

Titre: Égalisation adaptative et non invasive de la réponse temps-fréquence d'une petite salle
Title: Adaptive and non-invasive equalization of the time-frequency response of a small room

Auteur: Tristan Martin
Author:

Date: 2013

Type: Mémoire ou thèse / Dissertation or Thesis

Référence: Martin, T. (2013). Égalisation adaptative et non invasive de la réponse temps-fréquence d'une petite salle [Master's thesis, École Polytechnique de Montréal].
Citation: PolyPublie. <https://publications.polymtl.ca/1199/>

Document en libre accès dans PolyPublie

Open Access document in PolyPublie

URL de PolyPublie: <https://publications.polymtl.ca/1199/>
PolyPublie URL:

Directeurs de recherche: François-Raymond Boyer
Advisors:

Programme: Génie informatique
Program:

UNIVERSITÉ DE MONTRÉAL

ÉGALISATION ADAPTATIVE ET NON INVASIVE DE LA RÉPONSE
TEMPS-FRÉQUENCE D'UNE PETITE SALLE

TRISTAN MARTIN

DÉPARTEMENT DE GÉNIE INFORMATIQUE ET GÉNIE LOGICIEL
ÉCOLE POLYTECHNIQUE DE MONTRÉAL

MÉMOIRE PRÉSENTÉ EN VUE DE L'OBTENTION
DU DIPLÔME DE MAÎTRISE ÈS SCIENCES APPLIQUÉES
(GÉNIE INFORMATIQUE)
AOÛT 2013

UNIVERSITÉ DE MONTRÉAL

ÉCOLE POLYTECHNIQUE DE MONTRÉAL

Ce mémoire intitulé :

ÉGALISATION ADAPTATIVE ET NON INVASIVE DE LA RÉPONSE
TEMPS-FRÉQUENCE D'UNE PETITE SALLE

présenté par : MARTIN Tristan

en vue de l'obtention du diplôme de : Maîtrise ès Sciences Appliquées
a été dûment accepté par le jury d'examen constitué de :

M. LANGLOIS J.M. Pierre, Ph.D., président

M. BOYER François-Raymond, Ph.D., membre et directeur de recherche

M. CARDINAL Christian, Ph.D., membre

À ma mère . . .

REMERCIEMENTS

Mille fois merci à M. François-Raymond Boyer, Ph.D. Plus qu'un directeur de recherche, M. Boyer a su m'accompagner tout au long de ce travail, entre autres en prenant part à maintes discussions abordant une grande variété de sujets connexes à l'acoustique des salles et à l'audionumérique. M. Boyer a su partager son savoir-faire, bien m'encadrer dans mes travaux et publications, et surtout me guider dans ce présent travail de recherche.

RÉSUMÉ

Dans le cadre de cette recherche, on s'intéresse au son, à l'environnement dans lequel il se propage, à l'interaction entre l'onde de son et son canal de transmission ainsi qu'aux transformations induites par les composantes d'une chaîne audio. Le contexte précis étudié est celui de l'écoute musicale sur haut-parleurs.

Pour le milieu dans lequel l'onde se propage, comme pour tout canal de transmission, il existe des fonctions mathématiques permettant de caractériser les transformations induites par le canal sur un signal qui le traverse. Un signal électrique sert de signal d'excitation pour ce canal constitué en l'occurrence d'un amplificateur, d'un haut-parleur et de la salle dans laquelle a lieu l'écoute, qui selon ses caractéristiques, retourne en sortie à la position d'écoute une onde de son altérée. Réponse en fréquence, réponse à l'impulsion, fonction de transfert ; les mathématiques utilisées ne diffèrent en rien de celles servant communément à la caractérisation d'un canal de transmission ou à l'expression des fonctions liant les sorties d'un système linéaire à ses entrées.

Naturellement, il y a un but à cet exercice de modélisation : l'obtention de la réponse de la chaîne amplificateur/salle/haut-parleur rend possible sa correction. Il est commun dans bien des contextes d'écoute, qu'un filtre soit inséré dans la chaîne audio entre la source (exemple : lecteur CD) et le haut-parleur qui transforme le signal électrique en signal acoustique propagé dans la salle. Ce filtre, dit « égalisateur », a pour but de compenser en fréquences l'effet des composantes de la chaîne audio et de la salle sur le signal sonore y étant transmis. Ses propriétés découlent de celles de l'amplificateur, du haut-parleur et de la salle.

Bien qu'analytiquement rigoureuse, l'approche physique, centrée sur la modélisation physique du haut-parleur et sur l'équation de propagation de l'onde acoustique, est mal adaptée aux salles à géométrie complexe ou changeante au fil du temps. La seconde approche, la modélisation expérimentale, abordée dans ce travail, fait abstraction des propriétés physiques. La chaîne amplificateur/haut-parleur/salle y est plutôt vue comme une « boîte noire » comprenant entrées et sorties. Le problème étudié est celui de la caractérisation d'un système électro-acoustique ayant comme unique entrée un signal émis à travers un haut-parleur dans une salle, et comme unique sortie le signal capté par un microphone placé à la position d'écoute.

