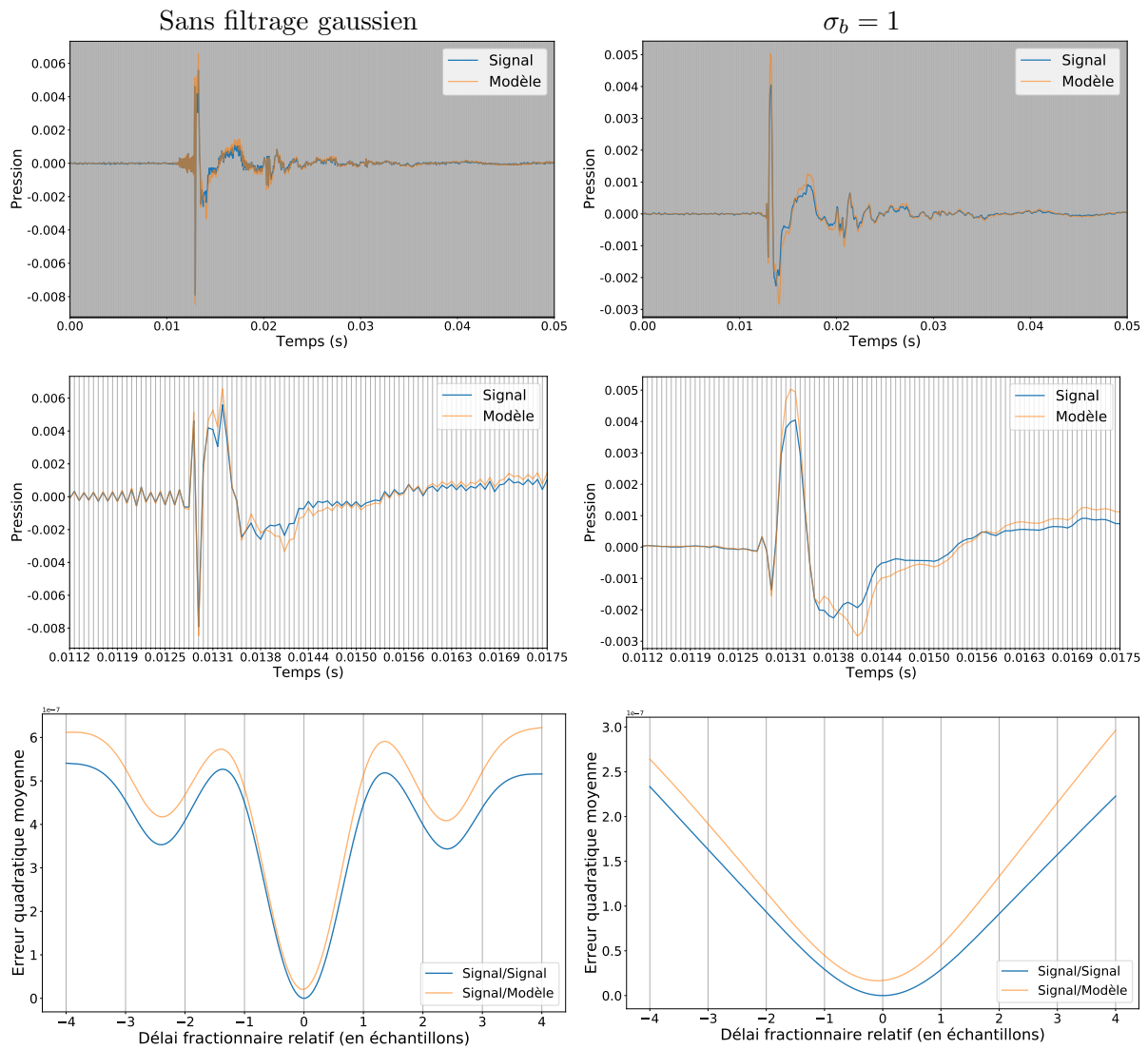


FIGURE 6.3 – Évolution de l’erreur quadratique moyenne en fonction du délai fractionnaire correspondant respectivement au critère sans filtrage gaussien (6.16) (colonne de gauche) et avec filtrage gaussien (6.18) (colonne de droite) pour un écho isolé dans le cas de sources et microphones aux réponses mesurées à dépendance directionnelle et fréquentielle.

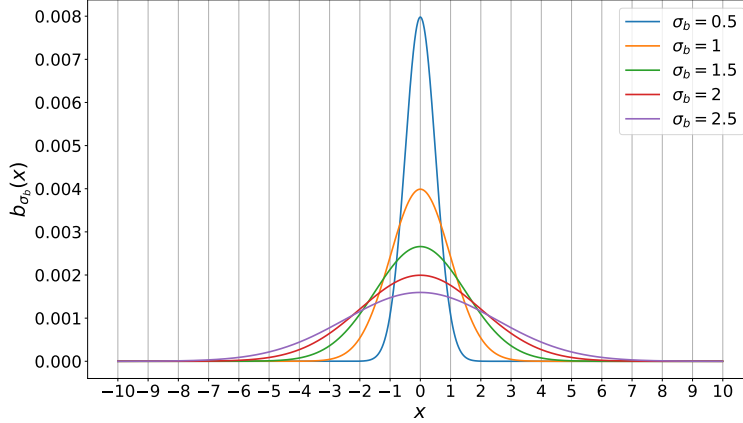


FIGURE 6.4 – Filtres de Gauss pour les différents écarts types considérés dans les expériences.

Afin de pallier à ce problème, un *filtre de Gauss* d'écart type σ_b , noté b_{σ_b} , est convolué au signal et au modèle afin de “lisser” ces irrégularités tout en conservant la validité de l'approche :

$$b_{\sigma_b}(x) = \frac{1}{\sqrt{2\pi}\sigma_b} \exp\left(-\frac{x^2}{2\sigma_b^2}\right). \quad (6.17)$$

Les exemples des figures 6.2 et 6.3 montrent que ce procédé permet d'obtenir une extension du bassin d'attraction au-delà de l'intervalle $[-\Delta_{max}, \Delta_{max}]$, pour $\sigma_b = 1$ échantillon. Aussi, les courbes d'erreur quadratique moyenne “Signal\Signal” (calculée avec v) et “Signal\Modèle” (calculée avec $\mathbf{v}_k = \delta_0$ tel que (6.10)) présentées dans ces exemples sont très similaires. Cela permet de conforter le choix fait de la valeur par défaut $\mathbf{v}_k = \delta_0$. La figure 6.4 représente les filtres de Gauss correspondant aux différentes valeurs d'écart type σ_b considérées dans les expériences.

Soit $\mathbf{B}_{\sigma_b} \in \mathbb{R}^{T,T}$ matrice de Toeplitz du vecteur \mathbf{b}_{σ_b} . Finalement, le critère utilisé pour l'estimation des délais par l'intermédiaire du changement de variable (6.15) et avec une convolution du signal et du modèle par une filtre de Gauss d'écart type σ_b (6.17) s'écrit :

$$\inf_{\mathbf{u}_m, \mathbf{v}_k = \delta_0} \|\mathbf{B}_{\sigma_b}(\mathbf{x}_m - \mathbf{H}_m \mathbf{C}_m(\Delta_{max} \cdot \tanh(\mathbf{u}_m))\mathbf{v})\|_2. \quad (6.18)$$

La figure 6.5 résume schématiquement la procédure itérative d'estimation des délais correctifs par optimisation du critère (6.18) avec un algorithme de descente de gradient. La variable γ correspond au taux d'apprentissage et z est la fonction du gradient du critère calculée à chaque itération pour la mise à jour des paramètres.

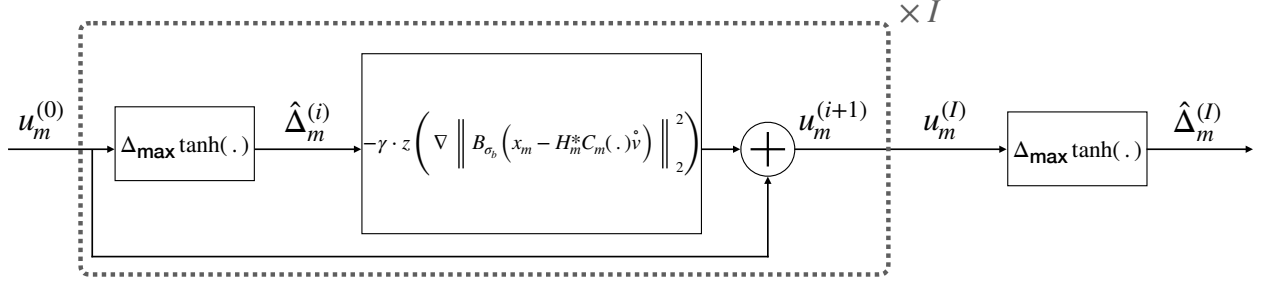


FIGURE 6.5 – Schéma de l'algorithme itératif de descente de gradient pour l'estimation des délais correctifs.

6.2.5 Estimation des réponses impulsionnelles des murs

La résolution du problème d'optimisation posé par (6.12) est effectuée en utilisant la méthode du gradient conjugué (Hestenes et al., 1952). Il s'agit d'un algorithme itératif largement utilisé pour résoudre efficacement des systèmes linéaires dont la matrice est réelle symétrique définie positive. Soit $\mathbf{b} = \mathbf{A}\mathbf{y}$ un tel système. La matrice du système ($\mathbf{A} \succ 0$) vérifie :

$$\begin{cases} \mathbf{A} = \mathbf{A}^T \text{ (Symétrie)} \\ \forall \mathbf{j} \in \mathbb{R}^N \setminus \{\mathbf{0}\}, \quad \mathbf{j}^T \mathbf{A} \mathbf{j} > 0 \text{ (Positivité)}. \end{cases}$$

Le principe général de la méthode consiste à construire de manière itérative un ensemble de vecteurs $(\mathbf{d}_1, \dots, \mathbf{d}_l)_{l \leq N}$ de \mathbb{R}^N appelés *directions*, pour estimer efficacement la solution de l'équation. Les $(\mathbf{d}_1, \dots, \mathbf{d}_l)_{l \leq N}$ sont construits de sorte à être 2 à 2 conjugués par rapport à \mathbf{A} , c'est-à-dire que pour tout $i \neq j$, $\mathbf{d}_i^T \mathbf{A} \mathbf{d}_j = 0$. Ce procédé permet ainsi d'accélérer la convergence vers le minimum. Lorsque $l = N$, un tel ensemble de vecteurs est une base de \mathbb{R}^N . La solution de l'équation $\mathbf{y}_{\text{solution}}$ peut donc être exprimée dans la base $(\mathbf{d}_1, \dots, \mathbf{d}_N)$:

$$\mathbf{y}_{\text{solution}} = \sum_{n=1}^N \alpha_n \mathbf{d}_n, \text{ avec } \alpha_i = \frac{\mathbf{d}_i^T \mathbf{b}}{\mathbf{d}_i^T \mathbf{A} \mathbf{d}_i} \text{ pour tout } i \in \llbracket 1, N \rrbracket.$$

L'algorithme donné par la méthode du gradient conjugué permet de choisir judicieusement les $(\mathbf{d}_1, \dots, \mathbf{d}_l)_{l \leq N}$ pour approximer la solution $\mathbf{y}_{\text{solution}}$ par une combinaison linéaire de ces directions. Lorsque le problème est bien posé, il est en général possible d'obtenir une bonne estimation de $\mathbf{y}_{\text{solution}}$ pour $l \ll N$.

La construction de cet ensemble de vecteurs repose sur la minimisation du critère convexe :

$$\operatorname{argmin}_{\mathbf{y}} \frac{1}{2} \mathbf{y}^T \mathbf{A} \mathbf{y} - \mathbf{y}^T \mathbf{b}. \quad (6.19)$$

En définissant $f(\mathbf{y}) = \frac{1}{2} \mathbf{y}^T \mathbf{A} \mathbf{y} - \mathbf{y}^T \mathbf{b}$, la minimisation du critère (6.19) correspond à la recherche du zéro du gradient de f qui s'exprime comme $\nabla f(\mathbf{y}) = \mathbf{A} \mathbf{y} - \mathbf{b}$. Le minimum du critère $\mathbf{y}_{\text{solution}}$ est donc une solution du problème inverse. En effet, $\nabla f(\mathbf{y}_{\text{solution}}) = 0 \implies \mathbf{A} \mathbf{y}_{\text{solution}} = \mathbf{b}$

L'idée clé de la méthode est d'utiliser à chaque itération la valeur actuelle du gradient et les directions prises aux itérations précédentes pour mettre à jour l'estimée de la solution. La direction utilisée pour la mise à jour est construite de sorte à être conjuguée aux directions des itérations précédentes par rapport à \mathbf{A} ainsi qu'à contribuer à la diminution du gradient. Un procédé analogue à celui de Gram-Schmidt pour l'orthonormalisation est utilisé pour garantir la conjugaison deux à deux des directions par rapport à \mathbf{A} au cours des itérations.

La méthode du gradient conjugué possède plusieurs avantages significatifs par rapport à d'autres méthodes itératives. Tout d'abord, elle converge en un nombre fini d'itérations pour les systèmes linéaires dont la matrice est symétrique et définie positive. De plus, l'algorithme fournit en général une bonne approximation de la solution en un nombre d'itérations relativement faible devant la taille du problème. De plus, en garantissant l'orthogonalité mutuelle des directions de recherche successives par rapport à la matrice du système, la méthode du gradient conjugué permet d'accélérer la convergence vers la solution lors de la descente du gradient.

À partir du modèle multi-observation (6.7), il est possible d'obtenir le système linéaire suivant :

$$\begin{aligned} \mathbf{x} &= \mathbf{H}^* \mathbf{C}(\Delta) \mathbf{v} \\ \implies (\mathbf{H}^* \mathbf{C}(\Delta))^T \mathbf{x} &= \underbrace{(\mathbf{H}^* \mathbf{C}(\Delta))^T \mathbf{H}^* \mathbf{C}(\Delta)}_{>0} \mathbf{v}. \end{aligned} \quad (6.20)$$

L'égalité (6.20) définit un système linéaire dont la matrice est réelle symétrique définie positive. Il est donc possible d'appliquer la méthode du gradient conjugué. L'estimation de \mathbf{v} par la méthode est obtenue par application de l'algorithme 6.1 détaillé ci-après :

Cette méthode d'estimation présente l'avantage de nécessiter uniquement comme paramètres le nombre d'itération de l'algorithme ainsi que la longueur des RIMs cumulées à estimer.

Algorithm 6.1 Méthode du gradient conjugué

- 1: **Entrée** : $\mathbf{x}, \mathbf{H}^*, \mathbf{C}(\Delta), n_{\text{iter}}$
 - 2: **Initialiser** \mathbf{v}
 - 3: $\mathbf{r} \leftarrow (\mathbf{H}^* \mathbf{C}(\Delta))^T (\mathbf{x} - \mathbf{H}^* \mathbf{C}(\Delta) \mathbf{v})$
 - 4: $\mathbf{p} \leftarrow \mathbf{r}$
 - 5: **Répéter** n_{iter} fois :
 - 6: $\alpha \leftarrow \frac{\mathbf{r}^T \mathbf{r}}{(\mathbf{p} \mathbf{H}^* \mathbf{C}(\Delta))^T \mathbf{H}^* \mathbf{C}(\Delta) \mathbf{p}} = \frac{\|\mathbf{r}\|_2^2}{\|\mathbf{H}^* \mathbf{C}(\Delta) \mathbf{p}\|_2^2}$
 - 7: $\mathbf{v} \leftarrow \mathbf{v} + \alpha \mathbf{p}$
 - 8: $\mathbf{r}_{\text{préc}} \leftarrow \mathbf{r}$
 - 9: $\mathbf{r} \leftarrow \mathbf{r} - \alpha (\mathbf{H}^* \mathbf{C}(\Delta))^T \mathbf{H}^* \mathbf{C}(\Delta) \mathbf{p}$
 - 10: $\beta \leftarrow \frac{\|\mathbf{r}\|_2^2}{\|\mathbf{r}_{\text{préc}}\|_2^2}$
 - 11: $\mathbf{p} \leftarrow \mathbf{r} + \beta \mathbf{p}$
 - 12: **Sortie** : \mathbf{v}
-

6.3 Expériences et résultats

6.3.1 Objectifs des expériences

Cette section présente une évaluation détaillée des performances de l’approche proposée en section 6.2 dans un cadre simulé “réaliste” utilisant des données *hybrides*, c’est-à-dire générées à partir de mesures réelles de réponses en direction et en fréquence de sources et de microphones, et en modélisant les incertitudes géométriques et le bruit de mesure.

En premier lieu, des expériences sont conduites afin d’étudier de manière approfondie, les performances de l’approche d’estimation des délais correctifs dans la sous-section 6.3.5 d’une part, et celles de l’approche d’estimation des RIMs dans la sous-section 6.3.6 d’autre part. Ensuite, la sous-section 6.3.7 révèle les résultats obtenus par l’approche d’optimisation en deux étapes qui consiste à estimer conjointement les RIMs après avoir estimé et appliqué les délais correctifs. Enfin, les sections 6.3.8 à 6.3.10 étudient respectivement l’influence sur ces résultats de différents facteurs que sont les réponses des sources et des microphones, le nombre de RISs utilisées et le niveau d’incertitude géométrique.

6.3.2 Paramètres de simulation

Les données utilisées pour les expériences sont générées en utilisant le simulateur d’acoustique de salle Pyroomacoustics (Scheibler et al., 2018) avec l’extension de Srivastava et al., 2022 permettant d’importer des mesures de réponses en direction et en fréquence de sources et de microphones réels pour la simulation. Les mesures des réponses d’un haut-parleur Genelec 8020 et d’un microphone omnidirectionnel AKG C414 publiées par Brandner et al., 2018 sont utilisées pour l’ensemble des expériences. La figure 2.10c présente un exemple d’une RIS simulée en utilisant ces mesures fournies par Brandner et al., 2018.

Chaque expérience considère un total de $N = 100$ salles en forme de pavé droit ($S = 6$ murs) dont les dimensions (longueur, largeur et hauteur) sont échantillonnées aléatoirement suivant une distribution uniforme dans l’intervalle $[3, 10] \times [3, 10] \times [2, 5]$ mètres. Les réponses des murs sont définies à partir de 6 coefficients d’absorption associés aux bandes d’octave logarithmiques centrées aux fréquences .125, .250, .5, 1, 2 et 4 kHz. Ces coefficients sont échantillonnées aléatoirement suivant la stratégie *RB* (*Reflexivity Biased*) proposée par Foy et al., 2021 et employée également pour les expériences des chapitres 4 et 5. Cette stratégie d’échantillonnage des profils d’absorption favorise les murs réfléchissants afin d’atteindre une représentativité à l’image des murs typiquement rencontrés en acoustique des salles. Il est à noter que ce biais de réflexivité peut contribuer à complexifier la résolution du problème étudié en augmentant globalement le nombre de échos énergétiques et donc le niveau de mélange dans les RISs. Pour chaque salle, un total de 25 RISs

sont simulées pour des emplacements libres de sources et de microphones dans la salle, échantillonnés suivant une distribution uniforme en respectant une distance minimale de 1 mètre aux murs ainsi qu’entre la source et l’un des microphones (ISO 3382-2, 2008).

Les résultats des expériences portant sur l’estimation mono-observation des délais correctifs sont obtenus en utilisant la totalité des RISs simulées (2500 au total par expérience). L’estimation multi-observation des RIMs est effectuée en utilisant pour chaque salle $M = 16$ RISs choisies arbitrairement parmi les 25 simulées. La sous-section 6.3.9 étudie l’évolution de ces résultats en fonction du nombre de RISs utilisées pour $M \in \{1, 4, 9, 16, 25\}$.

Les RISs sont simulées à la fréquence d’échantillonnage $\bar{f} = 16\text{kHz}$ permettant ainsi d’étudier le comportement acoustique des salles jusqu’à 8kHz. Il est ainsi possible de couvrir le spectre de la voix humaine ($< 3\text{kHz}$) ainsi que celui de l’essentiel des instruments de musique les plus communément employés ($< 5\text{kHz}$) (Möser et al., 2009).

Les RISs simulées incluent l’ensemble des échos spéculaires d’ordre inférieur ou égal à $Q_{\text{sim}} = 20$, au nombre de 11521 par RIS. Un bruit blanc gaussien additif leur est respectivement appliqué afin d’atteindre un PSNR (*Peak Signal-to-Noise Ratio*) égal à 50dB. Elles sont ensuite tronquées à l’instant $T/\bar{f} = 50\text{ms}$. L’ordre de réflexion maximal des sources-images visibles dans l’ensemble des segments de RIS ainsi créés est égal à 10. Sur l’ensemble des 2500 segments de RISs créés, le nombre minimal, moyen et maximal de sources-images visibles est respectivement de 38, 122 et 290.

L’incertitude géométrique est modélisée en considérant des erreurs aléatoires indépendantes sur la mesure des positions des sources et des microphones suivant la loi $\mathcal{N}(0, \sigma_{\text{geo}}^2)$, avec $\sigma_{\text{geo}} = 0.02\text{m}$. Cette modélisation de l’incertitude géométrique revient à considérer une erreur aléatoire sur la mesure des positions de chaque source-image et microphone suivant la loi $\mathcal{N}(0, \sigma_{\text{geo}}^2)$. En pratique, pour l’ensemble des sources-images, les temps d’arrivée utilisés comme donnée sont obtenus en ajoutant un délai aléatoire de loi $\mathcal{N}(0, 2\sigma_{\text{geo}}\bar{f}/c)$ aux temps d’arrivée exacts calculés à partir de la méthode des sources-images, avec $c = 343\text{ms}^{-1}$ la vitesse de propagation du son.

La figure 6.6 illustre les signaux correspondant aux données d’entrée ainsi qu’aux cibles du problème, sur l’exemple d’une RIS simulée jusqu’au premier ordre de réflexion. Rappelons les modèles mono-observation (6.6) et multi-observation (6.7) sur lesquels reposent notre approche :

$$\begin{aligned}\mathbf{x}_m &= \mathbf{H}_m^* \mathbf{C}_m(\boldsymbol{\Delta}_m) \mathbf{v} && \text{(mono-observation),} \\ \mathbf{x} &= \mathbf{H}^* \mathbf{C}(\boldsymbol{\Delta}) \mathbf{v} && \text{(multi-observation).}\end{aligned}$$

Pour chaque salle, les données utilisées pour résoudre le problème sont les M RISs $\{\mathbf{x}_m\}$ (figure 6.6a) ainsi que les réponses source-microphone localisées et atténuées conformément aux temps d’arrivée erronés $\{\mathbf{H}_m^*\}$ (figure 6.6b). Les cibles sont les délais correctifs $\{\boldsymbol{\Delta}_m\}$ ainsi que les RIMs

\mathbf{v} . La première étape consiste à utiliser les données du problème afin d'estimer, pour chaque RIS, les délais correctifs Δ_m dans l'objectif d'obtenir les réponses source-microphone localisées et atténuées conformément aux temps d'arrivée exacts $\mathbf{H}_m^* \mathbf{C}_m(\Delta_m)$. L'ensemble des délais correctifs estimés est noté $\hat{\Delta}$ (tel que $\hat{\Delta} = \{\hat{\Delta}_m\}$). La seconde étape porte sur l'estimation des RIMs cumulées \mathbf{v} par la méthode du gradient conjugué, à partir des RISs \mathbf{x} et des réponses source-microphone localisées et atténuées conformément aux temps d'arrivée estimés à l'issue de la première étape $\mathbf{H}^* \mathbf{C}(\hat{\Delta})$.