L'originalité de ce travail réside non seulement dans la technique développée pour en arriver à cette caractérisation, mais surtout dans les contraintes imposées dans la manière d'y arriver. La majorité des techniques documentées à ce jour font appel à des signaux d'excitation dédiés à la mesure ; des signaux dotés de caractéristiques favorables à la simplification du

calcul de réponse impulsionale qui en découle. Des signaux connus sont émis à travers un haut-parleur et la réponse à leur excitation est captée à l'aide d'un microphone à la position d'écoute. L'exercice de mesure lui-même pose problème, notamment, lorsqu'un auditoire est présent dans la salle. Aussi, la réponse de la salle peut changer entre le moment de la prise de mesure et l'écoute si la salle est reconfigurée, par exemple un rideau est tiré ou une estrade déplacée. Dans le cas d'une salle de spectacle, le haut-parleur utilisé peut varier selon le contexte.

Un recensement des travaux dans lesquels des solutions à ce problème sont suggérées fut effectué. Le principal objectif est de développer une méthode innovatrice permettant de capturer la réponse impulsionale de la chaîne audio à l'insu de l'auditoire. Pour ce faire, aucun signal dédié à la mesure ne doit être utilisé. La méthode développée permet la capture de la réponse impulsionale électro-acoustique en n'exploitant que les signaux musicaux. Le résultat est, un algorithme permettant la modélisation dynamique et en continu de la réponse d'une salle. Un filtre égalisateur numérique à réponse impulsionale finie doit être conçu, lui aussi capable de s'adapter dynamiquement au comportement de la salle, même lorsque celui-ci varie au fil du temps.

La familiarisation avec des concepts plus avancés de programmation C++ orienté objet étant de mise, une technique permettant d'exploiter des signaux musicaux afin d'obtenir la réponse impulsionale et la réponse en fréquence du système fut testée expérimentalement sous forme d'un module VST. L'excitation est procurée par les signaux musicaux émis sur haut-parleurs durant l'écoute. Une moyenne mobile pondérée reconstruit statistiquement, au fil du temps, la réponse de la salle sur toute la plage de fréquences audibles. Dans le but d'en quantifier la performance, la réponse en fréquence obtenue est comparée à celle obtenue par une méthode standard servant de référence. L'erreur quadratique moyenne sert de métrique d'erreur et montre que plus la musique défile, plus la réponse en fréquence obtenue s'apparente à la référence pour un même point d'écoute.

Une approche à résolution spectrale variable est utilisée pour construire, par bandes de fréquences, la réponse du filtre découlant de celle de la chaîne audio. La réponse en fréquence du système corrigée par le filtre égalisateur est plus plane que celle du système initial.

Des techniques explorées dans le cadre de ce travail de recherche ont mené à la publication d'un article scientifique dans une revue à comité de lecture et un article de conférence dans lesquels des méthodes similaires furent exploitées en génie des mines.

ABSTRACT

In this research, we are interested in sound, environment wherein it propagates, the interaction between the sound wave and a transmission channel, and the changes induced by the components of an audio chain. The specific context studied is that of listening to music on loudspeakers.

For the environment in which sound wave propagates, like for any transmission channel, there are mathematical functions used to characterize the changes induced by a channel on the signal therethrough. An electric signal serves as a input for a system, in this case consisting of an amplifier, a loudspeaker, and the room where the listening takes place, which according to its characteristics, returns as an output at the listening position, an altered sound wave. Frequency response, impulse response, transfer function, the mathematics used are no different from those used commonly for the characterization of a transmission channel or the expression of the outputs of a linear system to its inputs.

Naturally, there is a purpose to this modeling exercise: getting the frequency response of the amplifier/loudspeaker/room chain makes possible its equalization. It is common in many contexts of listening to find a filter inserted into the audio chain between the source (Eg CD player) and the amplifier/loudspeaker that converts the electrical signal to an acoustic signal propagated in the room. This filter, called “equalizer” is intended to compensate the frequency effect of the components of the audio chain and the room on the sound signal that will be transmitted. Properties for designing this filter are derived from those of the audio chain.

Although analytically rigorous, physical approach, focusing on physical modeling of the loudspeaker and the propagation equation of the acoustic wave is ill-suited to rooms with complex geometry and changing over time. The second approach, experimental modeling, and therefore that addressed in this work, ignores physical properties. The system audio chain is rather seen as a “black box” including inputs and outputs. The problem studied is the characterization of an electro-acoustic system as having a single input signal transmitted through a speaker in a room, and a single output signal picked up by a microphone at the listening position.