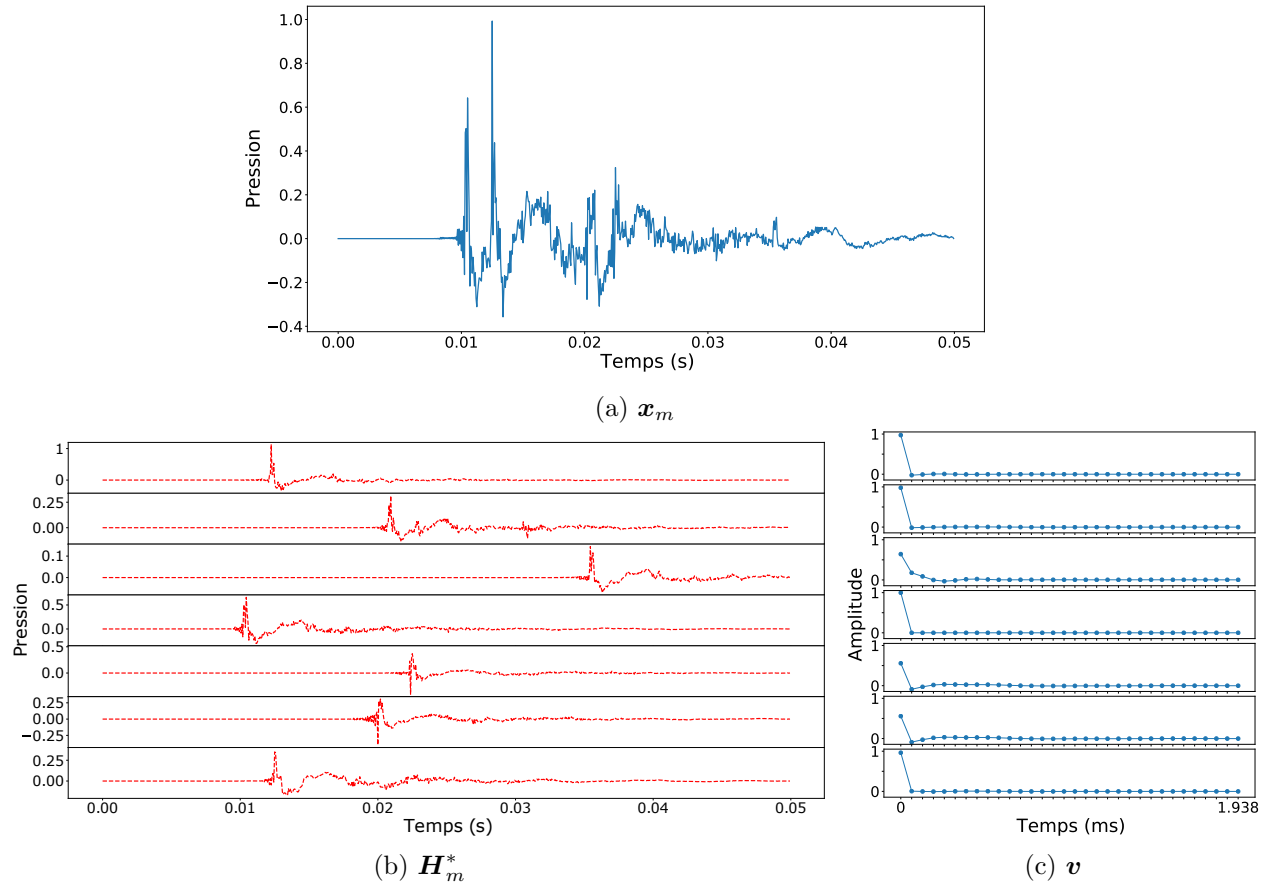


FIGURE 6.6 – (6.6a) Simulation jusqu'au premier ordre de réflexion d'une RIS x_m en utilisant des mesures de réponses en direction et en fréquence de sources et de microphones réels par Brandner et al., 2018. (6.6b) Chaque signal correspond à la RIS théorique qui serait obtenu si l'un des 6 murs était parfaitement réfléchissant, et tous les autres parfaitement absorbant, en présence d'erreurs géométriques. (6.6c) RIM pour chacune des 6 parois. La RIS (6.6a) correspond donc à la somme des six signaux en (6.6b) respectivement convolués avec les six signaux en (6.6c), aux délais issus des erreurs géométriques prêt : $\mathbf{x}_m = \mathbf{H}_m^* \mathbf{C}_m(\Delta_m) \mathbf{v}$ (voir l'équation 6.6). (6.6a) et (6.6b) correspondent aux données d'entrée de notre problème, alors que les délais correctifs Δ_m ainsi que les RIMs (6.6c) sont les inconnues du problème.

6.3.3 Paramètres de la méthode

Pour l'estimation des délais correctifs, l'intervalle des valeurs autorisées pour l'optimisation du critère, défini par Δ_{\max} , correspond à l'intervalle de confiance à 95% de la distribution des délais appliqués aux temps d'arrivée exacts pour modéliser l'incertitude géométrique, $\mathcal{N}(0, 2\sigma_{\text{geo}}\bar{f}/c)$. Dans ce cas, $\Delta_{\max} = 0.233\text{ms}$ soit 3.73 échantillons, équivalent à une distance parcourue par le son de 8cm.

Pour l'estimation des délais correctifs, l'optimisation des critères sans filtrage gaussien (6.16) et avec filtrage gaussien (6.18) est effectuée à l'aide de l'algorithme ADAM avec un taux d'apprentissage $\gamma = 1 \cdot 10^{-2}$. Le reste des paramètres sont ceux définis par défaut dans son implémentation en Pytorch. Il s'agit de la valeur de γ qui a permis d'obtenir la plus petite erreur quadratique moyenne sur l'estimation des RIMs pour l'ensemble des expériences effectuées avec $\gamma \in \{1 \cdot 10^{-1}, 1 \cdot 10^{-2}, 5 \cdot 10^{-3}, 1 \cdot 10^{-3}, 1 \cdot 10^{-4}\}$. Pour l'optimisation du critère avec filtrage gaussien (6.18), l'écart-type du filtre de Gauss est $\sigma_b = 1$. Cette valeur s'est avérée optimale lors d'expériences préliminaires parmi les valeurs $\sigma_b \in \{0.5, 1, 1.5, 2, 2.5\}$.

Le nombre maximal d'itérations de l'algorithme d'optimisation des délais correctifs est fixé à 500. Un arrêt anticipé intervient après un minimum de 60 itérations lorsque la valeur courante du critère est supérieure à 99% de celle observée 50 itérations en arrière (diminution du critère inférieure à 1% après 50 itérations). La taille des RIMs cumulées $\{\mathbf{v}_k\}$ est fixée à $L = 32$ échantillons, soit 2 ms, en reflet des RIMs de très courte durée sous l'hypothèse minimum phase (voir la sous-section 3.2.4). Le nombre d'itérations de l'algorithme du gradient conjugué est fixé à 1000 sur la base d'expériences préliminaires afin d'assurer la convergence de l'estimation.

6.3.4 Métriques d'évaluation

Estimation des délais correctifs Δ :

Les métriques définies pour évaluer la convergence et les résultats de l'estimation (mono-observation) des délais correctifs sont moyennées sur l'ensemble des RISs et l'ensemble des salles. Il s'agit de :

- L'erreur quadratique moyenne de reconstruction de la RIS issue du critère d'optimisation (6.16) ou (6.18), notée EQM_x . Il s'agit de la fonction de coût utilisée pour optimiser les délais correctifs;
- L'erreur absolue moyenne sur l'estimation des délais correctifs, notée EAM_Δ ;
- Le pourcentage d'estimations correctes des délais correctifs, défini pour une erreur absolue inférieure à 1 échantillon et noté $\text{CE}_\Delta\%$.

Il est important de noter que pour une même valeur des délais correctifs Δ , la valeur du critère EQM_x varie suivant la configuration d'expérience. En effet, la fonction de coût en Δ varie suivant la valeur par défaut choisie pour les $\{\mathbf{v}_k\}$ ainsi que suivant si un filtre de Gauss est utilisé ou non et la valeur de σ_b le cas échéant. Par conséquent, la métrique EQM_x est utilisée pour étudier le comportement de convergence individuellement pour chaque configuration d'expérience. Les métriques EAM_Δ et $\text{CE}_\Delta\%$ sont elles utilisées pour évaluer et comparer les résultats d'estimation obtenus par les différentes configurations d'expérience.

Estimation des RIMs cumulées \mathbf{v} :

Des métriques portant sur l'estimation des RIMs cumulées $\{\mathbf{v}_k\}$ ainsi que RIMs $\{\mathbf{v}_{(k)}\} = \{\mathbf{R}_k\}$ sont considérées. De plus, des métriques plus usuelles et évocatrices portant sur les densités spectrales d'énergie de ces estimations (carré du module de la TFD), et en particulier sur les profils d'absorption estimés ($\alpha_k = 1 - \tilde{v}_{(k)}$, voir équation (2.24)) sont également considérées. Ces densités spectrales d'énergie $\{\tilde{\mathbf{v}}_k\}$ sont constitués de $F = 17$ coefficients répartis sur une échelle linéaire des fréquences comprises entre 0 et 8kHz. Leurs valeurs sont écrêtées à 1 conformément au sens physique.

Les métriques définies pour évaluer la convergence et les résultats de l'estimation (multi-observation) des RIMs cumulées avec l'algorithme du gradient conjugué sont moyennées sur l'ensemble des salles. Il s'agit de :

- La norme infinie du résidu d'estimation, noté $\|\mathbf{r}\|_\infty = \sup_i |r_i|$ (voir algorithme 6.1) afin d'évaluer la convergence de l'estimation des RIMs cumulées ;
- L'erreur quadratique moyenne sur l'estimation des RIMs cumulées, notée $\text{EQM}_{\mathbf{v}_k}$;
- L'erreur absolue moyenne sur l'estimation des densités spectrales d'énergie des RIMs cumulées, notée $\text{EAM}_{\tilde{\mathbf{v}}_k}$;
- Le pourcentage d'estimations correctes des coefficients des densités spectrales d'énergie des RIMs cumulées, défini pour une erreur absolue inférieure à 0.1 et noté $\text{CE}_{\tilde{\mathbf{v}}_k}\%$;
- L'erreur quadratique moyenne sur l'estimation des RIMs, notée $\text{EQM}_{\mathbf{v}_{1:6}}$;
- L'erreur absolue moyenne sur l'estimation des profils d'absorption des murs, notée $\text{EAM}_{\alpha_{1:6}}$;
- Le pourcentage d'estimations correctes des coefficients des profils d'absorption des murs, défini pour une erreur absolue inférieure à 0.1 et noté $\text{CE}_{\alpha_{1:6}}\%$.

6.3.5 Estimation des délais correctifs

Cette sous-section étudie l'évolution des performances de l'approche d'estimation des délais correctifs obtenues respectivement par optimisation des 3 critères suivants :

- (i) sans filtrage gaussien (6.16) avec $\{\mathbf{v}_k = \boldsymbol{\delta}_0\}$;
- (ii) avec filtrage gaussien (6.18) avec $\{\mathbf{v}_k = \boldsymbol{\delta}_0\}$;
- (iii) avec filtrage gaussien (6.18) avec les RIMs cumulées exactes ;

Le critère (iii) correspond à un cas “oracle”, puisqu’en pratique les RIMs exactes ne sont pas connues. Son objectif est d’identifier les meilleurs résultats pouvant être espérés en alternant successivement l’estimation des délais correctifs avec le critère (ii) et celle des RIMs cumulées avec la méthode du gradient conjugué.

La figure 6.7 présente l’évolution de la fonction de coût EQM_x (erreur quadratique moyenne de reconstruction de la RIS) au cours des itérations de l’optimisation pour chacun des 3 critères définis.

La figure 6.7a montre qu’en l’absence de filtrage gaussien des signaux (6.16) et avec $\mathbf{v}_k = \boldsymbol{\delta}_0$ pour l’optimisation, la fonction de coût converge vers une valeur nettement supérieure à la valeur correspondant aux vrais délais correctifs ($\hat{\boldsymbol{\Delta}} = \boldsymbol{\Delta}$). Cette situation confirme l’existence de minima locaux pour ce critère et la tendance marquée de la méthode à converger vers l’un d’entre eux plutôt que vers son minimum global.

La figure 6.7b révèle que lorsque le filtrage gaussien est employé et que $\mathbf{v}_k = \boldsymbol{\delta}_0$ pour l’optimisation, la fonction de coût converge vers une valeur limite nettement inférieure à la valeur correspondant aux vrais délais correctifs ($\hat{\boldsymbol{\Delta}} = \boldsymbol{\Delta}$). Ce résultat démontre qu’il ne s’agit pas du minimum global de ce critère. Lorsque des RIMs parfaitement rigides sont considérées ($\mathbf{v}_k = \boldsymbol{\delta}_0$), le modèle comporte des échos nécessairement plus énergétiques que ceux présents dans le signal. Un piste d’explication peut être que l’optimisation du critère conduit à surestimer certains délais afin de faire décroître l’énergie des échos présents dans le modèle grâce à l’atténuation géométrique, voire d’en exclure certains des segments de RIS.

Enfin, la figure 6.7c démontre que lorsque le filtrage gaussien est employé et que les RIMs cumulées exactes sont utilisées pour l’optimisation, la fonction de coût converge vers un minimum global égal à la valeur correspondant aux vrais délais correctifs ($\hat{\boldsymbol{\Delta}} = \boldsymbol{\Delta}$), c’est-à-dire zéro. Ce résultat indique que le filtrage gaussien semble permettre de définir un critère qui favorise efficacement la convergence de la méthode vers un minimum global.

La figure 6.8 présente l’évolution de l’erreur absolue moyenne sur l’estimation des délais exprimés en échantillons ($\text{EAM}_{\boldsymbol{\Delta}}$, figure 6.8a) ainsi que du pourcentage d’estimations correctes, c’est-à-dire avec une erreur absolue inférieure à 1 échantillon ($\text{CE}_{\boldsymbol{\Delta}}\%$, figure 6.8b) au cours des itérations de l’optimisation pour chacun des trois critères définis.

Les figures 6.8a et 6.8b révèlent, lorsque des RIMs rigides sont considérées pour l’optimisation, que l’utilisation du filtrage gaussien permet d’améliorer significativement les résultats d’estimation

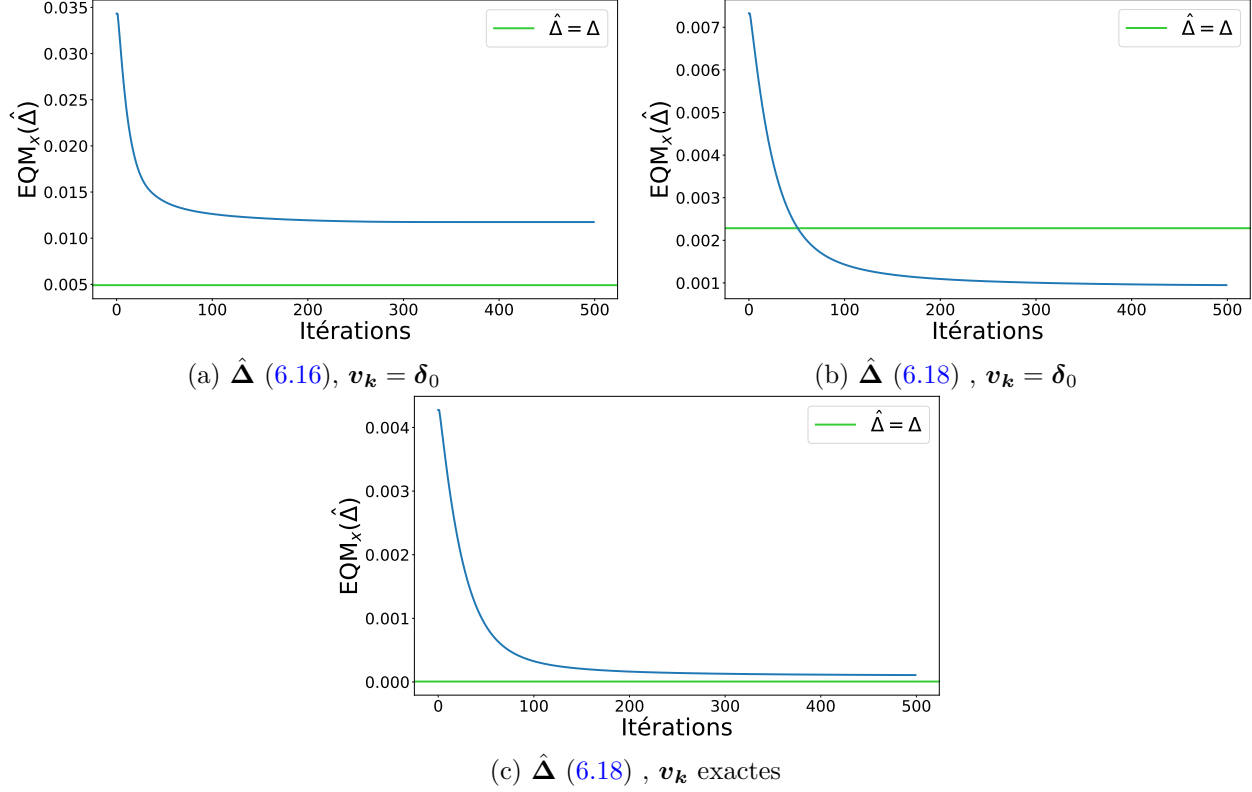


FIGURE 6.7 – Évolution de la fonction de coût (EQM_x) au cours des itérations de l’optimisation des critères “sans filtrage gaussien et $\mathbf{v}_k = \delta_0$ ” (6.7a), “avec filtrage gaussien et $\mathbf{v}_k = \delta_0$ ” (6.7b) et “avec filtrage gaussien et RIMs cumulées \mathbf{v}_k exactes” (6.7c). La ligne horizontale verte “ $\hat{\Delta} = \Delta$ ” indique pour chaque critère la valeur pour les délais correctifs exacts.

des délais. Ce constat est renforcé en particulier dans la première phase d’optimisation définie jusqu’à l’itération critique identifiée en figure 6.7c pour le critère avec filtrage gaussien et $\mathbf{v}_k = \delta_0$ telle que $\text{EQM}_x(\hat{\Delta}) = \text{EQM}_x(\Delta)$. La figure 6.8a montre pour ce critère (courbe orange) que l’erreur absolue moyenne sur l’estimation des délais diminue strictement jusqu’à cette itération critique environ puis augmente strictement. En parallèle, la figure 6.8b montre que le pourcentage des délais correctement estimés correspondant, après avoir atteint son maximum, présente une légère décroissance constante au cours des itérations. Cette décroissance reflète l’augmentation progressive de la proportion des délais estimés avec une erreur absolue supérieure à 1 échantillon.

Cette observation vient corroborer l’hypothèse faite précédemment suivant laquelle l’optimisation de ce critère (avec filtrage gaussien (6.18) et $\mathbf{v}_k = \delta_0$) conduirait à surestimer certains délais afin de réduire l’énergie du modèle considérant des RIMs parfaitement rigides. Toutefois, la valeur du délai correctif maximal autorisé Δ_{\max} permet de contrôler ce phénomène. En effet, dans une plage temporelle suffisamment réduite, l’influence de l’atténuation géométrique reste limitée. En ce sens, il est attendu a minima pour les zones les moins denses des RIS que l’alignement correct des échos reste privilégié. Dans cette hypothèse, le gain de performance obtenu grâce au filtrage

gaussien portent principalement sur les échos précoces des RISs, dont il est raisonnable de penser qu'ils jouent un rôle crucial pour l'estimation des RIMs. Les expériences de la sous-section 6.3.7 permettront d'évaluer l'impact final de l'utilisation du filtrage gaussien sur l'estimation des RIMs.

Enfin, la figure 6.8a démontre que l'utilisation du filtrage gaussien et des valeurs exactes des RIMs cumulées permet d'améliorer considérablement les résultats et d'obtenir une erreur absolue moyenne d'estimation des délais inférieure à 1 échantillon. Ce résultat suggère qu'une meilleure connaissance des RIMs permet d'améliorer l'estimation des délais, ce qui encourage le développement d'une approche alternée efficace. Aussi, pour ce critère, les métriques EAM_{Δ} et $CE_{\Delta}\%$ s'améliorent strictement au cours des itérations, ce qui indique que le phénomène de surestimation des délais évoqué tend à disparaître lorsque l'estimation des RIMs cumulées se précise.

Notons cependant que bien que l'utilisation de ce critère permette d'obtenir d'excellents résultats de reconstruction des RISs (figure 6.7c), près d'un quart des délais sont estimés avec une erreur absolue supérieure à 1 échantillon ($CE_{\Delta}\% = 74.1\%$). Ce constat indique qu'il existe une fraction non négligeable des échos pour laquelle une erreur significative sur l'estimation du délai produit un impact relativement faible sur l'erreur de reconstruction de la RIS. Plusieurs interprétations à ce phénomène sont possibles. Premièrement, ce constat peut refléter qu'une part importante des échos correspondent à des RIMs cumulées similaires. Dans ce cas, leurs permutations n'impactent ni l'erreur de reconstruction des RISs, ni l'erreur finale d'estimation des RIMs, ce qui ne serait donc pas un problème en pratique. Deuxièmement, cette observation peut concerner des échos peu énergétiques dont l'influence sur le calcul de l'erreur de reconstruction est faible. Dans cette situation, la compensation de l'atténuation géométrique pourrait éventuellement contribuer à augmenter leur poids dans l'optimisation. Enfin, il est possible d'intuiter que cette observation concerne essentiellement les échos les plus tardifs des segments de RISs, dont les caractéristiques tendent vers celles du champ diffus pour lequel le positionnement exact des échos a peu d'influence.

Finalement, lorsque l'incertitude géométrique est modélisée par une erreur sur la position de chaque microphone et source-image, de loi normale centrée d'écart-type égal à 2cm, cette étude prouve qu'un filtrage gaussien préalable des signaux contribue à améliorer significativement les résultats d'estimation des délais correctifs.