The originality of this work lies not only in the technique developed to arrive at this characterization, but especially in the constraints imposed in order to get there. The majority of technics documented to this date involve using excitation signals dedicated the measure; signals with favorable characteristics to simplify the calculation of the impulse response of the audio chain. Known signals are played through a loudspeaker and the room’s response to

excitation is captured with a microphone at the listening position. The measurement exercise itself poses problem, especially when there is an audience in the room. Also, the response of the room may change between the time of the measurement and time of listening. If the room is reconfigured for example, a curtain is pulled or the stage moved. In the case of a theater, the speaker used may vary depending on the context.

A survey of work in which solutions to this problem are suggested was made. The main objective is to develop an innovative method to capture the impulse response of an audio chain without the knowledge of the audience. To do this, no signal dedicated to the measurement should be used. The developed method allows the capture of the electro-acoustic impulse response exploiting only the music signals when it comes to a concert hall or using a movie sound track when a movie is a movie theater. As a result, an algorithm for modeling dynamically and continuously the response of a room. A finite impulse response filter acting as a digital equalizer must be designed and also able to dynamically adapt the behavior of the room, even when it varies over time.

Familiarization with more advanced programming concepts of object-oriented C++ being put, a technique to exploit music signals to obtain the impulse response and frequency response of the audio chain was implemented as a VST module and tested experimentally. The excitation is provided by music signals played through speakers. Using a weighted moving average reconstructed statistically over time, the response of the room on the entire audible frequency range is obtained. In order to quantify the performance the frequency response obtained is compared with that obtained by using a standard reference method. The mean square error is used as an error metric and shows that more music scrolls, more the frequency response obtained is similar to the reference one for the same listening position.

A multi spectral resolution method is used to build, for different frequency bands, the filter response arising from the inversion of the room/speaker frequency response. The resulting dynamically adapting filter has properties similar to those of the human ear, a significant spectral-resolution in lower frequencies, and high time-resolution at high frequencies. The response corrected by the filter system tends approaching to a pure pulse.

Techniques explored in the context of this research led to the publication of a scientific article in a peer reviewed journal and one conference paper in which similar methods were used for mining engineering applications.

TABLE DES MATIÈRES

DÉDICACE	iii
REMERCIEMENTS	iv
RÉSUMÉ	v
ABSTRACT	vii
TABLE DES MATIÈRES	ix
LISTE DES TABLEAUX	xi
LISTE DES FIGURES	xii
LISTE DES ANNEXES	xvii
LISTE DES SIGLES ET ABRÉVIATIONS	xviii
 CHAPITRE 1 INTRODUCTION	
1.1 Éléments de la problématique	1
1.2 Objectifs de recherche	5
1.3 Plan du mémoire	7
 CHAPITRE 2 DÉFINITIONS, CONCEPTS DE BASE ET RAPPELS	
2.1 Définitions	10
2.2 Concepts de base	11
2.3 Rappels	13
 CHAPITRE 3 REVUE DE LITTÉRATURE	
3.1 Historique de l'acoustique comme science	16
3.2 Le domaine fréquentiel	16
3.3 Modélisation expérimentale	18
3.4 Réponse impulsionnelle et réponse en fréquence	18
3.5 L'égalisation	19
3.6 Égalisation adaptative à l'insu de l'auditoire	21

CHAPITRE 4 DÉTAILS DE LA SOLUTION	23
4.1 Analogie : modélisation d'un moteur	24
4.2 La chaîne audio à modéliser	25
4.3 La réponse impulsionale	28
4.4 Technique de modélisation de référence	30
4.5 Technique de modélisation suggérée	33
4.6 Approche conventionnelle : domaine temporel	46
4.7 Approche alternative adaptée à la réponse en fréquence d'une salle	47
4.8 Conception du filtre égalisateur	49
4.9 Fonctionnement du filtre convolutif	56
CHAPITRE 5 IMPLÉMENTATION LOGICIELLE ET RÉSULTATS EXPÉRIMENTAUX	62
5.1 Définition du critère de performance	62
5.2 Prototypage logiciel de l'algorithme	64
5.3 Implémentation logicielle finale de l'algorithme	69
5.4 Environnement de test	72
5.5 Matériel utilisé	74
5.6 Fonctionnement du module VST dans l'hôte « VST-Host »	77
5.7 Ratio signal sur bruit de la chaîne	83
5.8 Obtention du modèle de référence	84
5.9 Performance de l'algorithme de modélisation suggéré	88
5.10 Performance du filtre égalisateur	96
CHAPITRE 6 APPLICATION DE LA MÉTHODE AU PROBLÈME DE L'ACOUSTIQUE DANS UN BROYEUR	108
CHAPITRE 7 CONCLUSION	112
7.1 Synthèse des travaux	112
7.2 Limitations de la solution proposée	113
7.3 Améliorations futures	114
7.4 Applications éventuelles	115
RÉFÉRENCES	116
ANNEXES	119

LISTE DES TABLEAUX

Tableau 4.1	Longueur et point de transition de chaque bande	52
Tableau 5.1	Séquencement des opérations de l'implémentation logicielle	71
Tableau 5.2	Caractéristiques de la salle de régie	73
Tableau 5.3	Paramètres du plugin de la figure 5.19	82

LISTE DES FIGURES

Figure 1.1	Six domaines touchés par l'étude de l'acoustique d'une salle	2
Figure 1.2	Système à l'étude	2
Figure 1.3	Système avec filtre égalisateur statique	3
Figure 1.4	Mesure de la réponse du système	6
Figure 1.5	Système avec filtre égaliseur adaptatif	7
Figure 1.6	Bloc de modélisation de la chaîne audio	8
Figure 1.7	Bloc de conception du filtre égalisateur	8
Figure 1.8	Bloc du filtre convolutif	8
Figure 4.1	Moteur électrique et tachymètre	24
Figure 4.2	Moteur à courant continu représenté dans le domaine temporel par sa réponse impulsionnelle $h(t)$	25
Figure 4.3	Chaîne audio et microphone de mesure	25
Figure 4.4	Chaîne audio	26
Figure 4.5	Système dans le domaine	26
Figure 4.6	Passage du domaine continu au domaine discret	27
Figure 4.7	Système discret	27
Figure 4.8	Réponse d'un système à l'impulsion	28
Figure 4.9	Balayage sinusoïdal de la plage de fréquence	31
Figure 4.10	Réponse de la chaîne audio au balayage sinusoïdal de la plage de fréquence	32
Figure 4.11	Approximation de la réponse d'un système discret par déconvolution circulaire	35
Figure 4.12	Implémentation de l'algorithme de modélisation de la chaîne audio . .	37
Figure 4.13	Bruit blanc gaussien de 0,186 ms (échelle de couleur en dBFS)	39
Figure 4.14	Spectrogramme d'une section de 0,186 seconde prise au hasard dans la chanson « Wickerman » (échelle de couleur en dBFS)	39
Figure 4.15	Chaîne audio idéale	40
Figure 4.16	Impact du plancher de bruit sur la mesure	41
Figure 4.17	Pondérations décroissantes pour différentes valeurs d' α	44
Figure 4.18	Choix d' α pour que 95% de la pondération provienne des 20 dernières secondes	45
Figure 4.19	Pondérations décroissantes appliquées aux anciens blocs de données selon la valeur d' α choisie	46
Figure 4.20	Implémentation de la moyenne mobile exponentielle pondérée	47

Figure 4.21	Erreurs sur le spectre d'amplitude	48
Figure 4.22	Réponse impulsionnelle de la banque de filtres en octave	53
Figure 4.23	Réponse en fréquence de la banque de filtres en octave	54
Figure 4.24	Fenêtres temporelles selon la bande de fréquences	55
Figure 4.25	Temps de coupure de la fenêtre appliquée à une bande et longueur effective de cette bande	56
Figure 4.26	Implémentation de la conception du filtre égalisateur	57
Figure 4.27	Implémentation du filtre convolutif	58
Figure 4.28	Filtre à réponse impulsionnelle finie	59
Figure 4.29	Méthode « Overlap-Add » pour la convolution linéaire	60
Figure 4.30	Chaîne audio avec filtre adaptatif pour l'égalisation des fréquences . .	61
Figure 5.1	Signaux audio dans l'environnement Matlab	65
Figure 5.2	Séquencement où le temps de calcul est maître	67
Figure 5.3	Réponse en fréquence obtenue avec le prototype d'implémentation Matlab	68
Figure 5.4	EQM la référence obtenue par balayage sinusoïdal et la méthode de modélisation suggérée dans le prototype d'implémentation Matlab . .	69
Figure 5.5	Séquencement où le calcul doit être terminé dans le temps alloué . . .	70
Figure 5.6	Signaux audio dans l'environnement VST	73
Figure 5.7	Salle de régie du Studio CME	74
Figure 5.8	Microphone de mesure ECM8000 de Behringer, image tirée de la documentation du fabricant (ecm, 2000)	75
Figure 5.9	Patron polaire omnidirectionnel du ECM8000, image tirée de la documentation du fabricant (ecm, 2000)	76
Figure 5.10	Réponse en fréquence du ECM8000, image tirée de la documentation du fabricant (ecm, 2000)	77
Figure 5.11	Vue avant de la carte de son Focusrite Scarlett 2i2, image tirée de la documentation du fabricant (2i2, 2011)	77
Figure 5.12	Configuration du périphérique audio « Focusrite »	78
Figure 5.13	Interface graphique VST-Host	79
Figure 5.14	Branchements des plugins dans le « VST-Host » pour la réalisation de tests expérimentaux	79
Figure 5.15	Interface graphique du plugin « SamSolo »	80
Figure 5.16	Entrée numérique microphone	80
Figure 5.17	Entrée numérique audio	81
Figure 5.18	Entrée-numérique	81
Figure 5.19	Interface graphique du plugin d'égalisation adaptative développé . .	82