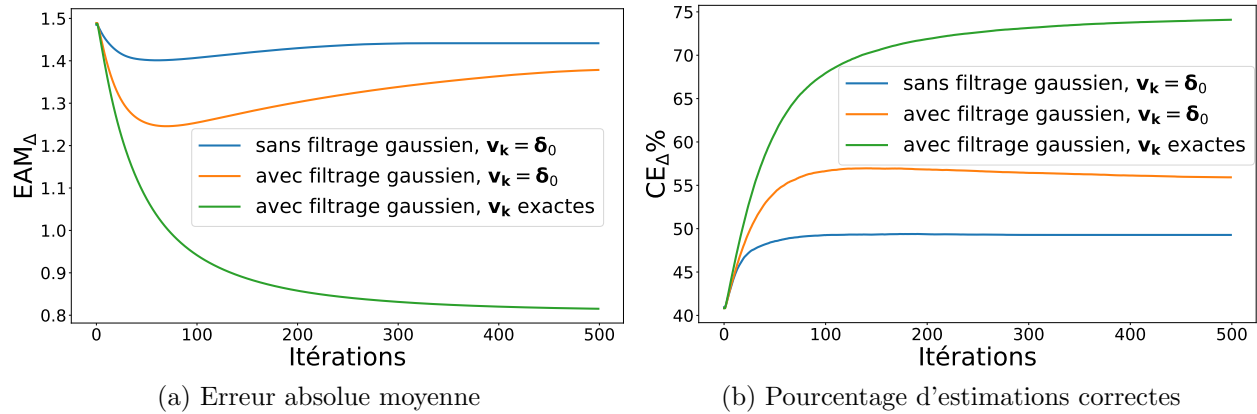


FIGURE 6.8 – Erreur absolue moyenne sur l’estimation des délais exprimés en échantillons (EAM_{Δ} , figure 6.8a) et pourcentage d’estimations correctes, c’est-à-dire avec une erreur absolue inférieure à 1 échantillon ($CE_{\Delta}\%$, figure 6.8b) au cours des itérations de l’optimisation des critères “sans filtrage gaussien (6.16) et $\mathbf{v}_k = \delta_0$ ” (courbes bleues), “avec filtrage gaussien (6.18) et $\mathbf{v}_k = \delta_0$ ” (courbes oranges) et “avec filtrage gaussien (6.18) et RIMs cumulées \mathbf{v}_k exactes” (courbes vertes).

6.3.6 Estimation des réponses impulsionnelles des murs avec connaissance exacte des temps d'arrivée des échos

Cette sous-section analyse le comportement de convergence et les résultats obtenus pour l'estimation des RIMs cumulées par la méthode du gradient conjugué, avec connaissance exacte des temps d'arrivée des échos. L'étude de ce cas "oracle" ($\hat{\Delta} = \Delta$), permet d'évaluer spécifiquement les performances de la méthode du gradient conjugué. Il peut correspondre à une hypothèse de connaissance exacte de la géométrie ou de compensation exacte des délais induits par les erreurs géométriques. Pour rappel, $M = 16$ RISs sont considérées dans chaque salle pour l'estimation.

La figure 6.9 présente le résidu d'estimation ($\|\mathbf{r}\|_\infty$, 6.9a) et l'erreur quadratique moyenne relative aux RIMs cumulées (EQM_{v_k} , 6.9b) au cours des itérations de la méthode du gradient conjugué. Certaines métriques sont calculées séparément pour les sources-images d'ordre égal à 1 ($q(\mathbf{k}) = 1$) et celles d'ordre strictement supérieur à 1 ($q(\mathbf{k}) > 1$).

La figure 6.9a permet d'observer que la norme infinie du résidu d'estimation des RIMs cumulées décroît globalement et tend vers 0 avec une convergence en moins de 200 itérations. Ce résultat encourageant indique que, avec connaissance exacte des temps d'arrivée, la méthode du gradient conjugué permet d'obtenir une estimation des RIMs cumulées qui résout précisément le problème multi-observation (6.7). Il est observé que bien que $\|\mathbf{r}\|_\infty$ décroisse globalement au cours des itérations, cette décroissance n'est pas stricte. Une explication possible est que la norme infinie d'un vecteur (coefficient maximal en valeur absolue) peut croître tandis que sa norme \mathbf{L}^1 (par exemple) décroît.

La figure 6.9b permet de constater que l'erreur quadratique moyenne sur l'estimation des RIMs (RIMs cumulées d'ordre égal à 1) décroît et tend vers 0 au cours des itérations de la méthode du gradient conjugué. Cependant, elle montre que celle relative aux RIMs cumulées d'ordre strictement supérieur 1 croît strictement à partir d'une certaine itération pour tendre vers une valeur élevée (supérieure à sa valeur initiale). Ce constat démontre que lorsque les temps d'arrivée sont connus exactement, la méthode permet d'estimer très précisément les RIMs sans parvenir pour autant à estimer l'ensemble des RIMs cumulées. Une interprétation possible est que lorsque l'ordre des réflexions augmente, le niveau de mélange des échos dans les RISs devient tel que le nombre de paramètres à estimer devient nettement supérieur au nombre de RISs utilisées (ici $M = 16$). Dans cette situation, la redondance d'information est suffisante pour estimer les RIMs bien que insuffisante pour permettre une estimation correcte de l'ensemble des RIMs cumulées.

Le tableau 6.1 présente l'erreur quadratique moyenne d'estimation des RIMs cumulées (EQM_{v_k}) ainsi que l'erreur absolue moyenne et le pourcentage d'estimations correctes relatifs aux densités spectrales d'énergie des RIMs cumulées estimées ($\text{EAM}_{\tilde{v}_k}$ et $\text{CE}_{\tilde{v}_k}$ %). Rappelons que la densité spectrale d'énergie correspond au carré du module de la TFD. Un coefficient de la densité spec-

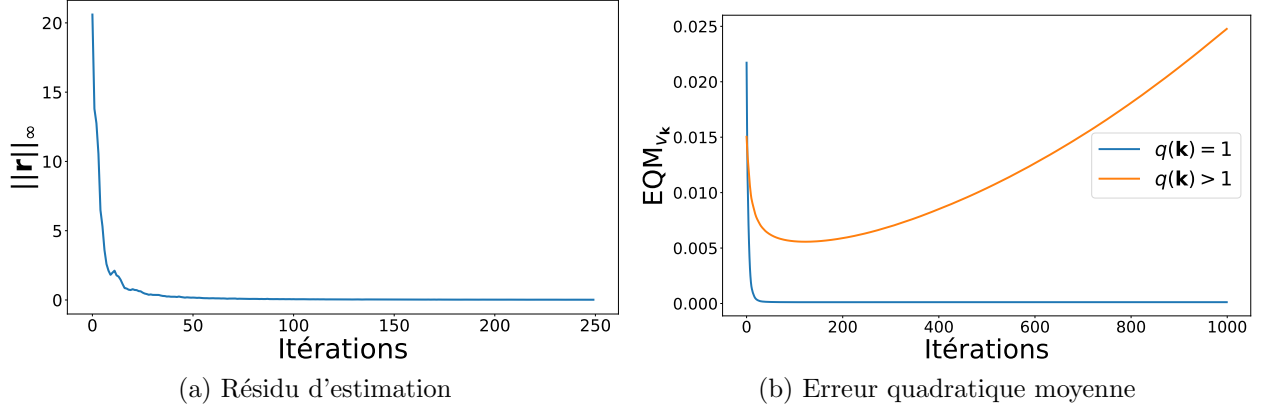


FIGURE 6.9 – Résidu d’estimation ($\|\mathbf{r}\|_\infty$, 6.9a) et erreur quadratique moyenne relative aux RIMs cumulées (EQM_{v_k} , 6.9b) au cours des itérations de la méthode du gradient conjugué. La figure 6.9b différencie l’évolution de l’ EQM_{v_k} pour les sources-images d’ordre égal à 1 d’une part (courbe bleu) et celles d’ordre strictement supérieur à 1 d’autre part.

trale d’énergie est considéré correctement estimé lorsque l’erreur absolue est inférieure à 0.1. Ces métriques sont calculées séparément pour les sources-images d’ordre égal à 1 ($q(\mathbf{k}) = 1$) et celles d’ordre strictement supérieur à 1 ($q(\mathbf{k}) > 1$). Rappelons également que la densité spectrale d’énergie d’une RIMs correspond à son profil de réflexion :

$$\begin{aligned}
 \mathbf{k} \in \mathcal{K}, \mathbf{k} = (k) &\implies |TFD\{\overbrace{v_k}^{\text{RIM}}\}|^2 = \tilde{v}_k, \text{ densité spectrale d'énergie} \\
 &= \omega_k, \text{ profil de réflexion} \\
 &= 1 - \alpha_k, \alpha_k \text{ profil d'absorption.}
 \end{aligned}$$

Le tableau 6.1 montre que l’erreur quadratique moyenne d’estimation des RIMs ($q(\mathbf{k}) = 1$) est inférieure de deux ordres de grandeur à celle relative aux RIMs cumulées des sources-images d’ordre strictement supérieur à 1 $q(\mathbf{k}) > 1$. En terme d’erreur sur les densités spectrales d’énergie, l’erreur absolue moyenne et le pourcentage d’estimation correctes passent respectivement de 0.025 et 97.1% pour $q(\mathbf{k}) = 1$ (estimation des profils d’absorption) à 0.253 et 39.2% pour $q(\mathbf{k}) > 1$.

La figure 6.10 présente l’évolution en fonction de la fréquence de l’erreur absolue moyenne et du pourcentage d’estimations correctes pour l’estimation des profils d’absorption des murs (EAM_{α_k} et $CE_{\alpha_k}\%$, $q(\mathbf{k}) = 1$ uniquement). Elle montre que pour cette expérience les deux métriques sont excellentes et stables à toutes les fréquences à l’exception du dernier échantillon fréquentiel pour lequel une forte diminution de la qualité de l’estimation est observée.

Finalement, cette étude démontre que lorsque les temps d’arrivée des échos sont considérés

	$q(\mathbf{k}) = 1$		$q(\mathbf{k}) > 1$	
	EQM $_{v_{\mathbf{k}}}$	EAM $_{\tilde{v}_{\mathbf{k}}}$ (CE $_{\tilde{v}_{\mathbf{k}}}$ %)	EQM $_{v_{\mathbf{k}}}$	EAM $_{\tilde{v}_{\mathbf{k}}}$ (CE $_{\tilde{v}_{\mathbf{k}}}$ %)
$\hat{\Delta} = \Delta$	$1.20 \cdot 10^{-4}$	0.025 (97.1%)	$2.48 \cdot 10^{-2}$	0.253 (39.2%)

TABLEAU 6.1 – Erreur quadratique moyenne relative aux RIMs cumulées (EQM $_{v_{\mathbf{k}}}$), erreur absolue moyenne et pourcentage d’estimations correctes relatifs aux spectres des RIMs cumulées (EAM $_{\tilde{v}_{\mathbf{k}}}$ et CE $_{\tilde{v}_{\mathbf{k}}}$ %), avec connaissance exacte des temps d’arrivée des échos ($\hat{\Delta} = \Delta$). Les métriques sont présentées séparément pour les sources-images d’ordre égal 1 ($q(\mathbf{k}) = 1$) et celle d’ordre strictement supérieur à 1 ($q(\mathbf{k}) > 1$).

connus exactement, la méthode du gradient conjugué permet d’estimer les RIMs avec une excellente précision. La figure 6.11 montre un exemple des résultats d’estimation des RIMs d’une salle obtenus pour cette configuration d’expérience. Cette salle est choisie en tant qu’exemple de référence pour comparer les performances obtenus avec les différents critères d’optimisation considérés par la suite.

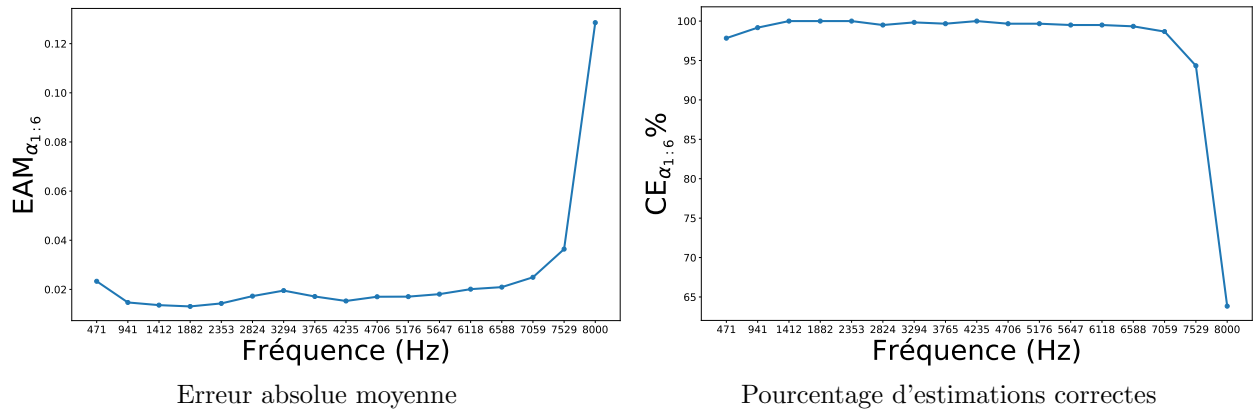


FIGURE 6.10 – Erreur absolue moyenne et pourcentage d’estimations correctes relatifs aux profils d’absorption des murs (EAM $_{\alpha_{1:6}}$ et CE $_{\alpha_{1:6}}$ %) en fonction de la fréquence, avec connaissance exacte des temps d’arrivée des échos ($\hat{\Delta} = \Delta$).

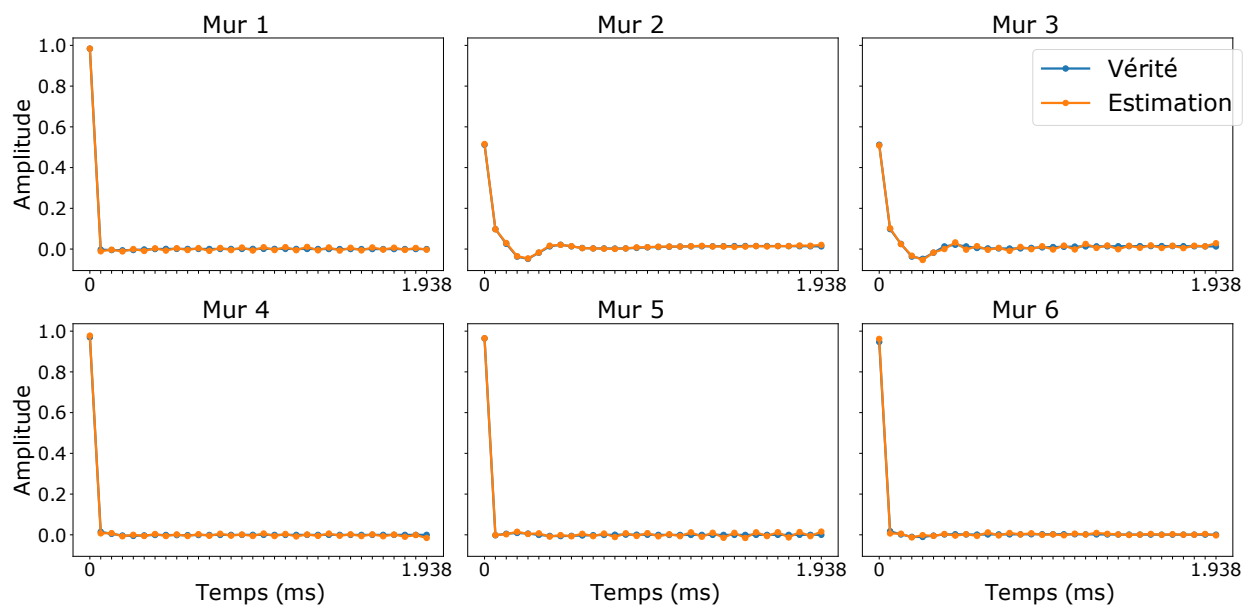


FIGURE 6.11 – Exemple des résultats d’estimation des RIMs d’une salle obtenus avec la méthode du gradient conjugué à partir de $M = 16$ RISs, avec connaissance exacte des temps d’arrivée des échos.

6.3.7 Correction des temps d'arrivée et estimation des réponses impulsionnelles des murs

Cette sous-section étudie les résultats obtenus avec l'approche d'optimisation en deux étapes proposée dans ce chapitre pour l'estimation des RIMs en présence d'erreurs sur le calcul des temps d'arrivée induites par l'incertitude géométrique.

Des expériences correspondant à différents modes d'estimation des délais correctifs Δ sont effectuées afin d'obtenir une compréhension détaillée du comportement et des résultats de la méthode proposée :

- la situation “oracle” qui correspond à une estimation parfaite des délais correctifs, notée “ $\hat{\Delta} = \Delta$ ” ;
- l'absence de correction des erreurs sur les temps d'arrivée, notée “ $\hat{\Delta} = 0$ ” ;
- l'estimation des délais correctifs par optimisation du critère sans filtrage gaussien (6.16) et $\mathbf{v}_k = \boldsymbol{\delta}_0$, notée “ $\hat{\Delta}$ (6.16), $\mathbf{v}_k = \boldsymbol{\delta}_0$ ” ;
- l'estimation des délais correctifs par optimisation du critère avec filtrage gaussien (6.18) et $\mathbf{v}_k = \boldsymbol{\delta}_0$, notée “ $\hat{\Delta}$ (6.18), $\mathbf{v}_k = \boldsymbol{\delta}_0$ ” ;
- la situation “oracle” qui correspond à l'estimation des délais correctifs par optimisation du critère avec filtrage gaussien (6.18) et les RIMs exactes, notée “ $\hat{\Delta}$ (6.18), \mathbf{v}_k exactes”.

Les séparations horizontales sont doublées dans les tableaux de résultats pour distinguer les cas “oracles” (“ $\hat{\Delta} = \Delta$ ” et “ $\hat{\Delta}$ (6.18), \mathbf{v}_k exactes”) des méthodes implémentables (“ $\hat{\Delta} = 0$ ”, “ $\hat{\Delta}$ (6.16), $\mathbf{v}_k = \boldsymbol{\delta}_0$ ”, “ $\hat{\Delta}$ (6.18), $\mathbf{v}_k = \boldsymbol{\delta}_0$ ”).

Le tableau 6.2 présente les résultats de ces différentes expériences avec les métriques suivantes (présentées en 6.3.4) :

- l'erreur absolue moyenne sur l'estimation des délais correctifs EAM_{Δ} et le pourcentage d'estimations correctes $CE_{\Delta}\%$ (erreur absolue inférieur à 1 échantillon) ;
- l'erreur quadratique moyenne sur l'estimation des RIMs $EQM_{v_{1:6}}$;
- l'erreur absolue moyenne sur l'estimation des profils d'absorption $EAM_{\alpha_{1:6}}$ et le pourcentage d'estimations correctes $CE_{\alpha_{1:6}}\%$ (erreur absolue inférieure à 0.1).

Tout d'abord, ce tableau montre qu'en l'absence de correction des temps d'arrivée erronés ($\hat{\Delta} = 0$), l'ensemble des métriques sont considérablement dégradées par rapport aux résultats obtenus lorsqu'ils sont exactement estimés ($\hat{\Delta} = \Delta$). L' $EQM_{v_{1:6}}$ se voit augmenter d'environ deux ordres de grandeur, l' $EAM_{\alpha_{1:6}}$ d'environ 50% et le $CE_{\alpha_{1:6}}\%$ diminuer de près de 80%.

De plus, il permet d’observer que l’utilisation du critère sans filtrage (6.16) avec $\mathbf{v}_k = \boldsymbol{\delta}_0$ pour l’estimation des délais correctifs améliore significativement l’ensemble des métriques par rapport au cas “ $\Delta = 0$ ”. Néanmoins, les performances de la méthode restent à ce stade insatisfaisantes. La figure 6.12 présente à titre d’exemple les résultats obtenus avec ce critère pour l’estimation des RIMs de la salle de référence. Elle illustre et confirme qu’il ne permet pas d’estimer correctement les RIMs.

Comparativement, les résultats obtenus en utilisant le critère avec filtrage gaussien (6.18) et $\mathbf{v}_k = \boldsymbol{\delta}_0$ démontrent l’importance et l’efficacité du filtrage gaussien des signaux. En effet, ce procédé permet approximativement de diviser l’ $EAM_{\alpha_{1.6}}$ par 2 et de multiplier le $CE_{\alpha_{1.6}}\%$ par 2. La figure 6.13 permet de constater sur l’exemple de référence l’amélioration significative des résultats d’estimation de l’ensemble RIMs obtenue en utilisant ce critère “ $\hat{\Delta}$ (6.18) , $\mathbf{v}_k = \boldsymbol{\delta}_0$ ”.

Enfin, les résultats obtenus en utilisant le critère avec filtrage gaussien (6.18) et considérant les RIMs cumulées exactes surpassent ceux obtenus avec le critère “ $\hat{\Delta}$ (6.18) , $\mathbf{v}_k = \boldsymbol{\delta}_0$ ”, avec une division par 3 environ de $L'EQM_{v_{1.6}}$, une $EAM_{\alpha_{1.6}}$ inférieure à 0.1 et un $CE_{\alpha_{1.6}}\%$ égal à 72.9%. Les métriques relatives à l’estimation des délais correctifs sont également nettement améliorées, ce qui démontre que l’utilisation des RIMs cumulées exactes permet d’obtenir une estimation significativement meilleure des délais correctifs puis des RIMs qu’avec $\mathbf{v}_k = \boldsymbol{\delta}_0$. Les résultats obtenus en utilisant le critère considérant du filtrage gaussien (6.18) et $\mathbf{v}_k = \boldsymbol{\delta}_0$ ainsi que la qualité de l’estimation des RIMs illustrée pour ce critère en figure 6.13 suggèrent qu’une approche alternée pourrait permettre de les améliorer encore davantage. La figure 6.14 montre sur l’exemple de référence les résultats d’estimation de l’ensemble RIMs obtenue en utilisant le critère “ $\hat{\Delta}$ (6.18) , \mathbf{v}_k exactes”. Une légère amélioration de l’estimation est constatée par rapport à celle illustrée en figure 6.13, avec le critère “ $\hat{\Delta}$ (6.18) , $\mathbf{v}_k = \boldsymbol{\delta}_0$ ”.

	EAM $_{\Delta}$ (CE $_{\Delta}$ %)	EQM $_{v_{1:6}}$	EAM $_{\alpha_{1:6}}$ (CE $_{\alpha_{1:6}}$ %)
$\hat{\Delta} = \Delta$	0 (100%)	$1.20 \cdot 10^{-4}$	0.025 (97.1%)
$\hat{\Delta} = 0$	1.489 (40.8%)	$2.59 \cdot 10^{-2}$	0.495 (15.5%)
$\hat{\Delta}$ (6.16), $v_k = \delta_0$	1.441 (49.3%)	$8.08 \cdot 10^{-3}$	0.317 (29.1%)
$\hat{\Delta}$ (6.18), $v_k = \delta_0$	1.378 (55.9%)	$3.74 \cdot 10^{-3}$	0.159 (56.2%)
$\hat{\Delta}$ (6.18), v_k exactes	0.815 (74.1%)	$1.31 \cdot 10^{-3}$	0.098 (72.9%)

TABLEAU 6.2 – Erreurs absolues moyennes et pourcentages d’estimations correctes sur l’estimation des délais correctifs (EAM $_{\Delta}$ et CE $_{\Delta}$ %) et sur l’estimation des profils d’absorption (EAM $_{\alpha_{1:6}}$ et CE $_{\alpha_{1:6}}$ %) ainsi que erreurs quadratiques moyennes sur l’estimation des RIMs (EQM $_{v_{1:6}}$), pour différents modes d’estimation des délais correctifs suivis de la méthode du gradient conjugué. Les modes considérés sont l’estimation exacte (“ $\hat{\Delta} = \Delta$ ”, temps d’arrivée exacts), l’estimation nulle (“ $\hat{\Delta} = 0$ ”, temps d’arrivée erronés non corrigés), l’optimisation sans filtrage gaussien avec $v_k = \delta_0$ (“ $\hat{\Delta}$ (6.16), $v_k = \delta_0$ ”), l’optimisation avec filtrage gaussien avec $v_k = \delta_0$ (“ $\hat{\Delta}$ (6.18), $v_k = \delta_0$ ”) et l’optimisation avec filtrage gaussien en utilisant les RIMs cumulées exactes (“ $\hat{\Delta}$ (6.18), v_k exactes”).