Figure 5.20	Enregistrement d'un bruit blanc à pleine amplitude utilisé pour le calcul du RSB	83
Figure 5.21	Enregistrement du bruit de fond utilisé pour le calcul du RSB	83
Figure 5.22	Rapport signal sur bruit de la chaîne audio utilisée pour la mesure	84
Figure 5.23	Signal de balayage sinusoïdal	85
Figure 5.24	Spectrogramme du signal de balayage sinusoïdal (échelle de couleur en dBFS)	85
Figure 5.25	Réponse au balayage sinusoïdal	86
Figure 5.26	Spectrogramme de la réponse au balayage sinusoïdal (échelle de couleur en dBFS)	86
Figure 5.27	Réponse impulsionale servant de référence obtenue par balayage sinusoïdal	87
Figure 5.28	Suppression du délai dans la réponse impulsionale obtenue par balayage sinusoïdal	87
Figure 5.29	Réponse impulsionale obtenue après 10 blocs de données traités	89
Figure 5.30	Réponse en fréquence obtenue après 10 blocs de données traités	90
Figure 5.31	Erreur quadratique moyenne sur la réponse en fréquence lorsque le filtre est désactivé	91
Figure 5.32	Erreur quadratique moyenne sur la réponse en fréquence lorsque le filtre est activé	92
Figure 5.33	Impact de l'activation du filtre sur l'erreur quadratique moyenne sur la réponse en fréquence	93
Figure 5.34	EQM (sans et avec filtre) accompagnée du spectrogramme de l'extrait de la pièce « Wikerman »	94
Figure 5.35	EQM (sans et avec filtre) accompagnée du spectrogramme de l'extrait de la pièce « Stricken »	94
Figure 5.36	Effet de la distance entre le haut-parleur et la position d'écoute sur la capacité à estimer la salle	95
Figure 5.37	Fenêtrage de la réponse impulsionale pour la bande A (9454 à 22050 Hz)	96
Figure 5.38	Fenêtrage de la réponse impulsionale pour la bande B (4663 à 9454 Hz)	97
Figure 5.39	Fenêtrage de la réponse impulsionale pour la bande C (2318 à 4663 Hz)	97
Figure 5.40	Fenêtrage de la réponse impulsionale pour la bande D (1158 à 2318 Hz)	98

Figure 5.41	Fenêtrage de la réponse impulsionnelle pour la bande E (580.6 à 1158 Hz)	98
Figure 5.42	Fenêtrage de la réponse impulsionnelle pour la bande F (292.7 à 580.6 Hz)	99
Figure 5.43	Fenêtrage de la réponse impulsionnelle pour la bande G (149.0 à 292.7 Hz)	99
Figure 5.44	Fenêtrage de la réponse impulsionnelle pour la bande H (77.12 à 149.0 Hz)	100
Figure 5.45	Fenêtrage de la réponse impulsionnelle pour la bande H (41.25 à 77.12 Hz)	100
Figure 5.46	Fenêtrage de la réponse impulsionnelle pour la bande H (0 à 41.25 Hz)	101
Figure 5.47	Réponse impulsionnelle complète (direct et réverbéré) et réponse impulsionnelle après fenêtrage par bandes d'octave (direct seulement) .	101
Figure 5.48	Réponse en fréquence complète (direct et réverbéré) et réponse en fréquence après fenêtrage par bandes (direct seulement)	102
Figure 5.49	Réponse en fréquence de la salle (direct seulement) et réponse en fréquence du filtre égalisateur	103
Figure 5.50	Réponse impulsionnelle de la salle (direct seulement) avant et après égalisation	103
Figure 5.51	Réponse en fréquence de la salle (direct seulement) avant et après égalisation	104
Figure 5.52	Réponse en fréquence de la salle (direct et réverbéré) avant et après égalisation	105
Figure 5.53	EQM entre la réponse en fréquence de la salle (direct seulement) et une réponse idéale (plate)	106
Figure 5.55	Capacité d'adaptation de la correction de la réponse de la salle lorsque celle-ci change soudainement	106
Figure 5.54	Effet de la distance entre le haut-parleur et la position d'écoute sur la capacité à égaliser la salle	107
Figure 6.1	Broyeur utilisé pour l'expérience (figure tirée de Hosseini <i>et al.</i> (2011))	109
Figure 6.2	Position du microphone et endroits percutés avec le marteau (figure tirée de Hosseini <i>et al.</i> (2011))	110
Figure 6.3	Force d'impact du marteau et pression acoustique résultante	111
Figure A.1	EQM (sans et avec filtre) accompagnée du spectrogramme de l'extrait de la pièce « Wickerman », du groupe « Iron Maiden », sur l'album « Brave New World », paru en 2000	119

Figure A.2	EQM (sans et avec filtre) accompagnée du spectrogramme de l'extrait de la pièce « Stricken », du groupe « Disturbed », sur l'album « Ten Thousand Fists », paru en 2005	120
Figure A.3	EQM (sans et avec filtre) accompagnée du spectrogramme de l'extrait de la pièce « Blood Brothers », du groupe « Iron Maiden », sur l'album « Brave New World », paru en 2000	120
Figure A.4	EQM (sans et avec filtre) accompagnée du spectrogramme de l'extrait de la pièce « Brave new world », du groupe « Iron Maiden », sur l'album « Brave New World », paru en 2000	121
Figure A.5	EQM (sans et avec filtre) accompagnée du spectrogramme de l'extrait de la pièce « Funny the way it is », du groupe « Dave Matthews Band's », sur l'album « Big Whiskey & the GrooGrux King », paru en 2009	121
Figure A.6	EQM (sans et avec filtre) accompagnée du spectrogramme de l'extrait de la pièce « Roxanne », du groupe « The Police », sur l'album « Outlandos d'Amour », paru en 1978	122

LISTE DES ANNEXES

Annexe A	EQM (sans et avec filtre) accompagnée du spectrogramme des six extraits musicaux	119
----------	--	-----

LISTE DES SIGLES ET ABRÉVIATIONS

A/N	Conversion analogique-numérique
dB	Décibel
dBFS	Décibel pleine échelle (de l'anglais « dB Full Scale »)
dB SPL	Décibel de pression acoustique (de l'anglais « dB Sound Pressure Level »)
éch	Échantillon(s)
EQ	Égalisateur
EQM	Erreur quadratique moyenne (en anglais « Mean Square Error » ou MSE)
f éch	Fréquence d'échantillonnage
FFT	Transformée de Fourier rapide (de l'anglais « Fast Fourier Transform »)
HP	Haut-parleur
Hz	Hertz
Mic	Microphone
MLS	Séquence de longueur maximale (de l'anglais « Maximum length sequence »)
MMEP	Moyenne mobile exponentielle pondérée
N/A	Conversion numérique-analogique
TFD	Transformée de Fourier discrète
TFDI	Transformée de Fourier discrète inverse
RI	Réponse impulsionnelle
RIF	Réponse impulsionnelle finie
RII	Réponse impulsionnelle infinie
RSB	Rapport signal sur bruit
VST	de l'anglais « Virtual Studio Technology »

CHAPITRE 1

INTRODUCTION

Il est question dans ce travail de la présentation d'une nouvelle technique permettant d'obtenir la réponse impulsionale et la réponse en fréquence d'une salle, ou d'une manière plus générale, une chaîne audionumérique complète. Par chaîne audio, on entend l'ensemble des composantes nécessaires à l'écoute de musique sur haut-parleur à partir d'une source audio numérique, ou d'une source analogique numérisée. La salle dans laquelle a lieu l'écoute fait elle aussi partie de cette chaîne. La technique suggérée doit être capable d'estimer la réponse de la chaîne audio (incluant la salle) durant l'écoute musicale et à l'insu de l'auditoire. Le but de l'obtention de cette réponse est la correction de cette dernière à l'aide d'un filtre égalisateur numérique inséré dans la chaîne dans le but de rendre la réponse à l'impulsion plus près d'une vraie impulsion et donc la réponse en fréquence plus plate. Le filtre assurant cette correction doit être adaptatif, c'est-à-dire que sa réponse doit s'adapter automatiquement à celle de la chaîne audio lorsque celle-ci change.

Ce chapitre d'introduction compte quatre sections. On commence dans 1.1 par définir plus en détail le problème traité, la manière avec laquelle il a été abordé dans le passé et la manière de l'aborder dans ce présent ouvrage. Ensuite, 1.2 contient les différents objectifs découlant de cette recherche. Les objectifs en lien direct avec la problématique y seront présentés mais aussi des objectifs secondaires découlant indirectement de la réalisation de ce travail. Finalement, 1.3 contient un court plan du mémoire.

1.1 Éléments de la problématique

Le sujet traité dans le cadre de cette recherche relève du domaine de l'acoustique ; de l'acoustique architecturale plus exactement. On s'intéresse au son, à l'environnement dans lequel il se propage, à l'interaction entre l'onde de son et son canal de transmission. Le contexte précis étudié est celui de l'écoute musicale sur haut-parleurs.