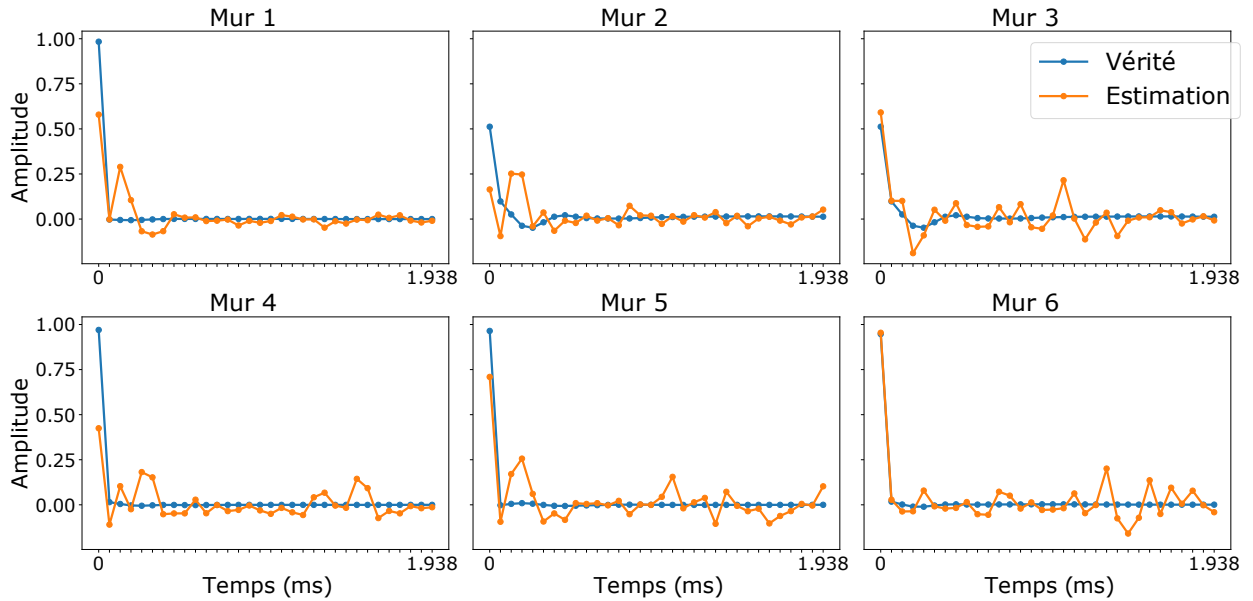


FIGURE 6.12 – Exemple des résultats d’estimation des RIMs d’une salle obtenus avec la méthode du gradient conjugué à partir de $M = 16$ RISs, pour une estimation des délais correctifs par optimisation sans filtrage gaussien (6.16) avec $v_k = \delta_0$.

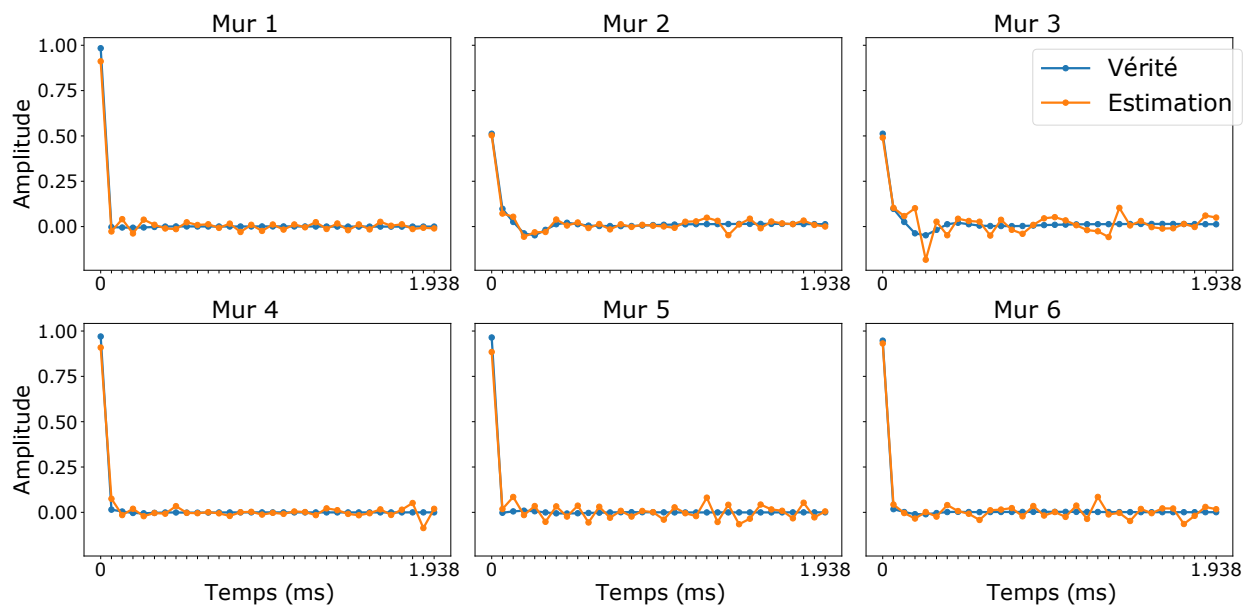


FIGURE 6.13 – Exemple des résultats d’estimation des RIMs d’une salle obtenus avec la méthode du gradient conjugué à partir de $M = 16$ RISs, pour une estimation des délais correctifs par optimisation avec filtrage gaussien (6.18) avec $\mathbf{v}_k = \delta_0$.

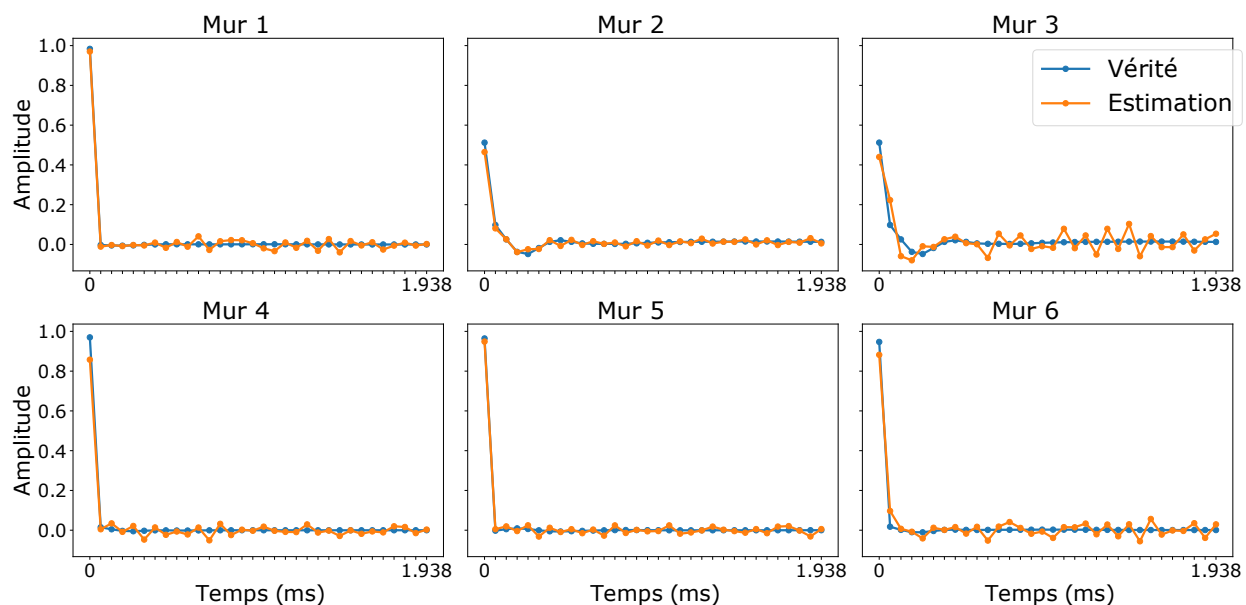


FIGURE 6.14 – Exemple des résultats d’estimation des RIMs d’une salle obtenus avec la méthode du gradient conjugué à partir de $M = 16$ RISs, pour une estimation des délais correctifs par optimisation avec filtrage gaussien (6.18) en utilisant les RIMs cumulées \mathbf{v}_k exactes.

La figure 6.15 présente l'évolution en fonction de la fréquence des métriques relatives à l'estimation des profils d'absorption ($EAM_{\alpha_{1:6}}$ 6.15a et $CE_{\alpha_{1:6}}\%$ 6.15b) pour l'utilisation des critères considérant du filtrage gaussien et respectivement des $\mathbf{v}_k = \boldsymbol{\delta}_0$ (courbes oranges) et les RIMs \mathbf{v}_k exactes (courbes vertes). Les lignes verticales en pointillés indiquent les frontières des bandes d'octave logarithmiques centrées sur les fréquences 0.125, 0.25, 0.5, 1, 2 et 4kHz.

Elle montre que les performances sont nettement détériorées au-delà de $f = 5657\text{Hz}$ (borne supérieure de la bande d'octave centrée sur 4kHz), en particulier pour le critère considérant des RIMs parfaitement rigides ($\mathbf{v}_k = \boldsymbol{\delta}_0$), avec une $EAM_{\alpha_{1:6}}$ égale à 0.235 et un $CE_{\alpha_{1:6}}\%$ égal à 41.1% (moins de la moitié des coefficients d'absorption sont correctement estimés). Ces résultats évoquent les oscillations "hautes fréquences" observables sur les figures 6.13 et 6.14 pour l'estimation de certaines des RIMs de la salle de référence en utilisant ces deux critères. Une piste d'amélioration de ces résultats peut être d'envisager l'utilisation d'une fonction de régularisation en association avec la méthode du gradient conjugué afin d'amoinrir ces oscillations. Toujours en utilisant le critère considérant du filtrage gaussien et des RIMs parfaitement rigides " $\hat{\Delta}$ (6.18), $\mathbf{v}_k = \boldsymbol{\delta}_0$ ", l' $EAM_{\alpha_{1:6}}$ et le $CE_{\alpha_{1:6}}\%$ pour l'ensemble des fréquences de Fourier comprises dans les 6 bandes d'octave considérées pour la simulation ($f \leq 5657\text{Hz}$) sont respectivement égaux à 0.128 et 62.5%. Les métriques associées à ce critère sont encore améliorées en restreignant l'étude aux 5 premières bandes d'octaves ($f \leq 2828\text{Hz}$) pour atteindre des résultats d'estimation très satisfaisant avec une $EAM_{\alpha_{1:6}}$ inférieure à 0.1 et 70% des coefficients d'absorption des murs correctement estimés. Par conséquent, il est attendu qu'une approche alternée efficace puisse permettre d'obtenir des performances intermédiaires à celles obtenues avec ces deux critères.

Finalement, cette étude démontre que l'amélioration significative des performances d'estimation des délais correctifs observée en section 6.3.5 pour l'utilisation du filtrage gaussien des signaux se répercute ensuite sur l'estimation des RIMs, l'améliorant aussi significativement. De plus, l'expérience "oracle" utilisant les RIMs cumulées exactes pour l'étape d'estimation des délais correctifs au lieu de les supposer parfaitement rigides montre des performances considérablement améliorées. Elle permet d'obtenir une indication sur les meilleurs résultats pouvant être espérés en alternant successivement l'estimation des délais correctifs puis celle des réponses des murs.

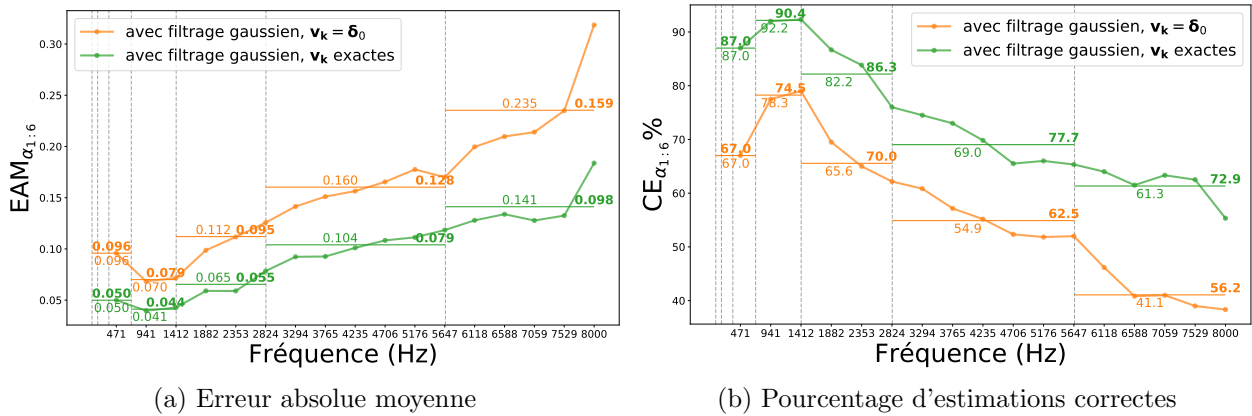


FIGURE 6.15 – Erreur absolue moyenne ($EAM_{\alpha_{1:6}}$ 6.15a) et pourcentage d'estimations correctes ($CE_{\alpha_{1:6}} \%$ 6.15b, erreur absolue inférieure à 0.1) relatifs aux profils d'absorption des murs en fonction de la fréquence pour l'utilisation des critères considérant du filtrage gaussien (6.18) et respectivement des RIMs parfaitement rigides ($\mathbf{v}_k = \delta_0$) (courbes oranges) et les RIMs \mathbf{v}_k exactes (courbes vertes). Les lignes verticales en pointillés indiquent les frontières des bandes d'octave logarithmique centrées sur les fréquences 0.125, 0.25, 0.5, 1, 2 et 4kHz. Les lignes horizontales et les valeurs en écriture normale indiquent la valeur de la métrique dans chacune des bandes d'octaves puis dans la plage de fréquences $[\sqrt{2} \cdot 4, 8]$ kHz. Les valeurs en gras indiquent la valeur de la métrique pour les fréquences inférieures à la borne supérieure de chaque bande d'octave puis à $\frac{f}{2} = 8$ kHz.

6.3.8 Influence des réponses des sources et des microphones

Cette sous-section présente les résultats d’expériences correspondant aux mêmes cas et paramètres que celles de la sous-section précédente (6.3.7) à l’exception que des réponses de sources et de microphones idéales sont considérées pour les simulations, c’est-à-dire qu’elles sont indépendantes de l’angle et de la fréquence et totalement neutres. Dans cette situation, les échos qui composent les RISs simulées sont de très courtes durées, à l’instar de l’exemple donné en figure 2.10a. C’est un scénario similaire à celui étudié dans le chapitre 5.

Ces résultats sont regroupés dans le tableau 6.3. Les meilleurs résultats obtenus pour ces expériences sont très similaires (légèrement améliorés) à leurs équivalents pour une simulation des RISs utilisant des mesures réelles de réponses en direction et en fréquence de sources et microphones, présentés dans le tableau 6.2 (cas “ $\hat{\Delta} = \Delta$ ”, “ $\hat{\Delta}$ (6.18), $\mathbf{v}_k = \delta_0$ ” et “ $\hat{\Delta}$ (6.18), \mathbf{v}_k exactes”). Ce constat confirme que cette approche permet de résoudre efficacement le problème inverse étudié pour un très haut niveau de réalisme de simulation rendant compte de l’effet considérable des sources et microphones avec l’utilisation de données réelles, et ce avec peu d’effet sur ses performances. Elle dépasse donc les limites de celle présentée dans le chapitre 5 et ouvre des perspectives d’application à des cas réels.

	EAM $_{\Delta}$ (CE $_{\Delta}$ %)	EQM $_{v_{1:6}}$	EAM $_{\alpha_{1:6}}$ (CE $_{\alpha_{1:6}}$ %)
$\hat{\Delta} = \Delta$	0 (100%)	$7.17 \cdot 10^{-5}$	0.022 (97.5%)
$\hat{\Delta}$ (6.18), $\mathbf{v}_k = \delta_0$	1.078 (65.1%)	$1.46 \cdot 10^{-3}$	0.134 (59.4%)
$\hat{\Delta}$ (6.18), \mathbf{v}_k exactes	0.818 (74.0%)	$7.23 \cdot 10^{-4}$	0.100 (69.0%)

TABLEAU 6.3 – Résultats d’expériences analogues à celles présentées en sous-section 6.3.6 à l’exception que les réponses de sources et de microphones considérées pour la simulation des données sont idéales. Erreurs absolues moyennes et pourcentages d’estimations correctes sur l’estimation des délais correctifs (EAM $_{\Delta}$ et CE $_{\Delta}$ %) et sur l’estimation des profils d’absorption (EAM $_{\alpha_{1:6}}$ et CE $_{\alpha_{1:6}}$ %) ainsi que erreurs quadratiques moyennes sur l’estimation des RIMs (EQM $_{v_{1:6}}$), pour différents modes d’estimation des délais correctifs suivis de la méthode du gradient conjugué. Les modes considérés sont l’estimation exacte (“ $\hat{\Delta} = \Delta$ ”, temps d’arrivée exacts), l’optimisation avec filtrage gaussien avec $\mathbf{v}_k = \delta_0$ (“ $\hat{\Delta}$ (6.18), $\mathbf{v}_k = \delta_0$ ”) et l’optimisation avec filtrage gaussien en utilisant les RIMs cumulées exactes (“ $\hat{\Delta}$ (6.18), \mathbf{v}_k exactes”).

6.3.9 Influence du nombre d'observations

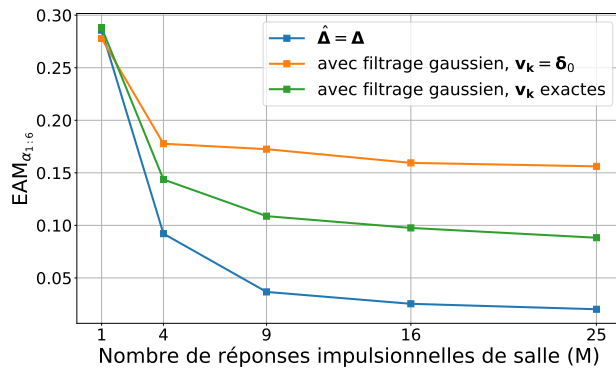
Cette sous-section reprends les cas et paramètres considérés dans les expériences présentées en sous-section (6.3.7) et étudie l'influence du nombre de RISs utilisées pour l'estimation, noté M . Les différentes valeurs de M considérées correspondent au nombre total de combinaisons source-microphone possibles pour un nombre égal de sources et de microphones allant de 1 à 5 ($M \in \{1, 4, 9, 16, 25\}$).

Le tableau 6.4 présente pour les différents cas considérés l'évolution de l'erreur quadratique moyenne d'estimation des RIMs ($\text{EQM}_{v_{1:6}}$) en fonction de M . La figure 6.16 représente pour chacun de ces cas l'évolution en fonction de M de l'erreur absolue moyenne d'estimation des profils d'absorption $\text{EAM}_{\alpha_{1:6}}$ et le pourcentage d'estimations correctes correspondant $\text{CE}_{\alpha_{1:6}}\%$ (erreur absolue inférieure à 0.1).

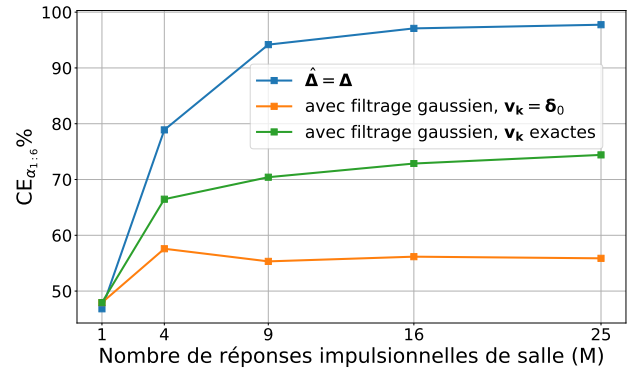
Les résultats du tableau 6.4 montrent que pour chacun des cas étudiés, l'augmentation de M entraîne une diminution systématique de l' $\text{EQM}_{v_{1:6}}$. La figure 6.16 révèle que ce constat également vérifié en ce qui concerne les métriques relatives à l'estimation des profils d'absorption $\text{EAM}_{\alpha_{1:6}}$ et $\text{CE}_{\alpha_{1:6}}\%$. Lorsque l'estimation des délais correctifs est suffisamment précise, l'augmentation du nombre RISs permet d'accroître la redondance d'information pour l'estimation des RIMs et ainsi d'améliorer les résultats.

	M				
	1	4	9	16	25
$\hat{\Delta} = \Delta$	4.19	$1.63 \cdot 10^{-2}$	$5.02 \cdot 10^{-4}$	$1.20 \cdot 10^{-4}$	$7.23 \cdot 10^{-5}$
$\hat{\Delta}$ (6.18), $\mathbf{v}_k = \delta_0$	7.77	$2.63 \cdot 10^{-1}$	$1.12 \cdot 10^{-2}$	$3.74 \cdot 10^{-3}$	$2.49 \cdot 10^{-3}$
$\hat{\Delta}$ (6.18), \mathbf{v}_k exactes	134	$9.46 \cdot 10^{-2}$	$3.66 \cdot 10^{-3}$	$1.31 \cdot 10^{-3}$	$8.14 \cdot 10^{-4}$

TABLEAU 6.4 – Erreurs quadratiques moyennes sur l'estimation des RIMs ($\text{EQM}_{v_{1:6}}$) en fonction du nombre de RISs utilisées pour l'estimation (M) pour différents modes d'estimation des délais correctifs suivis de la méthode du gradient conjugué. Les modes considérés sont l'estimation exacte (" $\hat{\Delta} = \Delta$ ", temps d'arrivée exacts), l'optimisation avec filtrage gaussien avec $\mathbf{v}_k = \delta_0$ (" $\hat{\Delta}$ (6.18), $\mathbf{v}_k = \delta_0$ ") et l'optimisation avec filtrage gaussien en utilisant les RIMs cumulées exactes (" $\hat{\Delta}$ (6.18), \mathbf{v}_k exactes").