L'acoustique est en fait une branche de la physique dont l'objet est l'étude du son, de sa génération à sa perception, en passant par sa propagation. Cette science riche comporte une panoplie de champs d'investigation : physiologie de l'audition, linguistique et communication, acoustique musicale, psychoacoustique, acoustique architecturale, traitement de signal et acoustique des salles, pour n'en nommer que quelques-uns.

L'acoustique des salles est une branche du domaine plus large de l'acoustique architecturale.

turale, connue comme étant le domaine scientifique et technologique qui vise à comprendre et maîtriser la qualité sonore des bâtiments. On distingue généralement trois phénomènes relatifs à l'acoustique des salles, l'aspect « objectif » étant celui visé par ce travail et montré dans la figure 1.1 :

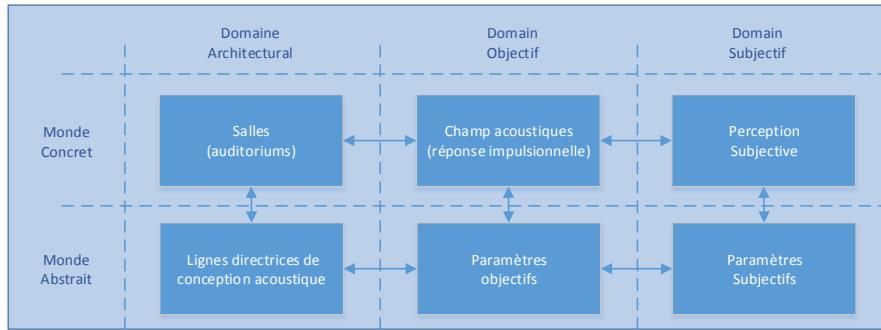


Figure 1.1 Six domaines touchés par l'étude de l'acoustique d'une salle

Pour tout canal de transmission, il existe des fonctions mathématiques permettant de caractériser les transformations induites par le canal sur un signal qui le traverse. Comme pour tout système, il existe des fonctions caractéristiques décrivant la réaction de ce dernier à une excitation donnée. Dans le présent travail, le canal de transmission (le système) étudié est le milieu dans lequel l'onde acoustique se propage. Dans le cas de l'écoute musicale sur haut-parleurs, il s'agit de la salle à l'intérieur de laquelle a lieu l'écoute. L'onde de son émise à travers un haut-parleur sert de signal d'excitation pour ce système qui, selon ses caractéristiques, retourne en sortie à la position d'écoute une onde de son altérée. La figure 1.2 montre le système à l'étude.

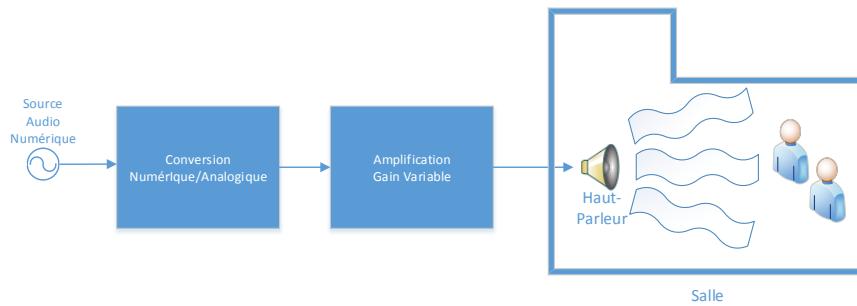


Figure 1.2 Système à l'étude

Facteur d'atténuation, réponse en fréquence, réponse à l'impulsion, fonction de transfert ; les mathématiques permettant de décrire l'impact d'une salle sur l'onde de son qui s'y propage ne diffèrent en rien de celles communément utilisées pour la caractérisation d'un canal de transmission ou pour exprimer les fonctions liant les sorties d'un système à ses entrées.

Naturellement, il y a un but à cet exercice de modélisation : l'obtention de la réponse de la salle rend possible sa correction. Il est commun dans bien des contextes d'écoute, qu'un filtre soit inséré dans la chaîne audio entre la source (exemple : lecteur CD) et les haut-parleurs transformant le signal électrique en signal acoustique propagé dans la salle. Ce filtre, dit « égalisateur », a pour but de compenser l'effet de la salle sur le signal sonore y étant transmis. Dans un tel cas, le filtre doit disposer de propriétés découlant de celles de la salle. La figure 1.3 montre l'insertion d'un filtre égalisateur numérique dans une chaîne audio.