(a) Erreur absolue moyenne



(b) Pourcentage d'estimations correctes

FIGURE 6.16 – Erreurs absolues moyennes et pourcentages d'estimations correctes sur l'estimation des profils d'absorption ($EAM_{\alpha_{1:6}}$ et $CE_{\alpha_{1:6}}\%$) en fonction du nombre de RISs utilisées pour l'estimation (M) pour différents modes d'estimation des délais correctifs suivis de la méthode du gradient conjugué. Les modes considérés sont l'estimation exacte ($\hat{\Delta} = \Delta$, temps d'arrivée exacts), l'optimisation avec filtrage gaussien avec $\mathbf{v}_k = \delta_0$ ($\hat{\Delta}$ (6.18), $\mathbf{v}_k = \delta_0$) et l'optimisation avec filtrage gaussien en utilisant les RIMs cumulées exactes ($\hat{\Delta}$ (6.18), \mathbf{v}_k exactes).

6.3.10 Influence du niveau de bruit géométrique

Cette sous-section présente les résultats d'expériences correspondant aux mêmes cas et paramètres que celles de la sous-section 6.3.7, à l'exception que l'incertitude géométrique est modélisée cette fois-ci sur les positions des sources, des microphones mais aussi des murs par des erreurs aléatoires indépendantes de loi $\mathcal{N}(0, \sigma_{\text{geo}}^2)$, avec $\sigma_{\text{geo}} = 0.02\text{m}$.

Il s'agit de la modélisation de l'incertitude géométrique employée pour les expériences du chapitre 5, qui rend compte du cumul des erreurs sur les positions des murs lors du calcul des positions de sources-images avec la méthode des sources-images. Le niveau d'incertitude correspondant est nettement plus élevée que pour les expériences précédemment effectuées dans ce chapitre.

Pour toute source-image $\mathbf{k} \in \mathcal{K}$, le temps d'arrivée erroné est obtenu en ajoutant au temps d'arrivée exact un délai aléatoire de loi $\mathcal{N}(0, \sigma_{q(\mathbf{k})}^2)$, avec $\sigma_{q(\mathbf{k})}$ qui dépend de l'ordre de réflexion $q(\mathbf{k})$ comme dans (5.8). La valeur de Δ_{max} considérée correspond à l'intervalle de confiance à 95% de la distribution normale des délais qui s'appliquent aux sources-images d'ordre égal à 2. Dans cette situation, $\Delta_{\text{max}} = 0.466\text{ms}$ soit 7.46 échantillons, équivalent à une distance parcourue par le son de 16cm.

Le tableau 6.5 présente les résultats obtenus pour les différents cas considérés avec cette modélisation de l'incertitude géométrique. Il montre une dégradation des résultats obtenus pour l'ensemble des expériences conduites. Cependant, il est possible de considérer que pour les deux critères avec filtrage gaussien (“ $\hat{\Delta}$ (6.18), $\mathbf{v}_{\mathbf{k}} = \delta_0$ ” et “ $\hat{\Delta}$ (6.18), $\mathbf{v}_{\mathbf{k}}$ exactes”), les métriques d'évaluation relatives à l'estimation des RIMs et des profils d'absorption restent comparables à celles des expériences de la sous-section 6.3.7. Ce constat suggère que la méthode parvient à compenser correctement les délais d'une partie des échos les plus utiles pour l'estimation des RIMs par la méthode du gradient

	EAM $_{\Delta}$ (CE $_{\Delta}$ %)	EQM $_{v_{1:6}}$	EAM $_{\alpha_{1:6}}$ (CE $_{\alpha_{1:6}}$ %)
$\hat{\Delta} = \Delta$	0 (100%)	$1.22 \cdot 10^{-4}$	0.025 (97.1%)
$\hat{\Delta}$ (6.18), $\mathbf{v}_{\mathbf{k}} = \delta_0$	4.041 (32.5%)	$5.02 \cdot 10^{-3}$	0.179 (52.4%)
$\hat{\Delta}$ (6.18), $\mathbf{v}_{\mathbf{k}}$ exactes	3.744 (38.4%)	$4.18 \cdot 10^{-3}$	0.139 (61.6%)

TABLEAU 6.5 – Résultats d'expériences analogues à celles présentées en sous-section 6.3.6 à l'exception qu'une modélisation plus réaliste des effets de l'incertitude géométrique sur les temps d'arrivée erronés est effectuée. Erreurs absolues moyennes et pourcentages d'estimations correctes sur l'estimation des délais correctifs (EAM $_{\Delta}$ et CE $_{\Delta}$ %) et sur l'estimation des profils d'absorption (EAM $_{\alpha_{1:6}}$ et CE $_{\alpha_{1:6}}$ %) ainsi que erreurs quadratiques moyennes sur l'estimation des RIMs (EQM $_{v_{1:6}}$), pour différents modes d'estimation des délais correctifs suivis de la méthode du gradient conjugué. Les modes considérés sont l'estimation exacte (“ $\hat{\Delta} = \Delta$ ”, temps d'arrivée exacts), l'optimisation avec filtrage gaussien avec $\mathbf{v}_{\mathbf{k}} = \delta_0$ (“ $\hat{\Delta}$ (6.18), $\mathbf{v}_{\mathbf{k}} = \delta_0$ ”) et l'optimisation avec filtrage gaussien en utilisant les RIMs cumulées exactes (“ $\hat{\Delta}$ (6.18), $\mathbf{v}_{\mathbf{k}}$ exactes”).

conjugué. Il est possible d'intuiter qu'il s'agit d'échos d'ordre relativement faible. En effet, compte tenu de la modélisation de l'incertitude géométrique, les délais correspondant sont moins importants que ceux appliqués aux échos d'ordre élevé. De plus, ils apparaissent généralement dans les premiers instants des RISs où la densité des échos est plus faible.

6.4 Conclusion et perspectives

Dans ce chapitre, le problème inverse de l'estimation des réponses impulsionnelles des murs d'une salle à partir de réponses impulsionnelles de salle, des mesures associées des réponses des sources et microphones et des paramètres géométriques aux erreurs près de l'ordre de quelques centimètres a été étudié.

Pour résoudre ce problème, une approche d'optimisation dans le domaine temporel portant sur la compensation des erreurs induites par l'incertitude géométrique sur les temps d'arrivée des échos ainsi que sur l'estimation des réponses impulsionnelles des murs est proposée. Elle repose sur une extension du modèle de réponse impulsionnelle de salle issu de la méthode des sources-images qui intègre la modélisation des erreurs sur le calcul des temps d'arrivée exacts des échos. Elle considère d'abord l'estimation de ces délais par optimisation d'un critère de reconstruction des réponses impulsionnelles de salle avec l'algorithme ADAM en supposant des RIMs parfaitement rigides puis l'estimation des réponses impulsionnelles des murs (cumulées) avec la méthode du gradient conjugué.

Une étude détaillée de ses performances est effectuée à partir d'expériences conduites dans un cadre simulé réaliste, utilisant des mesures réelles de réponses en direction et en fréquence de sources et de microphones pour la simulation, et modélisant les incertitudes géométriques ainsi que le bruit de mesure. L'analyse des résultats de ces expériences montre que l'évaluation des profils d'absorption calculés à partir des réponses impulsionnelles des murs estimées permet de mieux rendre compte de leur qualité. Les profils d'absorption calculés sont décrits sur une échelle linéaire de 17 fréquences entre 0 et 8000Hz.

Cette étude démontre que lorsque les temps d'arrivée des échos sont considérés connus exactement, la méthode du gradient conjugué permet d'estimer les réponses impulsionnelles des murs avec une excellente précision. Lorsque l'incertitude géométrique est modélisée par une erreur sur la position de chaque microphone et source-image de loi normale centrée d'écart-type égal à 2cm, elle prouve qu'un filtrage gaussien préalable des signaux contribue à améliorer significativement les résultats d'estimation des délais correctis. Elle montre également que ce procédé permet ensuite d'améliorer considérablement l'estimation des réponses impulsionnelles des murs. En utilisant 16 réponses impulsionnelles de salle pour des emplacements arbitraires de sources et de microphones, l'estimation des réponses des murs permet d'atteindre une erreur absolue moyenne relative aux profils d'absorption égale à 0.159 avec 56% des coefficients estimés avec une erreur inférieure à 0.1. Les valeurs de ces deux métriques passent respectivement à 0.128 et 62.5% pour les fréquences inférieures à 5657Hz et atteignent 0.095 et 70% pour les fréquences inférieures à 2828Hz.

Des expériences spécifiques démontrent que les résultats obtenus sont améliorés par une augmentation du nombre d'observations et qu'ils restent relativement stables face à l'effet source-

microphone obtenu à partir des mesures réelles ainsi qu'à une modélisation plus réaliste et plus intense de l'incertitude géométrique.

Une expérience utilisant les réponses exactes des murs pour l'étape d'estimation des délais correctis au lieu de les supposer parfaitement rigides a été menée. Son objectif était d'identifier les meilleurs résultats pouvant être espérés en alternant successivement l'estimation des délais correctifs puis celle des réponses des murs. Elle révèle que les performances d'estimation obtenues sont considérablement améliorées avec une erreur absolue moyenne relative aux profils d'absorption entre 0 et 8000Hz égale à 0.098 et 72.9% des coefficients estimés avec une erreur inférieure à 0.1.

Le temps d'exécution moyen de l'approche est de 9min4s avec un processeur Intel i7-10610U composé de 8 coeurs de fréquences égales à 1.80 GHz et de 6min8s en utilisant deux cartes graphiques NVIDIA GeForce RTX 2080 Ti (calculé sur 100 salles). L'approche d'estimation proposée dans ce chapitre est donc beaucoup plus rapide que celle par apprentissage supervisée virtuel présentée au chapitre 4, pour laquelle la durée de simulation des données nécessaires à l'entraînement pour une unique salle est estimée à 39h38min. De plus, cette approche démontre son efficacité pour des réponses impulsionnelles de salle dont l'ensemble des échos sont de durées considérables et superposés. Elle dépasse donc les limites de celle présentée dans le chapitre 5 qui n'est pas applicable dans une telle situation. Enfin, elle ouvre des perspectives d'application à des cas réels. En effet, elle obtient des résultats d'estimation encourageant pour une application à des données hybrides très similaires aux données réelles, synthétisées à partir de signaux mesurés. Enfin, elle définit un estimateur qui ne dépend d'aucune donnée d'apprentissage, contrairement à l'approche du chapitre 4 pour laquelle se posent des questions de généralisation de l'apprentissage supervisé virtuel à l'estimation sur des données réelles.

Des travaux futurs pourront porter sur le développement d'une approche alternée à partir des 2 étapes d'optimisation définies dans ce chapitre. Ils pourront également tenter d'améliorer encore le niveau de réalisme en permettant cette fois la simulation des données en utilisant des mesures réelles des réponses impulsionnelles de murs. Enfin, ils pourront s'intéresser à l'utilisation de fonctions de régularisation pour améliorer l'estimation des réponses impulsionnelles des murs par la méthode du gradient conjugué, potentiellement en basant ces fonctions sur des modèles traditionnels de physique des matériaux à l'instar de [Mondet et al., 2020](#).

6.5 Discussions

Tout d’abord, dans notre approche, la compensation de l’incertitude géométrique est effectuée en considérant un problème d’optimisation paramétré avec les erreurs sur les temps d’arrivée (délais correctifs) plutôt que celles sur les paramètres géométriques directement. Ce choix de modélisation permet d’axer cette compensation sur la concordance entre le signal (la RIS) et son modèle pour aller au-delà des hypothèses posées par la méthode des sources-images de murs parfaitement plans et de sources et microphones ponctuels. En réalité, les murs présentent des courbures et les sources et microphones sont des volumes dont il peut être complexe voire impossible d’identifier l’unique centre acoustique servant de référence pour la position. De ce fait, il existe une limite sur la possibilité d’estimer très précisément une géométrie réelle avec cette paramétrisation simplifiée.

De plus, notre approche est formulée exclusivement dans le domaine temporel. La résolution dans le domaine de Fourier est communément utilisée en traitement du signal car elle permet de transformer les convolutions en multiplications pour une simplification des équations et une accélération des calculs. Cependant, elle ne permet pas de réduire la taille du système par troncature temporelle des signaux, comme cela est fait dans nos expériences. En effet, la validité du théorème de convolution nécessite une fenêtre de Fourier grande devant la durée des échos. Cependant, dans le cas général de sources et microphones à dépendance directionnelle et fréquentielle, la longueur des échos et leur niveau de superposition sont tels que cela implique de considérer la TFD de la RIS complète. Lors d’expériences préliminaires à ce chapitre de thèse, nous nous sommes intéressé à l’estimation des RIMs par optimisation d’un critère dans le domaine de Fourier complexe avec l’*algorithme des directions alternées* (Boyd et al., 2011). Ces expériences ont démontré la capacité de la méthode définie à résoudre efficacement le problème pour des RISs simulées jusqu’au deuxième ordre de réflexion. Cependant, elles ont révélé que les résultats obtenus par la méthode s’effondraient lorsque le système n’incluait pas la totalité des échos présents dans le signal ainsi que lorsque la partie tardive des RISs était écartée par troncature, comme par multiplication par une fenêtre d’apodisation. Enfin, elles ont montré qu’à partir du troisième ordre de réflexion, la taille du système complet était telle que les ressources de calcul à disposition devenaient insuffisantes pour supporter l’exécution de la méthode.

Enfin, des efforts considérables ont été produits afin d’implémenter une approche pour l’estimation (mono-observation) des délais correctifs avec des réseaux de neurones Transformers. L’ambition était de tirer parti de la capacité de ces modèles à gérer des séquences de tailles variables en entrée et en sortie, à exploiter l’information des positions des éléments dans une séquence, ainsi qu’à tenir compte du “contexte” dans le signal temporel pour la prédiction de chaque délai correctif. Cependant, nous ne sommes pas parvenu à obtenir de résultats avec cette piste de recherche. L’approche par optimisation proposée dans ce chapitre nous a permis d’aller au-delà des obstacles que nous avons rencontrés avec les réseaux de neurones Transformers.

7.1 Bilan

Ce manuscrit présente les travaux effectués au cours de cette thèse intitulée “Apprentissage automatique et optimisation pour la détermination des propriétés acoustiques d’une salle à partir de signaux audio”. Cette thèse s’intéresse en particulier à l’estimation des propriétés d’absorption de chaque mur d’une salle à partir de plusieurs réponses impulsionnelles de salle acquises pour des positions libres de sources et de microphones ainsi que des paramètres géométriques et de mesures des réponses des appareils associés. Plusieurs méthodes sont proposées pour la résolution de ce problème inverse ouvert et complexe en visant progressivement une description plus fine des propriétés d’absorption des murs, une application à des données plus réalistes et une plus grande simplicité de mise en oeuvre.

Estimation pour une géométrie spécifique par apprentissage supervisé virtuel :

Une première approche est proposée au chapitre 4 pour l’estimation des profils d’absorption des murs d’une salle décrits dans six bandes d’octaves à partir d’une seule réponse impulsionnelle de salle et de la connaissance approximative des paramètres géométriques associés. Cette estimation est réalisée avec un réseau de neurones profond entraîné sur une base de données simulée et annotée de réponses impulsionnelles de salle correspondant à cette même géométrie spécifique (apprentissage supervisé virtuel). Cette méthodologie est adoptée en raison de la difficulté de renseigner l’information géométrique en entrée des réseaux de neurones.

Cette approche est évaluée dans un cadre simulé considérant la modélisation des réflexions spéculaires et diffuses, d’erreurs sur la mesure des paramètres géométriques de l’ordre de quelques centimètres et du bruit de mesure. Les réponses des sources et des microphones sont considérées idéales, c’est-à-dire indépendantes de la direction et de la fréquence.

Les expériences conduites montrent qu’un réseau neuronal convolutif entraîné par cette approche permet d’obtenir une estimation fiable et robuste des coefficients d’absorption en bande d’octave, jusqu’au plus haut niveau de réalisme atteignable en combinant les différents types de

distorsions considérées. De plus, elles confirment l'influence importante des erreurs géométriques sur les résultats d'estimation et donc l'importance de leur prise en compte pour le développement de méthodes robustes en vue d'applications réelles. Enfin, elles révèlent que la partie précoce (son direct et premiers échos) est la composante la plus importante des réponses impulsionnelles de salle pour l'estimation des profils d'absorption des murs par cette approche.

Néanmoins, cette approche présente plusieurs limites. Principalement, elle nécessite pour chaque nouvelle géométrie considérée de renouveler complètement la simulation d'une base de données d'apprentissage et l'entraînement d'un réseau de neurones profond, qui constituent deux étapes extrêmement chronophages (la durée de l'étape de simulation seule est estimée à 39h38min). De plus, elle soulève les problématiques classiques de l'apprentissage automatique et en particulier virtuellement supervisé, notamment à propos de la généralisation de l'apprentissage sur des données synthétiques à des tests sur des données réelles, ou encore de la diversité et de la représentativité du jeu d'entraînement.

Ces limites motivent pour la suite de la thèse le développement d'approches "informées" par la géométrie, robustes aux erreurs géométriques et définissant un estimateur qui ne dépend d'aucune donnée synthétique.

Estimation informée par la géométrie par optimisation sur les spectrogrammes des réponses impulsionnelles de salle :

Une deuxième approche est proposée au chapitre 5 pour l'estimation des profils d'absorption des murs d'une salle, décrits sur une échelle linéaire des fréquences, à partir des spectrogrammes de plusieurs réponses impulsionnelles de salle et de la connaissance approximative des paramètres géométriques associés. À chaque fréquence, une fonction objectif des paramètres d'intérêt est optimisée sur un ensemble de fenêtres temporelles des spectrogrammes supposées inclure entièrement un unique écho afin d'éviter les difficultés causées par leur superposition dans le domaine énergétique. Ces fenêtres dites pertinentes sont sélectionnées à partir de l'information géométrique approximative.

Cette approche est évaluée dans un cadre simulé considérant la modélisation des réflexions spéculaires, d'erreurs sur la mesure des paramètres géométriques de l'ordre de quelques centimètres et du bruit de mesure. Les réponses des sources et des microphones sont considérées idéales, c'est-à-dire indépendantes de la direction et de la fréquence.

L'étude simulée démontre que la version probabiliste et robuste de l'approche basée sur l'algorithme RANSAC permet d'estimer précisément les coefficients d'absorption des murs de la salle au-dessus de 400 Hz, ce en présence d'erreurs géométriques et de bruit de mesure. De plus, elle révèle que l'augmentation du nombre de réponses impulsionnelles de salle utilisées pour l'estima-

tion permet de réduire significativement les erreurs d'estimation et d'améliorer la robustesse de l'approche.

Cependant, le fait que cette approche repose sur la possibilité d'isolement des échos au sein des réponses impulsionnelles de salle constitue une limitation importante de son champ d'application. En effet, dans la réalité, l'effet source-microphone tend généralement à augmenter la durée des échos de telle sorte qu'aucun ne puisse être isolé par un fenêtrage temporel. Ce phénomène est lié à la variation en fréquence des réponses des appareils. Une autre limite du champ d'application de cette approche est qu'elle ne permet pas en l'état de traiter la variation en direction des réponses des sources et des microphones utilisés.

Ces conclusions orientent les prochains travaux vers le développement d'une approche bénéficiant des mêmes atouts que celle-ci et considérant l'omniprésence des superpositions entre les échos dans les réponses impulsionnelles de salle ainsi que la dépendance directionnelle et fréquentielle des réponses des appareils utilisés.

Approche par optimisation dans le domaine temporel :

Enfin, une troisième approche est proposée au chapitre 6 pour l'estimation des réponses impulsionnelles des murs d'une salle à partir de réponses impulsionnelles de salle, des paramètres de la géométrie connus approximativement et des mesures des réponses en direction et en fréquence des sources et des microphones. La réponse impulsionnelle d'un mur constitue une description plus riche de ses propriétés acoustiques que le profil d'absorption car elle renseigne son effet sur la phase du son incident, en plus de celui sur son énergie.

L'estimation est réalisée par une approche d'optimisation dans le domaine temporel définissant deux sous-problèmes complémentaires. Le premier porte sur l'estimation des erreurs sur les temps d'arrivée des échos calculés avec la méthode des sources images causées par l'incertitude géométrique en optimisant sous contraintes un critère défini dans le domaine temporel continu avec l'algorithme ADAM. Ce critère est défini à partir d'une extension du modèle de réponse impulsionnelle de salle issu de la méthode des sources images permettant de rendre compte de ces délais sur les temps d'arrivée des échos et suppose initialement des murs parfaitement rigides. Le second sous-problème s'intéresse à l'estimation des réponses impulsionnelles des murs de la salle, formulée comme un problème d'optimisation linéaire et multi-observation dans le domaine temporel, avec la méthode du gradient conjugué.

Une étude détaillée des performances de l'approche est effectuée dans un cadre simulé hautement réaliste utilisant des données hybrides générées à partir de réelles mesures de réponses en direction et en fréquence de sources et de microphones, ainsi que modélisant les incertitudes géométriques et le bruit de mesure. Les superpositions entre les échos sont omniprésentes au sein des réponses

impulsionnelles de salle générées.

Cette étude démontre que lorsque les temps d'arrivée des réflexions sont considérés connus exactement, la méthode du gradient conjugué permet d'estimer les réponses impulsionnelles des murs avec une excellente précision. Lorsque l'incertitude géométrique est modélisée par une erreur de l'ordre de quelques centimètres sur la position de chaque microphone et source image, elle prouve qu'un filtrage gaussien préalable des signaux contribue à améliorer significativement les résultats d'estimation des erreurs sur les temps d'arrivée des échos. Elle montre également que cette amélioration bénéficie ensuite considérablement à l'estimation des réponses impulsionnelles des murs, permettant ainsi d'obtenir des résultats très satisfaisants. Des expériences spécifiques démontrent que les résultats obtenus sont améliorés par une augmentation du nombre d'observations et qu'ils restent relativement stables face à l'effet source-microphone obtenu à partir des mesures réelles ainsi qu'à une modélisation plus réaliste et intense de l'incertitude géométrique.

Une expérience utilisant les réponses impulsionnelles exactes des murs pour l'estimation des erreurs sur les temps d'arrivée des échos au lieu de les supposer parfaitement rigides est menée. Les performances d'estimation des délais correctifs ainsi que celles des réponses impulsionnelles des murs en sont nettement améliorées, indiquant ainsi qu'une approche alternée issue des deux sous-problèmes d'optimisation présentés pourrait permettre une résolution encore meilleure du problème inverse étudié.

Finalement, cette approche permet de surpasser les limites des deux précédentes. Premièrement, il s'agit d'une approche informée par la géométrie et capable de gérer les erreurs géométriques. De plus, elle démontre son efficacité pour des réponses impulsionnelles de salle dont l'ensemble des échos sont superposés et permet de gérer la dépendance en direction et en fréquence des réponses des sources et des microphones.

7.2 Perspectives de travaux

De nombreux défis restent à relever afin d'améliorer la fiabilité des approches développées en vue d'applications réelles, d'atteindre un niveau encore supérieur de description des propriétés acoustiques des murs et d'étendre le champ d'application des méthodes à des cas plus complexes.

Premièrement, par soucis de réalisme, les expériences menées dans cette thèse portent sur des données simulées en modélisant l'action des murs sur les ondes acoustiques par des filtres à phase minimale conçus en utilisant des mesures en laboratoire de profils d'absorption en bandes d'octave de matériaux typiques en acoustiques des salles. Cependant, il apparaît que la quantité d'information utilisée pour synthétiser ces réponses reste relativement restreinte et que la modélisation employée est assez simpliste pour prétendre atteindre une représentation hautement réaliste. Des travaux futurs pourront donc porter sur l'amélioration et l'enrichissement de la synthèse des réponses impulsionnelles des murs pour la simulation, voire sur l'utilisation de mesures réelles afin d'évaluer la fiabilité des approches en vue d'application à des cas concrets.

De plus, les approches proposées dans cette thèse ainsi que les données simulées utilisées pour leurs évaluations considèrent des surfaces aux propriétés acoustiques identiques quelque soit l'angle d'incidence du son. Cependant, il est reconnu que la réponse impulsionnelle d'un mur dépend de l'angle d'incidence et que sa variation peut être significative pour un nombre important de matériaux. Des perspectives de travaux pourront donc concerner la modélisation de la dépendance angulaire des réponses impulsionnelles des murs (incidence oblique) pour la simulation ainsi que pour l'estimation.

Aussi, les expériences conduites dans cette thèse portent sur des données simulées correspondant à des salles en forme de pavé droit uniquement, vides de tout objet et dont les murs sont uniformes acoustiquement. Par conséquent, des extensions des simulateurs, des approches proposées dans cette thèse ainsi que le développement de nouvelles techniques permettant d'étendre la complexité des configurations de salle traitées pourront être envisagées.

Pour finir, des travaux tels que [Dokmanić et al., 2013](#) ou [Sprunck et al., 2022](#) investiguent la résolution du problème inverse dual portant sur l'estimation des paramètres géométriques à partir de réponses impulsionnelles de salle et *d'a priori* sur les propriétés acoustiques des murs. À plus long terme, des recherches pourront tenter de combiner des approches de résolution pour ces deux problèmes complémentaires dans le but de les unifier, de relâcher certains a priori voire de s'en affranchir.

7.3 Perspectives d'applications

Les travaux accomplis à l'issue de cette thèse contribuent à simplifier le processus du diagnostic acoustique d'une salle, à le détailler à l'échelle des propriétés acoustiques des murs individuels, à l'enrichir avec une description fine de leur effet sur l'énergie, la phase et la fréquence du son, ainsi qu'à améliorer sa fiabilité.

À terme, des avancées supplémentaires dans cette direction de recherche pourraient conduire à rendre possible l'obtention rapide et automatique de l'ensemble des paramètres acoustiques d'une salle permettant de modéliser et de simuler son champ sonore avec une très grande précision. Il serait alors possible de mieux comprendre le champ sonore dans une salle, d'agir dessus efficacement, de le répliquer ou encore d'utiliser cette connaissance pour servir d'autres applications en traitement du signal audio. L'idée d'un tel progrès laisse imaginer des évolutions et de nouvelles applications en lien avec l'acoustique des salles dont voici quelques exemples.

De telles avancées pourraient contribuer à transformer la conception des solutions de réhabilitation acoustique en facilitant l'étude de l'évolution du champ sonore dans une salle provoquée par la modification des propriétés acoustiques d'un de ses murs, notamment avec des tests d'écoute très haute fidélité.

Une telle précision de connaissance des propriétés acoustiques d'une salle pourraient bénéficier à la résolution d'autres applications phares en traitement du signal audio. Par exemple, elle pourrait contribuer à améliorer les performances de tâches telles que la diarisation, la séparation ou le rehaussement d'un signal de parole (par exemple dans une salle de réunion). Une telle maîtrise de l'acoustique d'une salle pourrait aussi permettre de compenser totalement ou de moduler son effet sur des signaux qui y sont enregistrés.

De tels travaux pourraient bénéficier au domaine de la Réalité Augmentée Audio. Par exemple, ils pourraient permettre d'améliorer la cohérence d'un contenu audio produit dans le casque d'un utilisateur avec l'environnement acoustique réel dans lequel il se trouve.

Enfin, une telle percée pourrait rendre possible la création de "jumeaux numériques" acoustiques de salle, qui constituerait une réplique numérique de l'état courant des propriétés acoustiques et du champ sonore de lieux historiques ou d'exception. Une telle innovation pourrait trouver des applications en lien avec notamment la conservation et l'accès au patrimoine culturel, la réalité virtuelle ou encore la conception d'effets de réverbération numérique réalistes pour la production audio.

8.1 Notions de base en traitement du signal

L'objectif de cette section est de rappeler très succinctement les notions et formules de base en traitement du signal intervenant dans les approches présentées dans ce manuscrit de thèse.

8.1.1 Signaux continus

Transformée de Fourier :

Soit $x(t)$ un signal temporel continu. Sa représentation équivalente dans le domaine fréquentiel, notée $\underline{x}(f)$, est obtenue par application de la Transformée de Fourier $\mathcal{F}\{\cdot\}$ définie par :

$$F\{x\}(f) = \underline{x}(f) = \int_{-\infty}^{\infty} x(t) \cdot e^{-j2\pi ft} dt.$$

Transformée de Fourier Inverse :

Le signal temporel $x(t)$ peut être retrouvé à partir de sa représentation fréquentielle $\underline{x}(f)$ par application de la Transformée de Fourier Inverse $\mathcal{F}^{-1}\{\cdot\}$ définie par :

$$\mathcal{F}^{-1}\{\underline{x}\}(t) = x(t) = \int_{-\infty}^{\infty} \underline{x}(f) \cdot e^{j2\pi ft} df.$$

Produit de convolution :

Soit $x(t)$ et $y(t)$ deux signaux temporels continus. Le produit de convolution " $x * y$ " des deux signaux est défini comme :

$$(x * y)(t) = (y * x)(t) = \int_{-\infty}^{\infty} x(\tau) \cdot y(t - \tau) d\tau.$$

Il représente la réponse du système linéaire caractérisé par $x(t)$ à l'entrée $y(t)$, ou vice versa.

Théorème de convolution :

Si $x(t)$ et $y(t)$ sont deux signaux temporels continus de carré intégrable, alors le théorème de convolution établit une relation entre le produit de convolution dans le domaine temporel et la multiplication dans le domaine fréquentiel par l'intermédiaire des transformées directe et inverse de Fourier :

$$\begin{cases} \mathcal{F}\{x * y\}(f) &= (\underline{x} \cdot \underline{y})(f) \\ \mathcal{F}^{-1}\{\underline{x} \cdot \underline{y}\}(t) &= (x * y)(t). \end{cases} \quad (8.1)$$

8.1.2 Signaux discrets

Transformée de Fourier Discrète :

Soit $x[t]$ un signal temporel discret. Sa représentation équivalente dans le domaine fréquentiel, notée $\underline{x}[f]$, est obtenue par application de la Transformée de Fourier Discrète TFD $\{.\}$ définie par :

$$\text{TFD}\{x\}[f] = \underline{x}[f] = \sum_{t=-\infty}^{\infty} x[t] \cdot e^{-j2\pi ft}.$$

Transformée de Fourier Discrète Inverse :

Le signal temporel discret $x[t]$ peut être retrouvé à partir de sa représentation fréquentielle discrète $\underline{x}[f]$ par application de la Transformée de Fourier Inverse Discrète TFD $^{-1}\{.\}$ définie par :

$$\text{TFD}^{-1}\{\underline{x}\}[t] = x[t] = \frac{1}{N} \sum_{f=-\infty}^{\infty} \underline{x}[f] \cdot e^{j2\pi ft}.$$

Produit de convolution :

Soit $x[t]$ et $y[t]$ deux signaux temporels discrets. Le produit de convolution " $x * y$ " des deux

signaux discrets est défini comme :

$$(x * y)[t] = \sum_{\tau=-\infty}^{\infty} x[\tau] \cdot y[t - \tau].$$

Il représente la réponse du système linéaire caractérisé par $x[t]$ à l'entrée $y[t]$, ou vice versa.

Théorème de convolution :

Le théorème de convolution appliqué au signaux discret $x[t]$ et $y[t]$ établit les relations suivantes :

$$\begin{cases} \text{TFD}\{x * y\}[f] &= (\underline{x} \cdot \underline{y})[f] \\ \text{TFD}^{-1}\{\underline{x} \cdot \underline{y}\}[t] &= (x * y)[t]. \end{cases} \quad (8.2)$$

Densité spectrale d'énergie :

La densité spectrale d'énergie d'un signal temporel discret $x[t]$ est définie comme le carré du module de sa Transformée de Fourier :

$$\tilde{x}[f] = |\underline{x}[f]|^2 = |\text{TFD}\{x\}[f]|^2.$$

Elle renseigne la distribution de l'énergie du signal en fonction de la fréquence.

Transformée de Fourier à court terme :

La Transformée de Fourier à court terme (TFCT) d'un signal temporel discret $x[t]$, notée $\underline{x}[t, f]$ est définie par :

$$\underline{x}[t, f] = \sum_{n=-\infty}^{\infty} x[t] \cdot w[n - t] \cdot e^{-jfn},$$

où $w[t]$ est appelée *fenêtre d'appodisation*. Il s'agit d'un signal dont le support temporel est de courte durée et qui est utilisée pour annuler les coefficients du signal $x[t]$ en dehors d'une zone temporelle d'intérêt. Cette fenêtre dispose généralement de propriétés permettant de limiter l'apparition d'oscillations parasites liées à un fenêtrage trop brutal du signal.

La TCFT est très utile pour analyser des signaux dont les propriétés fréquentielles varient au cours du temps (signaux non stationnaires).

Spectrogramme :

Le spectrogramme du signal temporel discret $x[t]$, noté $\tilde{x}[t, f]$, est défini comme le carré du module de sa transformée de Fourier à court terme :

$$\tilde{x}[t, f] = |\underline{x}[t, f]|^2.$$

Il s'agit d'une représentation très utilisée en traitement du signal pour étudier et visualiser l'évolution de l'énergie d'un signal au cours de temps et dans différentes bandes de fréquences (respectivement sur des échelles linéaires).

Liste des figures

2.1	Schéma illustrant l'interaction entre le son émis par une source sonore dans un milieu homogène (en blanc) et un mur d'épaisseur finie (en gris). Lors de cette interaction, le son est réfléchi, dissipé et transmis par le mur au milieu suivant (en beige).	11
2.2	Onde plane incidente \underline{p}_i et onde plane réfléchie \underline{p}_r pour une réflexion spéculaire ($\theta_i = \theta_r$)	13
2.3	Exemple d'une RIS mesurée issue de la base de données d'Echorate (Di Carlo et al., 2021). Le premier graphe représente la RIS entière. Le second graphe représente les premières 50ms de la RIS.	16
2.4	Schéma de la composition d'une réponse impulsionnelle de salle.	17
2.5	Illustration en deux dimensions de la méthode des sources-images faisant apparaître le son direct, une réflexion spéculaire d'ordre 1 et une réflexion spéculaire d'ordre 2 avec les sources-images associées.	18
2.6	Illustration d'une situation d'occlusion acoustique. Les lettres A et B représentent deux murs, R le récepteur, S la source, S_A , S_B des sources-images d'ordre 1 et S_{AB} , S_{BA} des sources-images d'ordre 2. (Gauche) Les trajets $S \rightarrow A \rightarrow R$ et $S \rightarrow A \rightarrow B \rightarrow R$ sont tous les deux valides. (Droite) Le trajet spéculaire $S \rightarrow A \rightarrow B \rightarrow R$ est invalide car $R \rightarrow S_{BA}$ n'intersecte pas A. Source : https://reuk.github.io/wayverb/image_source.html	21
2.7	Modèle général de la réflexion acoustique. Une partie du son incident est absorbée par le mur (en gris), l'autre partie est divisée entre une réflexion spéculaire et une réflexion diffuse.	21
2.8	Schéma d'une chaîne d'acquisition de RIS par émission d'un signal sonore "sine sweep". Avant d'être enregistré par le microphone, le signal émis par la source est affecté successivement par la source, la salle puis le microphone. La RIS est obtenue à partir du signal mesuré en en déconvoluant le sine sweep.	23
2.9	Représentation de la directivité pour le cas d'un instrument omnidirectionnel (gauche) et directionnel (droite).	24

2.10	Illustration de différentes problématiques intervenant pour l'estimation des profils d'absorption des surfaces d'une salle à partir de RISs	25
2.11	Illustration de l'optimisation d'un critère non convexe par application d'un algorithme basique de descente de gradient. Quatre itérations de l'algorithme (symbolisées par des ronds) sont illustrées à partir de deux valeurs initiales distinctes (symbolisées par des carrés). La trajectoire verte montre une convergence vers le minimum global du critère (étoile verte) tandis que la trajectoire verte converge vers un minimum local du critère (étoile rouge)	33
2.12	Schéma d'un perceptron multicouche	34
3.1	Photographies du dispositif expérimental de mesure installé dans l'Espace de projection de l'Ircam (figure 3.1a) et de la chambre anéchoïque de l'Ircam (figure 3.1b) lors de la campagne de mesure organisée en collaboration avec l'équipe Espaces acoustiques et cognitifs de l'Ircam (Novembre 2022).	46
3.2	Schéma illustrant dans le plan temps-fréquence les régions privilégiées pour la simulation des composantes correspondantes des RISs à l'aide des méthodes ondulatoires, déterministes et stochastiques (Thomas, 2017 ; Badeau, 2019).	50
4.1	Visualisation 3D des géométries G1 et G2 (détaillées dans le tableau 4.1). La source est représentée par une sphère et le microphone est représenté par une croix.	58
4.2	Erreur absolue moyenne (EAM) sur l'estimation des coefficients d'absorption en bande d'octave de l'ensemble des parois d'une salle par des modèles de type MLP et CNN entraînés et testés respectivement sur les bases de données simulées associées à la géométrie fixe G1 présentés dans le tableau 4.3	62
4.3	Erreurs absolues moyennes (EAM) par bande d'octave du CNN pour la base de données G1-SDEB (voir tableau 4.3).	63
4.4	Erreurs absolues moyennes du CNN sur l'estimation profils d'absorption en bande d'octave des 6 parois d'un salle pour la base de données G1-SDEB. Les erreurs obtenues en utilisant uniquement la partie précoce ou uniquement la partie tardive des RISs sont comparées pour différents instants de coupure. Pour chaque paroi, les temps d'arrivée du son direct et de l'écho du premier ordre associé sont indiqués (calcul par la méthode des sources-images appliquée à la géométrie du tableau 4.1).	65

5.1	Illustration d'une situation de validation et d'une situation de non validation de la condition définie par \mathcal{J}_k^* par la fenêtre temporelle $[t : t + 2F - 1]$ d'une RIS x_m délimitée par des crochets noirs. Les intervalles \mathcal{W}_t^- et \mathcal{W}_t^+ correspondant sont respectivement délimités par des crochets verts et des crochets rouges. Le premier schéma présente une situation valide où la fenêtre $[t : t + 2F - 1]$ permet d'isoler et d'inclure complètement l'écho associé à k . Le second schéma présente une situation non valide où la fenêtre $[t : t + 2F - 1]$ ne permet ni d'isoler l'écho k , ni de l'inclure complètement.	73
5.2	Illustration de la robustesse dans le cas d'une régression linéaire en deux dimensions. Le modèle linéaire associé à la droite rouge est estimé par le critère des moindres carrées sur l'ensemble des observations. Le modèle linéaire associé à la droite verte est lui estimé par le critère des moindres carrées sur un ensemble d'inliers, constitué typiquement à l'issue de l'algorithme RANSAC. Ce modèle permet de généraliser convenablement la tendance linéaire suivie par la majorité des observations sans influence des outliers, c'est-à-dire de manière robuste aux outliers.	78
5.3	Exemple d'estimation des profils d'absorption des 6 murs d'une salle en forme de pavé droit ($S = 6$ murs) avec la méthode de sélection robuste basée sur RANSAC. L'estimation est effectuée en minimisant (5.5) sur les ensembles $\mathcal{J}_k^{\text{ransac}}$ constitués à partir de $M = 9$ RISs (3 sources et 3 microphones). Les RISs présentent un PSNR égal à 50dB et l'incertitude géométrique un écart-type σ_{geo} égale à 2cm. L'ordre maximal de modélisation Q_{max} , c'est-à-dire considérée pour identifier les fenêtres pertinentes est égal à 2. L'ordre Q des sources-images associées aux fenêtres pertinentes utilisées pour l'estimation est égal à 2.	83
5.4	Erreur absolue moyenne (EAM) et pourcentage des estimations correctes (CE) obtenus avec la procédure de sélection robuste basée sur RANSAC et $Q = 2$ pour différents nombres de sources et microphones et différents niveau de bruit de mesure et d'incertitude géométrique	84
6.1	Sinus cardinal fenêtré $\phi[\cdot - \Delta_{m,k}]$ pour un délai fractionnaire échantillonné à la fréquence $\bar{f} = 16\text{kHz}$ et un délai fractionnaire $\Delta_{m,k} = 0.18$ ms, équivalent à 2.87 échantillons.	91
6.2	Évolution de l'erreur quadratique moyenne en fonction du délai fractionnaire correspondant respectivement au critère sans filtrage gaussien (6.16) (colonne de gauche) et avec filtrage gaussien (6.18) (colonne de droite) pour un écho isolé dans le cas de sources et microphones aux réponses idéales.	98

6.3	Évolution de l’erreur quadratique moyenne en fonction du délai fractionnaire correspondant respectivement au critère sans filtrage gaussien (6.16) (colonne de gauche) et avec filtrage gaussien (6.18) (colonne de droite) pour un écho isolé dans le cas de sources et microphones aux réponses mesurées à dépendance directionnelle et fréquentielle.	99
6.4	Filtres de Gauss pour les différents écarts types considérés dans les expériences.	100
6.5	Schéma de l’algorithme itératif de descente de gradient pour l’estimation des délais correctifs.	101
6.6	(6.6a) Simulation jusqu’au premier ordre de réflexion d’une RIS x_m en utilisant des mesures de réponses en direction et en fréquence de sources et de microphones réels par Brandner et al., 2018. (6.6b) Chaque signal correspond à la RIS théorique qui serait obtenu si l’un des 6 murs était parfaitement réfléchissant, et tous les autres parfaitement absorbant, en présence d’erreurs géométriques. (6.6c) RIM pour chacune des 6 parois. La RIS (6.6a) correspond donc à la somme des six signaux en (6.6b) respectivement convolués avec les six signaux en (6.6c), aux délais issues des erreurs géométriques prêt : $\mathbf{x}_m = \mathbf{H}_m^* \mathbf{C}_m(\Delta_m) \mathbf{v}$ (voir l’équation 6.6). (6.6a) et (6.6b) correspondent aux données d’entrée de notre problème, alors que les délais correctifs Δ_m ainsi que les RIMs (6.6c) sont les inconnues du problème.	106
6.7	Évolution de la fonction de coût (EQM_x) au cours des itérations de l’optimisation des critères “sans filtrage gaussien et $\mathbf{v}_k = \delta_0$ ” (6.7a), “avec filtrage gaussien et $\mathbf{v}_k = \delta_0$ ” (6.7b) et “avec filtrage gaussien et RIMs cumulées \mathbf{v}_k exactes” (6.7c). La ligne horizontale verte “ $\hat{\Delta} = \Delta$ ” indique pour chaque critère la valeur pour les délais correctifs exacts.	110
6.8	Erreur absolue moyenne sur l’estimation des délais exprimés en échantillons (EAM_Δ , figure 6.8a) et pourcentage d’estimations correctes, c’est-à-dire avec une erreur absolue inférieure à 1 échantillon ($\text{CE}_\Delta\%$, figure 6.8b) au cours des itérations de l’optimisation des critères “sans filtrage gaussien (6.16) et $\mathbf{v}_k = \delta_0$ ” (courbes bleues), “avec filtrage gaussien (6.18) et $\mathbf{v}_k = \delta_0$ ” (courbes oranges) et “avec filtrage gaussien (6.18) et RIMs cumulées \mathbf{v}_k exactes” (courbes vertes).	112
6.9	Résidu d’estimation ($\ \mathbf{r}\ _\infty$, 6.9a) et erreur quadratique moyenne relatifs aux RIMs cumulées (EQM_{v_k} , 6.9b) au cours des itérations de la méthode du gradient conjugué. La figure 6.9b différencie l’évolution de l’ EQM_{v_k} pour les sources-images d’ordre égal à 1 d’une part (courbe bleu) et celles d’ordre strictement supérieur à 1 d’autre part.	114

6.10	Erreur absolue moyenne et pourcentage d'estimations correctes relatifs aux profils d'absorption des murs ($EAM_{\alpha_{1:6}}$ et $CE_{\alpha_{1:6}}\%$) en fonction de la fréquence, avec connaissance exacte des temps d'arrivée des échos ($\hat{\Delta} = \Delta$).	115
6.11	Exemple des résultats d'estimation des RIMs d'une salle obtenus avec la méthode du gradient conjugué à partir de $M = 16$ RISs, avec connaissance exacte des temps d'arrivée des échos.	116
6.12	Exemple des résultats d'estimation des RIMs d'une salle obtenus avec la méthode du gradient conjugué à partir de $M = 16$ RISs, pour une estimation des délais correctifs par optimisation sans filtrage gaussien (6.16) avec $\mathbf{v}_k = \boldsymbol{\delta}_0$	119
6.13	Exemple des résultats d'estimation des RIMs d'une salle obtenus avec la méthode du gradient conjugué à partir de $M = 16$ RISs, pour une estimation des délais correctifs par optimisation avec filtrage gaussien (6.18) avec $\mathbf{v}_k = \boldsymbol{\delta}_0$	120
6.14	Exemple des résultats d'estimation des RIMs d'une salle obtenus avec la méthode du gradient conjugué à partir de $M = 16$ RISs, pour une estimation des délais correctifs par optimisation avec filtrage gaussien (6.18) en utilisant les RIMs cumulées \mathbf{v}_k exactes.	120
6.15	Erreur absolue moyenne ($EAM_{\alpha_{1:6}}$ 6.15a) et pourcentage d'estimations correctes ($CE_{\alpha_{1:6}}\%$ 6.15b, erreur absolue inférieure à 0.1) relatifs aux profils d'absorption des murs en fonction de la fréquence pour l'utilisation des critères considérant du filtrage gaussien (6.18) et respectivement des RIMs parfaitement rigides ($\mathbf{v}_k = \boldsymbol{\delta}_0$) (courbes oranges) et les RIMs \mathbf{v}_k exactes (courbes vertes). Les lignes verticales en pointillés indiquent les frontières des bandes d'octave logarithmiques centrées sur les fréquences 0.125, 0.25, 0.5, 1, 2 et 4kHz. Les lignes horizontales et les valeurs en écriture normale indiquent la valeur de la métrique dans chacune des bandes d'octaves puis dans la plage de fréquences $[\sqrt{2} \cdot 4, 8]$ kHz. Les valeurs en gras indiquent la valeur de la métrique pour les fréquences inférieures à la borne supérieure de chaque bande d'octave puis à $\frac{f}{2} = 8$ kHz.	122
6.16	Erreurs absolues moyennes et pourcentages d'estimations correctes sur l'estimation des profils d'absorption ($EAM_{\alpha_{1:6}}$ et $CE_{\alpha_{1:6}}\%$) en fonction du nombre de RISs utilisées pour l'estimation (M) pour différents modes d'estimation des délais correctifs suivis de la méthode du gradient conjugué. Les modes considérés sont l'estimation exacte ($\hat{\Delta} = \Delta$, temps d'arrivée exacts), l'optimisation avec filtrage gaussien avec $\mathbf{v}_k = \boldsymbol{\delta}_0$ ($\hat{\Delta}$ (6.18) , $\mathbf{v}_k = \boldsymbol{\delta}_0$) et l'optimisation avec filtrage gaussien en utilisant les RIMs cumulées exactes ($\hat{\Delta}$ (6.18), \mathbf{v}_k exactes).	125

Liste des tableaux

4.1	Géométries fixes G1 et G2 utilisées pour les expériences (en mètres).	58
4.2	Plages des coefficients d'absorption (abs) et de diffusion (diff) par type de paroi et par bande d'octave.	59
4.3	Description des huit des bases de données simulées utilisés pour les expériences. . . .	60
4.4	Erreurs absolues moyennes (EAM) du CNN pour l'estimation des coefficients d'absorption en bande d'octave pour les bases de données G1-SDEB, G2-SDEB et \mathcal{G} -SDEB. 64	64
5.1	Erreur absolue moyenne (EAM) et pourcentage des estimations correctes (CE%) obtenus en minimisant (5.5) sur différents sous-ensembles de fenêtres de RISs. L'écart-type de l'erreur sur la mesure des paramètres géométriques est $\sigma_{\text{geo}} \in \{0, 2\}$ cm. L'ordre maximal des échos dont les fenêtres sont considérées pour l'estimation est $Q \in \{1, 2\}$. L'ordre maximal de modélisation des sources-images est $Q_{\text{max}} = 2$. Le PSNR des RISs est fixé à 50dB.	82
6.1	Erreur quadratique moyenne relative aux RIMs cumulées (EQM_{v_k}), erreur absolue moyenne et pourcentage d'estimations correctes relatifs aux spectres des RIMs cumulées ($\text{EAM}_{\tilde{v}_k}$ et $\text{CE}_{\tilde{v}_k} \%$), avec connaissance exacte des temps d'arrivée des échos ($\hat{\Delta} = \Delta$). Les métriques sont présentées séparément pour les sources-images d'ordre égal 1 ($q(\mathbf{k}) = 1$) et celle d'ordre strictement supérieur à 1 ($q(\mathbf{k}) > 1$).	115
6.2	Erreurs absolues moyennes et pourcentages d'estimations correctes sur l'estimation des délais correctifs (EAM_{Δ} et $\text{CE}_{\Delta} \%$) et sur l'estimation des profils d'absorption ($\text{EAM}_{\alpha_{1,6}}$ et $\text{CE}_{\alpha_{1,6}} \%$) ainsi que erreurs quadratiques moyennes sur l'estimation des RIMs ($\text{EQM}_{v_{1,6}}$), pour différents modes d'estimation des délais correctifs suivis de la méthode du gradient conjugué. Les modes considérés sont l'estimation exacte (" $\hat{\Delta} = \Delta$ ", temps d'arrivée exacts), l'estimation nulle (" $\hat{\Delta} = 0$ ", temps d'arrivée erronés non corrigés), l'optimisation sans filtrage gaussien avec $\mathbf{v}_k = \delta_0$ (" $\hat{\Delta}$ (6.16), $\mathbf{v}_k = \delta_0$ "), l'optimisation avec filtrage gaussien avec $\mathbf{v}_k = \delta_0$ (" $\hat{\Delta}$ (6.18), $\mathbf{v}_k = \delta_0$ ") et l'optimisation avec filtrage gaussien en utilisant les RIMs cumulées exactes (" $\hat{\Delta}$ (6.18), \mathbf{v}_k exactes").	119

- 6.3 Résultats d'expériences analogues à celles présentées en sous-section 6.3.6 à l'exception que les réponses de sources et de microphones considérées pour la simulation des données sont idéales. Erreurs absolues moyennes et pourcentages d'estimations correctes sur l'estimation des délais correctifs (EAM_{Δ} et $CE_{\Delta}\%$) et sur l'estimation des profils d'absorption ($EAM_{\alpha_{1:6}}$ et $CE_{\alpha_{1:6}}\%$) ainsi que erreurs quadratiques moyennes sur l'estimation des RIMs ($EQM_{v_{1:6}}$), pour différents modes d'estimation des délais correctifs suivis de la méthode du gradient conjugué. Les modes considérés sont l'estimation exacte (" $\hat{\Delta} = \Delta$ ", temps d'arrivée exacts), l'optimisation avec filtrage gaussien avec $\mathbf{v}_k = \delta_0$ (" $\hat{\Delta}$ (6.18), $\mathbf{v}_k = \delta_0$ ") et l'optimisation avec filtrage gaussien en utilisant les RIMs cumulées exactes (" $\hat{\Delta}$ (6.18), \mathbf{v}_k exactes"). 123
- 6.4 Erreurs quadratiques moyennes sur l'estimation des RIMs ($EQM_{v_{1:6}}$) en fonction du nombre de RISs utilisées pour l'estimation (M) pour différents modes d'estimation des délais correctifs suivis de la méthode du gradient conjugué. Les modes considérés sont l'estimation exacte (" $\hat{\Delta} = \Delta$ ", temps d'arrivée exacts), l'optimisation avec filtrage gaussien avec $\mathbf{v}_k = \delta_0$ (" $\hat{\Delta}$ (6.18), $\mathbf{v}_k = \delta_0$ ") et l'optimisation avec filtrage gaussien en utilisant les RIMs cumulées exactes (" $\hat{\Delta}$ (6.18), \mathbf{v}_k exactes"). 124
- 6.5 Résultats d'expériences analogues à celles présentées en sous-section 6.3.6 à l'exception qu'une modélisation plus réaliste des effets de l'incertitude géométrique sur les temps d'arrivée erronés est effectuée. Erreurs absolues moyennes et pourcentages d'estimations correctes sur l'estimation des délais correctifs (EAM_{Δ} et $CE_{\Delta}\%$) et sur l'estimation des profils d'absorption ($EAM_{\alpha_{1:6}}$ et $CE_{\alpha_{1:6}}\%$) ainsi que erreurs quadratiques moyennes sur l'estimation des RIMs ($EQM_{v_{1:6}}$), pour différents modes d'estimation des délais correctifs suivis de la méthode du gradient conjugué. Les modes considérés sont l'estimation exacte (" $\hat{\Delta} = \Delta$ ", temps d'arrivée exacts), l'optimisation avec filtrage gaussien avec $\mathbf{v}_k = \delta_0$ (" $\hat{\Delta}$ (6.18), $\mathbf{v}_k = \delta_0$ ") et l'optimisation avec filtrage gaussien en utilisant les RIMs cumulées exactes (" $\hat{\Delta}$ (6.18), \mathbf{v}_k exactes").126

Liste des acronymes

- CE%** Pourcentage des estimations correctes. [81](#)
- CNN** Convolutional network (Réseau neuronal convolutif). [60](#)
- DRR** Direct to Reverberant Ratio (Rapport son direct sur son réverbéré). [45](#)
- EAM** Erreur absolue moyenne. [61](#)
- EQM** Erreur quadratique moyenne. [60](#)
- FDTD** Finite Difference Time Domain. [47](#)
- FEM** Finite Element Method. [47](#)
- GPU** Graphic Processing Unit (Processeur graphique ou carte graphique). [47](#)
- MLP** Multilayer perceptron (Perceptron multicouche). [60](#)
- PSNR** Peak signal-to-noise Ratio (Rapport signal sur bruit maximal). [26](#)
- RANSAC** Random sample consensus. [78](#)
- RIM** Réponse impulsionnelle de mur. [13](#)
- RIS** Réponse impulsionnelle de salle. [15](#)
- SNR** Signal-to-noise ratio (Rapport signal sur bruit). [26](#)
- TFCT** Transformée de Fourier court à terme. [70](#)
- TFD** Transformée de Fourier discrète. [30](#)

Notations

Notations

$x(\cdot), X(\cdot)$ signaux continus

$x[\cdot], X[\cdot]$ signaux discrets

\mathbf{x} vecteur

\mathbf{X} matrice

$\underline{x}(f), \underline{x}[f]$ transformée de Fourier (continue et discrète)

$\tilde{x}[f]$ densité spectrale d'énergie discrète (carrée du module de la Transformée de Fourier)

$\tilde{x}[t, f]$ spectrogramme (carrée du module de la Transformée de Fourier à court terme)

Références

- ABEL, Jonathan S ; HUANG, Patty, 2006. A simple, robust measure of reverberation echo density. In : *Audio Engineering Society Convention 121*. Audio Engineering Society.
- ALLARD, Jean ; ATALLA, Noureddine, 2009. *Propagation of sound in porous media : modelling sound absorbing materials*. John Wiley & Sons.
- ALLEN, Jont B ; BERKLEY, David A, 1979. Image method for efficiently simulating small-room acoustics. *The Journal of the Acoustical Society of America*. T. 65, n° 4, p. 943-950.
- ANTONELLO, Niccolò ; WATERSCHOOT, Toon van ; MOONEN, Marc ; NAYLOR, P, 2015. Evaluation of a Numerical Method for Identifying Surface Acoustic Impedances in a Reverberant Room. In : *Proc. of the 10th European Congress and Exposition on Noise Control Engineering*, p. 1-6.
- AOSHIMA, Nobuharu, 1981. Computer-generated pulse signal applied for sound measurement. *The Journal of the Acoustical Society of America*. T. 69, n° 5, p. 1484-1488.
- ASTM, E, 1998. 1050-98. *Standard Test Method for Impedance and Absorption of Acoustical Materials Using Tube, Two Microphones and A Digital Frequency Analysis System*, American Society for Testing and Materials.
- BADEAU, Roland, 2019. Common mathematical framework for stochastic reverberation models. *The Journal of the Acoustical Society of America*. T. 145, n° 4, p. 2733-2745.
- BALINT, Jamilla ; BERZBORN, Marco ; NOLAN, Mélanie ; VORLÄNDER, Michael, 2023. Measuring Sound Absorption : The Hundred-Year Debate on the Reverberation Chamber Method. *Acoustics Today*. T. 19, p. 13. Disp. à l'adr. DOI : [10.1121/AT.2023.19.3.13](https://doi.org/10.1121/AT.2023.19.3.13). ©2023 Acoustical Society of America.
- BARRY, Timothy M, 1974. Measurement of the absorption spectrum using correlation/spectral density techniques. *The Journal of the Acoustical Society of America*. T. 55, n° 6, p. 1349-1351.
- BARUMERLI, Roberto ; BIANCHI, Daniele ; GERONAZZO, Michele ; AVANZINI, Federico, 2021. SofaMyRoom : a fast and multiplatform" shoebox" room simulator for binaural room impulse response dataset generation. *arXiv preprint arXiv :2106.12992*.
- BERTIN, Nancy ; BADEAU, Roland ; VINCENT, Emmanuel, 2010. Enforcing harmonicity and smoothness in Bayesian non-negative matrix factorization applied to polyphonic music transcription. *IEEE Transactions on Audio, Speech, and Language Processing*. T. 18, n° 3, p. 538-549.

- BERTIN, Nancy ; KITIĆ, Srđan ; GRIBONVAL, Rémi, 2016. Joint estimation of sound source location and boundary impedance with physics-driven cosparsity regularization. In : *2016 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, p. 6340-6344.
- BERZBORN, Marco ; NOLAN, Mélanie ; FERNANDEZ-GRANDE, Efrén ; VORLÄNDER, Michael, 2019. *On the directional properties of energy decay curves*. Universitätsbibliothek der RWTH Aachen.
- BOLLES, Robert C ; FISCHLER, Martin A, 1981. A RANSAC-based approach to model fitting and its application to finding cylinders in range data. In : *IJCAI*. T. 1981, p. 637-643.
- BORISH, Jeffrey, 1984. Extension of the image model to arbitrary polyhedra. *The Journal of the Acoustical Society of America*. T. 75, n° 6, p. 1827-1836.
- BORK, Ingolf, 2000. A comparison of room simulation software—the 2nd round robin on room acoustical computer simulation. *Acta Acustica united with Acustica*. T. 86, n° 6, p. 943-956.
- BOYD, Stephen ; PARIKH, Neal ; CHU, Eric ; PELEATO, Borja ; ECKSTEIN, Jonathan et al., 2011. Distributed optimization and statistical learning via the alternating direction method of multipliers. *Foundations and Trends® in Machine Learning*. T. 3, n° 1, p. 1-122.
- BRANDÃO, Eric ; LENZI, Arcanjo ; PAUL, Stephan, 2015. A review of the in situ impedance and sound absorption measurement techniques. *Acta Acustica united with Acustica*. T. 101, n° 3, p. 443-463.
- BRANDNER, Manuel ; FRANK, Matthias ; RUDRICH, Daniel, 2018. DirPat—Database and viewer of 2D/3D directivity patterns of sound sources and receivers. In : *Audio Engineering Society Convention 144*. Audio Engineering Society.
- BRINKMANN, Fabian ; ASPÖCK, Lukas ; ACKERMANN, David ; LEPA, Steffen ; VORLÄNDER, Michael ; WEINZIERL, Stefan, 2019. A round robin on room acoustical simulation and auralization. *The Journal of the Acoustical Society of America*. T. 145, n° 4, p. 2746-2760.
- BRINKMANN, Fabian ; ASPÖCK, Lukas ; ACKERMANN, David ; OPDAM, Rob ; VORLÄNDER, Michael ; WEINZIERL, Stefan, 2021. A benchmark for room acoustical simulation. Concept and database. *Applied Acoustics*. T. 176, p. 107867.
- BRYAN, Nicholas J, 2020. Impulse response data augmentation and deep neural networks for blind room acoustic parameter estimation. In : *ICASSP 2020-2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, p. 1-5.
- CAMPBELL, D ; PALOMAKI, K ; BROWN, G, 2005. A MATLAB simulation of "shoebox" room acoustics for use in research and teaching. *Computing and Information Systems*. T. 9, n° 3, p. 48.
- CARLO, Diego Di ; TANDEITNIK, Pinchas ; FOY, Cedric ; BERTIN, Nancy ; DELEFORGE, Antoine ; GANNOT, Sharon, 2021. dEchorate : a calibrated room impulse response dataset for echo-aware signal processing. *EURASIP Journal on Audio, Speech, and Music Processing*. T. 2021, n° 1, p. 1-15.

- CHEN, Changan ; JAIN, Unnat ; SCHISLER, Carl ; GARI, Sebastia Vicenc Amengual ; ALHALAH, Ziad ; ITHAPU, Vamsi Krishna ; ROBINSON, Philip ; GRAUMAN, Kristen, 2020. SoundSpaces : Audio-Visual Navigaton in 3D Environments. In : *ECCV*.
- CHEN, Changan ; SCHISLER, Carl ; GARG, Sanchit ; KOBERNIK, Philip ; CLEGG, Alexander ; CALAMIA, Paul ; BATRA, Dhruv ; ROBINSON, Philip W ; GRAUMAN, Kristen, 2022. SoundSpaces 2.0 : A Simulation Platform for Visual-Acoustic Learning. In : *NeurIPS 2022 Datasets and Benchmarks Track*.
- CHUNG, JY ; BLASER, DA, 1980a. Transfer function method of measuring in-duct acoustic properties. I. Theory. *The Journal of the Acoustical Society of America*. T. 68, n° 3, p. 907-913.
- CHUNG, JY ; BLASER, DA, 1980b. Transfer function method of measuring in-duct acoustic properties. II. Experiment. *The Journal of the Acoustical Society of America*. T. 68, n° 3, p. 914-921.
- CIABURRO, Giuseppe ; IANNACE, Gino, 2021. Acoustic characterization of rooms using reverberation time estimation based on supervised learning algorithm. *Applied Sciences*. T. 11, n° 4, p. 1661.
- COX, Trevor ; D'ANTONIO, Peter, 2016. *Acoustic absorbers and diffusers : theory, design and application*. CRC press.
- CRAMOND, AJ ; DON, CG, 1984. Reflection of impulses as a method of determining acoustic impedance. *The Journal of the Acoustical Society of America*. T. 75, n° 2, p. 382-389.
- CROCCO, Marco ; DEL BUE, Alessio, 2016. Estimation of TDOA for room reflections by iterative weighted l1 constraint. In : *2016 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, p. 3201-3205.
- CUMMINGS, A, 1991. Impedance tube measurements on porous media : The effects of air-gaps around the sample. *Journal of sound and vibration*. T. 151, n° 1, p. 63-75.
- DAVIES, Huw G, 1973. Noise propagation in corridors. *The Journal of the Acoustical Society of America*. T. 53, n° 5, p. 1253-1262.
- DAVIES, JC ; MULHOLLAND, KA, 1979. An impulse method of measuring normal impedance at oblique incidence. *Journal of Sound and Vibration*. T. 67, n° 1, p. 135-149.
- DEFRANCE, G ; POLACK, J-D, 2010. Estimating the mixing time of concert halls using the eXtensible Fourier Transform. *Applied Acoustics*. T. 71, n° 9, p. 777-792.
- DI CARLO, Diego ; TANDEITNIK, Pinchas ; FOY, Cédric ; DELEFORGE, Antoine ; BERTIN, Nancy ; GANNOT, Sharon, 2021. dEchorate : a Calibrated Room Impulse Response Database for Echo-aware Signal Processing. *arXiv preprint arXiv :2104.13168*.
- DIAZ-GUERRA, David ; MIGUEL, Antonio ; BELTRAN, Jose R, 2021. gpuRIR : A python library for room impulse response simulation with GPU acceleration. *Multimedia Tools and Applications*. T. 80, p. 5653-5671.

- DILUNGAN, Stéphane ; DELEFORGE, Antoine ; FOY, Cédric ; FAISAN, Sylvain, 2021. Learning-based estimation of individual absorption profiles from a single room impulse response with known positions of source, sensor and surfaces. In : *INTER-NOISE and NOISE-CON Congress and Conference Proceedings*. Institute of Noise Control Engineering. T. 263, p. 5623-5630. N° 1.
- DILUNGAN, Stéphane ; DELEFORGE, Antoine ; FOY, Cédric ; FAISAN, Sylvain, 2022. Geometry-informed estimation of surface absorption profiles from room impulse responses. In : *2022 30th European Signal Processing Conference (EUSIPCO)*. IEEE, p. 867-871.
- DOKMANIĆ, Ivan ; LU, Yue M ; VETTERLI, Martin, 2011. Can one hear the shape of a room : The 2-D polygonal case. In : *2011 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, p. 321-324.
- DOKMANIĆ, Ivan ; PARHIZKAR, Reza ; WALTHER, Andreas ; LU, Yue M ; VETTERLI, Martin, 2013. Acoustic echoes reveal room shape. *Proceedings of the National Academy of Sciences*. T. 110, n° 30, p. 12186-12191.
- DUANGPUMMET, Suradej ; KARNJANA, Jessada ; KONGPRAWECHNON, Waree ; UNOKI, Masashi, 2022. Blind estimation of speech transmission index and room acoustic parameters based on the extended model of room impulse response. *Applied Acoustics*. T. 185, p. 108372.
- DUPONT, Samuel ; MELON, Manuel ; BERRY, Alain, 2020. Characterization of acoustic material at oblique incidence using a spherical microphone array. *The Journal of the Acoustical Society of America*. T. 147, n° 5, p. 3613-3625.
- EATON, James ; GAUBITCH, Nikolay D ; MOORE, Alastair H ; NAYLOR, Patrick A, 2015. The ACE challenge—Corpus description and performance evaluation. In : *2015 IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA)*. IEEE, p. 1-5.
- EATON, James ; GAUBITCH, Nikolay D ; MOORE, Alastair H ; NAYLOR, Patrick A, 2016. Estimation of room acoustic parameters : The ACE challenge. *IEEE/ACM Transactions on Audio, Speech, and Language Processing*. T. 24, n° 10, p. 1681-1693.
- EYRING, Carl F, 1930. Reverberation time in “dead” rooms. *The Journal of the Acoustical Society of America*. T. 1, n° 2A, p. 217-241.
- FARINA, A ; GALAVERNA, P ; TRUFFELLI, G, 2007. “Ramsete : Room acoustics modeling on pc. URL <http://www.ramsete.com>.
- FARINA, Angelo, 2000. Simultaneous measurement of impulse response and distortion with a swept-sine technique. In : *Audio engineering society convention 108*. Audio Engineering Society.
- FARINA, Angelo, 2007. Advancements in impulse response measurements by sine sweeps. In : *Audio engineering society convention 122*. Audio Engineering Society.
- FERNANDEZ-GRANDE, Efren ; BIANCO, Michael J ; GANNOT, Sharon ; GERSTOFT, Peter, 2021. DTU three-channel room impulse response dataset for direction of arrival estimation 2020. *IEEE Dataport*.

- FOY, Cédric ; DELEFORGE, Antoine ; DI CARLO, Diego, 2021. Mean absorption estimation from room impulse responses using virtually supervised learning. *The Journal of the Acoustical Society of America*. T. 150, n° 2, p. 1286-1299.
- FU, Zhong-hua ; LI, Jian-wei, 2016. GPU-based image method for room impulse response calculation. *Multimedia Tools and Applications*. T. 75, p. 5205-5221.
- GARAI, Massimo, 1993. Measurement of the sound-absorption coefficient in situ : the reflection method using periodic pseudo-random sequences of maximum length. *Applied acoustics*. T. 39, n° 1-2, p. 119-139.
- GARDNER, Matt W ; DORLING, SR, 1998. Artificial neural networks (the multilayer perceptron)—a review of applications in the atmospheric sciences. *Atmospheric environment*. T. 32, n° 14-15, p. 2627-2636.
- GÓMEZ BOLAÑOS, Javier ; DELIKARIS-MANIAS, Symeon ; PULKKI, Ville ; ESKELINEN, Joonas ; HÆGGSTRÖM, Edward, 2014. Laser-induced acoustic point source for accurate impulse response measurements within the audible bandwidth. *The Journal of the Acoustical Society of America*. T. 135, n° 6, EL298-EL303.
- GOODFELLOW, Ian J. ; POUGET-ABADIE, Jean ; MIRZA, Mehdi ; XU, Bing ; WARDE-FARLEY, David ; OZAIR, Sherjil ; COURVILLE, Aaron ; BENGIO, Yoshua, 2014. *Generative Adversarial Networks*. Disp. à l'adr. arXiv : [1406.2661](https://arxiv.org/abs/1406.2661) [stat.ML].
- GUIDORZI, PAOLO ; BARBARESI, LUCA ; D'ORAZIO, DARIO ; GARAI, MASSIMO, 2015. Impulse responses measured with MLS or Swept-Sine signals applied to architectural acoustics : an in-depth analysis of the two methods and some case studies of measurements inside theaters. *Energy Procedia*. T. 78, p. 1611-1616.
- HABETS, Emanuel AP, 2006. Room impulse response generator. *Technische Universiteit Eindhoven, Tech. Rep.* T. 2, n° 2.4, p. 1.
- HADAD, Elijor ; HEESE, Florian ; VARY, Peter ; GANNOT, Sharon, 2014. Multichannel audio database in various acoustic environments. In : *2014 14th International Workshop on Acoustic Signal Enhancement (IWAENC)*. IEEE, p. 313-317.
- HALD, Jørgen ; SONG, Wookeun ; HADDAD, Karim ; JEONG, Cheol-Ho ; RICHARD, Antoine, 2019. In-situ impedance and absorption coefficient measurements using a double-layer microphone array. *Applied Acoustics*. T. 143, p. 74-83.
- HAMILTON, Brian, 2021. *PFFDTD software* [<https://github.com/bsxfun/pffDTD>].
- HECHT-NIELSEN, Robert, 1992. Theory of the backpropagation neural network. In : *Neural networks for perception*. Elsevier, p. 65-93.
- HEINZ, Renate, 1993. Binaural room simulation based on an image source model with addition of statistical methods to include the diffuse sound scattering of walls and to predict the reverberant tail. *Applied Acoustics*. T. 38, n° 2-4, p. 145-159.

- HESTENES, Magnus R ; STIEFEL, Eduard et al., 1952. Methods of conjugate gradients for solving linear systems. *Journal of research of the National Bureau of Standards*. T. 49, n° 6, p. 409-436.
- HOCHREITER, Sepp ; SCHMIDHUBER, Jürgen, 1997. Long short-term memory. *Neural computation*. T. 9, n° 8, p. 1735-1780.
- HOLLIN, KA ; JONES, MH, 1977. The measurement of sound absorption coefficient in situ by a correlation technique. *Acta Acustica united with Acustica*. T. 37, n° 2, p. 103-110.
- ICK, Christopher ; MEHRABI, Adib ; JIN, Wenyu, 2023. Blind Acoustic Room Parameter Estimation Using Phase Features. In : *ICASSP 2023-2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, p. 1-5.
- INTERNATIONAL, ASTM, 2022. ASTM C423, Standard Test Method for Sound Absorption and Sound Absorption Coefficients by the Reverberation Room Method. *Subcommittee E33.01, editor. ASTM B. Stand. Vol. 04.06 Therm. Insul. Build. Environ. Acoust.* T. 4, n° 6, p.1-11.
- IRCAM, [s. d.] [<https://www.ircam.fr>
Espace de projection et chambre anéchoïque : <https://www.ircam.fr/lircam/le-batiment>].
- STANDARDIZATION (ISO), International Organization for, 2003. ISO 354 :2003, Acoustics – Measurement of Sound Absorption in a Reverberation Room.
- JEONG, Cheol-Ho, 2013. Converting Sabine absorption coefficients to random incidence absorption coefficients. *The Journal of the Acoustical Society of America*. T. 133, n° 6, p. 3951-3962.
- JEONG, Cheol-Ho ; LEE, Doheon ; SANTURETTE, Sébastien ; IH, Jeong-Guon, 2014. Influence of impedance phase angle on sound pressures and reverberation times in a rectangular room. *The Journal of the Acoustical Society of America*. T. 135, n° 2, p. 712-723.
- JEUB, Marco ; SCHAFER, Magnus ; VARY, Peter, 2009. A binaural room impulse response database for the evaluation of dereverberation algorithms. In : *2009 16th International Conference on Digital Signal Processing*. IEEE, p. 1-5.
- KINGMA, Diederik P ; BA, Jimmy, 2014. Adam : A method for stochastic optimization. *arXiv preprint arXiv :1412.6980*.
- KUTTRUFF, Heinrich, 1995. A simple iteration scheme for the computation of decay constants in enclosures with diffusely reflecting boundaries. *The Journal of the Acoustical Society of America*. T. 98, n° 1, p. 288-293.
- LANOYE, Reinhilde ; BOECKX, Laurens ; DE GEETERE, Lieven ; VERMEIR, Gerrit ; LAURIKS, Walter, 2004. Experience with different free field techniques to evaluate the surface impedance. In : *ICA*, p. 1983-1986.
- LANOYE, Reinhilde ; VERMEIR, Gerrit ; LAURIKS, Walter ; KRUSE, Roland ; MELLERT, Volker, 2006. Measuring the free field acoustic impedance and absorption coefficient of sound absorbing materials with a combined particle velocity-pressure sensor. *The Journal of the Acoustical Society of America*. T. 119, n° 5, p. 2826-2831.

- LECUN, Yann; BOSER, Bernhard; DENKER, John S; HENDERSON, Donnie; HOWARD, Richard E; HUBBARD, Wayne; JACKEL, Lawrence D, 1989. Backpropagation applied to handwritten zip code recognition. *Neural computation*. T. 1, n° 4, p. 541-551.
- LENTZ, Tobias; SCHRÖDER, Dirk; VORLÄNDER, Michael; ASSENMACHER, Ingo, 2007. *EURASIP journal on advances in signal processing*. T. 2007, p. 1-19.
- LOONEY, David; GAUBITCH, Nikolay D, 2020. Joint estimation of acoustic parameters from single-microphone speech observations. In : *ICASSP 2020-2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, p. 431-435.
- LUO, Yi; YU, Jianwei, 2022. *FRA-RIR : Fast Random Approximation of the Image-source Method*. Disp. à l'adr. arXiv : [2208.04101](https://arxiv.org/abs/2208.04101) [eess.AS].
- MASZTALSKI, Piotr; MATUSZEWSKI, Mateusz; PIASKOWSKI, Karol; ROMANIUK, Michał, 2020. StoRIR : Stochastic room impulse response generation for audio data augmentation. *arXiv preprint arXiv :2008.07231*.
- MOMMERTZ, Eckard, 1995. Angle-dependent in-situ measurements of reflection coefficients using a subtraction technique. *Applied Acoustics*. T. 46, n° 3, p. 251-263.
- MONDET, Boris; BRUNSKOG, Jonas; JEONG, Cheol-Ho; RINDEL, Jens Holger, 2020. From absorption to impedance : Enhancing boundary conditions in room acoustic simulations. *Applied Acoustics*. T. 157, p. 106884.
- MÖSER, Michael; MÖSER, Michael, 2009. Fundamentals of room acoustics. *Engineering Acoustics : An Introduction to Noise Control*, p. 217-236.
- MÜLLER, Swen; MASSARANI, Paulo, 2001. Transfer-function measurement with sweeps. *Journal of the Audio Engineering Society*. T. 49, n° 6, p. 443-471.
- NAKAMURA, Satoshi; HIYANE, Kazuo; ASANO, Futoshi; NISHIURA, Takanobu; YAMADA, Takeshi, 2000. Acoustical Sound Database in Real Environments for Sound Scene Understanding and Hands-Free Speech Recognition. In : *LREC*, p. 965-968.
- NAVA, Gabriel Pablo; YASUDA, Yosuke; SATO, Yoichi; SAKAMOTO, Shinichi, 2009. On the in situ estimation of surface acoustic impedance in interiors of arbitrary shape by acoustical inverse methods. *Acoustical science and technology*. T. 30, n° 2, p. 100-109.
- NOLAN, Mélanie, 2020. Estimation of angle-dependent absorption coefficients from spatially distributed in situ measurements. *The Journal of the Acoustical Society of America*. T. 147, n° 2, EL119-EL124.
- NOLAN, Mélanie; BERZBORN, Marco; FERNANDEZ-GRANDE, Efren, 2020. Isotropy in decaying reverberant sound fields. *The Journal of the Acoustical Society of America*. T. 148, n° 2, p. 1077-1088.

- OKAWA, Yukiko ; WATANABE, Yasuaki ; IKEDA, Yusuke ; OIKAWA, Yasuhiro, 2021. Estimation of acoustic impedances in a room using multiple sound intensities and ftd method. In : *Advances in Acoustics, Noise and Vibration-Proceedings of the 27th International Congress on Sound and Vibration, ICSV*.
- PAATERO, Pentti ; TAPPER, Unto, 1994. Positive matrix factorization : A non-negative factor model with optimal utilization of error estimates of data values. *Environmetrics*. T. 5, n° 2, p. 111-126.
- PASZKE, Adam ; GROSS, Sam ; MASSA, Francisco ; LERER, Adam ; BRADBURY, James ; CHANAN, Gregory ; KILLEEN, Trevor ; LIN, Zeming ; GIMELSHEIN, Natalia ; ANTIGA, Luca et al., 2019. Pytorch : An imperative style, high-performance deep learning library. *Advances in neural information processing systems*. T. 32.
- PEREZ, Ricardo Falcon ; GÖTZ, Georg ; PULKKI, Ville, 2019. Machine-learning-based estimation of reverberation time using room geometry for room effect rendering. In : *Proceedings of the 23rd International Congress on Acoustics : integrating 4th EAA Euroregio*. T. 9, p. 13.
- PRINN, Albert G, 2023. A review of finite element methods for room acoustics. In : *Acoustics*. MDPI. T. 5, p. 367-395. N° 2.
- PyTorch Documentation : torch.nn.functional*, [s. d.] [<https://pytorch.org/docs/stable/nn.functional.html>].
- RAGHUVANSHI, Nikunj ; LLOYD, Brandon ; GOVINDARAJU, Naga ; LIN, Ming C, 2009. Efficient numerical acoustic simulation on graphics processors using adaptive rectangular decomposition. In : *Proceedings of the EAA Symposium on Auralization*.
- RATNARAJAH, Anton ; TANG, Zhenyu ; MANOCHA, Dinesh, 2020. IR-GAN : Room impulse response generator for far-field speech recognition. *arXiv preprint arXiv :2010.13219*.
- RATNARAJAH, Anton ; ZHANG, Shi-Xiong ; YU, Meng ; TANG, Zhenyu ; MANOCHA, Dinesh ; YU, Dong, 2022. FAST-RIR : Fast neural diffuse room impulse response generator. In : *ICASSP 2022-2022 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, p. 571-575.
- REMAGGI, Luca ; JACKSON, Philip JB ; COLEMAN, Philip ; WANG, Wenwu, 2016. Acoustic reflector localization : Novel image source reversion and direct localization methods. *IEEE/ACM Transactions on Audio, Speech, and Language Processing*. T. 25, n° 2, p. 296-309.
- RIBEIRO, Flavio ; FLORENCIO, Dinei ; BA, Demba ; ZHANG, Cha, 2011. Geometrically constrained room modeling with compact microphone arrays. *IEEE Transactions on Audio, Speech, and Language Processing*. T. 20, n° 5, p. 1449-1460.
- RICHARD, Antoine ; FERNANDEZ-GRANDE, Efren ; BRUNSKOG, Jonas ; JEONG, Cheol-Ho, 2017. Estimation of surface impedance at oblique incidence based on sparse array processing. *The Journal of the Acoustical Society of America*. T. 141, n° 6, p. 4115-4125.

- RIFE, Douglas D ; VANDERKOOY, John, 1989. Transfer-function measurement with maximum-length sequences. *Journal of the Audio Engineering Society*. T. 37, n° 6, p. 419-444.
- RÖBER, Niklas ; SPINDLER, Martin ; MASUCH, Maic, 2006. Waveguide-based Room Acoustics through Graphics Hardware. In : *ICMC*.
- ROSENBLATT, Frank, 1958. The perceptron : a probabilistic model for information storage and organization in the brain. *Psychological review*. T. 65, n° 6, p. 386.
- RUST, Romana ; XYDIS, Achilleas ; HEUTSCHI, Kurt ; PERRAUDIN, Nathanael ; CASAS, Gonzalo ; DU, Chaoyu ; STRAUSS, Jürgen ; EGGENSCHWILER, Kurt ; PEREZ-CRUZ, Fernando ; GRAMAZIO, Fabio et al., 2021. A data acquisition setup for data driven acoustic design. *Building Acoustics*. T. 28, n° 4, p. 345-360.
- SABINE, Wallace Clement, 1922. *Collected papers on acoustics*. Harvard university press.
- SAVIOJA, Lauri ; SVENSSON, U Peter, 2015. Overview of geometrical room acoustic modeling techniques. *The Journal of the Acoustical Society of America*. T. 138, n° 2, p. 708-730.
- SAVIOJA, Lauri ; XIANG, Ning, 2019. Introduction to the special issue on room acoustic modeling and auralization. *The Journal of the Acoustical Society of America*. T. 145, n° 4, p. 2597-2600.
- SCHEIBLER, Robin ; BEZZAM, Eric ; DOKMANIĆ, Ivan, 2018. Pyroomacoustics : A python package for audio room simulation and array processing algorithms. In : *2018 IEEE international conference on acoustics, speech and signal processing (ICASSP)*. IEEE, p. 351-355.
- SCHIMMEL, Steven M ; MULLER, Martin F ; DILLIER, Norbert, 2009. A fast and accurate “shoe-box” room acoustics simulator. In : *2009 IEEE International Conference on Acoustics, Speech and Signal Processing*. IEEE, p. 241-244.
- SCHMID, Jonas M ; ESER, Martin ; MARBURG, Steffen, 2023. Bayesian Approach for the In Situ Estimation of the Acoustic Boundary Admittance. *Journal of Theoretical and Computational Acoustics*. T. 31, n° 04, p. 2350013.
- SCHRÖDER, Dirk, 2011. *Physically based real-time auralization of interactive virtual environments*. T. 11. Logos Verlag Berlin GmbH.
- SCHRÖDER, Dirk ; DROSS, Philipp ; VORLÄNDER, Michael, 2007. A fast reverberation estimator for virtual environments. In : *Audio engineering society conference : 30th international conference : intelligent audio environments*. Audio Engineering Society.
- SCHRÖDER, Dirk ; VORLÄNDER, Michael, 2011. RAVEN : A real-time framework for the auralization of interactive virtual environments. In : *Forum acusticum*. Aalborg Denmark, p. 1541-1546.
- SCHROEDER, Manfred R, 1965. New method of measuring reverberation time. *The Journal of the Acoustical Society of America*. T. 37, n° 6.Supplement, p. 1187-1188.
- SCHROEDER, Manfred R, 1979. Integrated-impulse method measuring sound decay without using impulses. *The Journal of the Acoustical Society of America*. T. 66, n° 2, p. 497-500.

- SILTANEN, Samuel ; LOKKI, Tapio ; SAVIOJA, Lauri, 2010. Rays or waves? Understanding the strengths and weaknesses of computational room acoustics modeling techniques. In : *Proceedings of the International Symposium on Room Acoustics, ISRA*, p. 29-31.
- SKÅLEVIK, Magne, 2011. Schroeder frequency revisited. In : *Proceedings of Forum Acusticum*. T. 2011.
- SPRUNCK, Tom ; DELEFORGE, Antoine ; PRIVAT, Yannick ; FOY, Cédric, 2022. Gridless 3d recovery of image sources from room impulse responses. *IEEE Signal Processing Letters*. T. 29, p. 2427-2431.
- SRIVASTAVA, Prerak ; DELEFORGE, Antoine ; VINCENT, Emmanuel, 2021. Blind room parameter estimation using multiple multichannel speech recordings. In : *2021 IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA)*. IEEE, p. 226-230.
- SRIVASTAVA, Prerak ; DELEFORGE, Antoine ; VINCENT, Emmanuel, 2022. Realistic sources, receivers and walls improve the generalisability of virtually-supervised blind acoustic parameter estimators. In : *2022 International Workshop on Acoustic Signal Enhancement (IWAENC)*. IEEE, p. 1-5.
- STAN, Guy-Bart ; EMBRECHTS, Jean-Jacques ; ARCHAMBEAU, Dominique, 2002. Comparison of different impulse response measurement techniques. *Journal of the Audio engineering society*. T. 50, n° 4, p. 249-262.
- STANDARDIZATION (ISO), International Organization for, 1994a. ISO 10534-1, Acoustics - Determination of sound absorption coefficient and impedance in impedance tubes - Part1 : Method using standing wave ratio.
- STANDARDIZATION (ISO), International Organization for, 1994b. ISO 10534-2 , Acoustics - Determination of sound absorption coefficient and impedance in impedance tubes - Part1 : Transfer-function method.
- STANDARDIZATION (ISO), International Organization for, 2008. ISO 3382-2 : 2008 , Acoustics. Measurements of room acoustics parameters. Part2 : Reverberation time in ordinary rooms.
- STENDER, Merten ; ADAMS, Christian ; WEDLER, Mathies ; GREBEL, Antje ; HOFFMANN, Nobert, 2021. Explainable machine learning determines effects on the sound absorption coefficient measured in the impedance tube. *The Journal of the Acoustical Society of America*. T. 149, n° 3, p. 1932-1945.
- SUZUKI, Yoiti ; ASANO, Futoshi ; KIM, Hack-Yoon ; SONE, Toshio, 1995. An optimum computer-generated pulse signal suitable for the measurement of very long impulse responses. *The Journal of the Acoustical Society of America*. T. 97, n° 2, p. 1119-1123.
- SVENSSON, Peter ; KRISTIANSEN, Ulf R, 2002. Computational modelling and simulation of acoustic spaces. In : *Audio engineering society conference : 22nd international conference : Virtual, synthetic, and entertainment audio*. Audio Engineering Society.

- SZÖKE, Igor ; SKÁCEL, Miroslav ; MOŠNER, Ladislav ; PALIESEK, Jakub ; ČERNOCKÝ, Jan, 2019. Building and evaluation of a real room impulse response dataset. *IEEE Journal of Selected Topics in Signal Processing*. T. 13, n° 4, p. 863-876.
- TANG, Zhenyu ; ARALIKATTI, Rohith ; RATNARAJAH, Anton Jeran ; MANOCHA, Dinesh, 2022. GWA : A large high-quality acoustic dataset for audio processing. In : *ACM SIGGRAPH 2022 Conference Proceedings*, p. 1-9.
- THOMAS, Matthew Reuben, 2017. *Wayverb : A Graphical Tool for Hybrid Room Acoustics Simulation*. Thèse de doct. University of Huddersfield.
- THYDAL, Tobias ; PIND, Finnur ; JEONG, Cheol-Ho ; ENGSIG-KARUP, Allan P, 2021. Experimental validation and uncertainty quantification in wave-based computational room acoustics. *Applied Acoustics*. T. 178, p. 107939.
- TORRAS-ROSELL, Antoni ; JACOBSEN, Finn, 2010. Measuring long impulse responses with pseudorandom sequences and sweep signals. *Proc. 39th Int. Congr. Noise Control Eng. (INTER-NOISE 2010), Lisbon, Portugal*.
- VASWANI, Ashish ; SHAZEER, Noam ; PARMAR, Niki ; USZKOREIT, Jakob ; JONES, Llion ; GOMEZ, Aidan N ; KAISER, Lukasz ; POLOSUKHIN, Illia, 2017. Attention is all you need. *Advances in neural information processing systems*. T. 30.
- VERCAMMEN, Martijn, 2010. Improving the accuracy of sound absorption measurement according to ISO 354. In : *Proceedings of the International Symposium on Room Acoustics*. Melbourne, Australia, p. 29-31.
- VIGRAN, Tor Erik ; KELDERS, Luc ; LAURIKS, Walter ; LECLAIRE, Philippe ; JOHANSEN, TF, 1997. Prediction and measurements of the influence of boundary conditions in a standing wave tube. *Acta Acustica United with Acustica*. T. 83, n° 3, p. 419-423.
- VORLÄNDER, Michael, 2013. Computer simulations in room acoustics : Concepts and uncertainties. *The Journal of the Acoustical Society of America*. T. 133, n° 3, p. 1203-1213.
- WABNITZ, Andrew ; EPAIN, Nicolas ; JIN, Craig ; VAN SCHAIK, André, 2010. Room acoustics simulation for multichannel microphone arrays. In : *Proceedings of the International Symposium on Room Acoustics*. Citeseer, p. 1-6.
- WEBB, Craig J ; BILBAO, Stefan, 2011. Computing room acoustics with CUDA-3D FDTD schemes with boundary losses and viscosity. In : *2011 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, p. 317-320.
- XYDIS, Achilleas ; PERRAUDIN, Nathanaël ; RUST, Romana ; HEUTSCHI, Kurt ; CASAS, Gonzalo ; GROGNOUZ, Oksana Riba ; EGGENSCHWILER, Kurt ; KOHLER, Matthias ; PEREZ-CRUZ, Fernando, 2023. GIR dataset : A geometry and real impulse response dataset for machine learning research in acoustics. *Applied Acoustics*. T. 208, p. 109333.

- YU, Wangyang ; KLEIJN, W Bastiaan, 2020. Room Acoustical Parameter Estimation From Room Impulse Responses Using Deep Neural Networks. *IEEE/ACM Transactions on Audio, Speech, and Language Processing*. T. 29, p. 436-447.
- YUZAWA, Mutsushige, 1975. A method of obtaining the oblique incident sound absorption coefficient through an on-the-spot measurement. *Applied Acoustics*. T. 8, n° 1, p. 27-41.
- ZHANG, Shuying ; LEE, Joonhee, 2021. Diffuseness quantification in a reverberation chamber and its variation with fine-resolution measurements. *Buildings*. T. 11, n° 11, p. 519.

Apprentissage automatique et optimisation pour la détermination des propriétés acoustiques d'une salle à partir de signaux audio

Résumé

L'objectif de cette thèse est de contribuer à améliorer significativement le diagnostic acoustique de salles *in situ* en développant des nouvelles approches à la croisée de l'apprentissage automatique, de l'optimisation, du traitement du signal audio et de la physique acoustique. Les travaux présentés dans ce manuscrit portent sur l'estimation des propriétés d'absorption acoustique de chacune des parois d'une salle à partir de plusieurs réponses impulsionnelles de salle acquises pour des positions libres de sources et de microphones, ainsi que de mesures des paramètres géométriques et des réponses des appareils associés. Plusieurs méthodes sont proposées pour la résolution de ce problème inverse complexe en visant progressivement une description plus fine des propriétés d'absorption des parois, une application à des données simulées plus réalistes et une plus grande simplicité de mise en oeuvre. Ces travaux aboutissent à la présentation d'une approche d'estimation des réponses impulsionnelles des parois d'une salle par optimisation dans le domaine temporel. Celle-ci est fondée sur une nouvelle extension du modèle de réponse impulsionnelle de salle issu de la méthode des sources-images permettant de rendre compte des erreurs sur la mesure des paramètres géométriques, en plus de la dépendance spatio-fréquentielle des réponses des sources et des microphones.

Mots-clés: Apprentissage Automatique, Optimisation, Traitement du signal audio, Acoustique des salles, Problème inverse, Absorption, Réponse impulsionnelle de salle, Réponse impulsionnelle de mur

Abstract

The aim of this thesis is to contribute to significantly improve *in situ* acoustic diagnosis of rooms by developing new approaches at the crossroads of machine learning, optimization, audio signal processing and acoustics. The work presented in this manuscript focuses on estimating the acoustic absorption properties of each wall in a room from multiple room impulse responses acquired for free positions of sources and microphones, along with measurements of geometric parameters and responses of the associated devices. Several methods are proposed to solve this complex inverse problem, progressively aiming for a more detailed description of the absorption properties of the walls, application to more realistic simulated data, and greater ease of implementation. This work culminates in the presentation of an approach for estimating the wall's impulse responses by optimization in the time domain. This approach is based on a new extension of the room impulse response model derived from the image source method, which accounts for errors in the measurement of the geometric parameters, in addition to the spatial frequency dependence of the source and microphone responses.

Keywords: Machine learning, Optimization, Audio signal processing, Room acoustics, Inverse problem, Absorption, Room impulse response, Wall impulse response