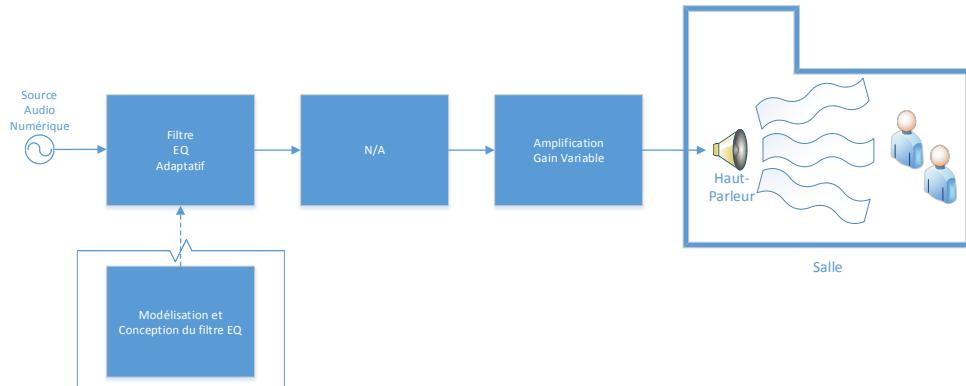


Figure 1.3 Système avec filtre égalisateur statique

Deux écoles principales de pensée existent en ce qui a trait à la caractérisation de l'acoustique d'une salle : la modélisation physique et la modélisation expérimentale. La première suggère une approche analytique centrée sur la physique de la propagation de l'onde acoustique. Certaines méthodes de modélisation qui en découlent s'attardent sur l'interaction entre l'onde acoustique et les matériaux composant les différentes surfaces qu'expose l'intérieur de la salle dans le but de caractériser la capacité de la salle à dissiper de l'énergie acoustique. Une telle approche peut être appliquée à l'acoustique des petites salles.

Quand considérons-nous une salle petite ? Il est tout d'abord crucial de mentionner que le fait de considérer une salle petite est arbitraire. Typiquement, une salle est acoustiquement dite petite lorsque ses dimensions sont de même ordre de grandeur que la longueur d'onde acoustique qui s'y propage.

Dans chaque salle existent des endroits appelés nuls où certaines fréquences sont fortement

atténuées. Ces nuls ne sont souvent observés qu'à une distance égale aux multiples impairs du quart de la longueur d'onde des murs, plafond et plancher, soit $1/4 \lambda$, $3/4 \lambda$, $5/4 \lambda$, $9/4 \lambda$ etc. Similairement, des gains élevés sont observables à des distances des murs, plafond et plancher, égales aux multiples paires de la longueur d'onde, soit $2/4 \lambda$, $4/4 \lambda$, $6/4 \lambda$, etc. Ces longueurs d'onde sont appelées « mode de résonance », ou « mode propres ». Cet effet se fait sentir d'une manière encore plus prononcée lorsqu'un point de l'espace est à la fois à la même distance du mur, du plancher et du plafond. Cette situation se présente particulièrement dans le cas de petites salles à géométrie cubique où le centre de la pièce se trouve à la même distance des six surfaces qui la délimitent. Pour une petite salle parallélépipède, les fondamentales des modes de résonance se présentent généralement sur une bande de fréquence particulièrement basse, mais tout de même audible chez l'humain, soit de 20 à 150 Hz. C'est le fait que ces modes de résonance se situent dans cette plage de fréquences qui rend la salle petite, d'un point de vue acoustique.

La modélisation physique est préférée lorsque la géométrie de la salle peut être approximée par la combinaison de formes bien connues, par exemple une salle rectangulaire ou parabolique. Cependant, bien qu'analytiquement rigoureuse, l'approche physique est mal adaptée aux salles à géométrie complexe et changeante au fil du temps, par exemple une salle de spectacle à géométrie complexe, dans laquelle les décors changent, des portes s'ouvrent, d'autres se ferment, les gens se déplacent.

Par opposition, la seconde école de pensée favorise les méthodes faisant abstraction des propriétés physiques. La salle y est plutôt vue comme une « boîte noire » faite d'entrées, sources sonores telles que des haut-parleurs, et de sorties, par exemple des microphones capturant le son perçu à une position donnée. Les méthodes de modélisation en découlant caractérisent le milieu de propagation expérimentalement par la comparaison de l'onde de son émise en entrée à celle observée en sortie. Ces méthodes sont beaucoup mieux adaptées aux milieux de propagation à géométrie complexe.

La figure 1.4 illustre l'exercice de modélisation suggérée par les techniques standards de modélisation expérimentale permettant d'obtenir la réponse en fréquence du système permettant de configurer un filtre égalisateur tel que celui de la figure 1.3. L'exercice de modélisation expérimentale fait abstraction des propriétés physiques du canal de transmission et du milieu de propagation acoustique. Il ne s'intéresse qu'à la transformation induite par ce canal sur un signal de test y étant émis.

Le présent travail porte sur les techniques de modélisation expérimentale et ce sont principalement ces dernières qui sont couvertes par la revue de littérature effectuée. Le problème étudié est celui de la caractérisation d'un système acoustique ayant comme unique entrée un signal émis à travers un haut-parleur dans une salle, et comme unique sortie le signal

capté par un microphone placé à la position d'écoute. Une méthode innovatrice permettant de modéliser l'acoustique de la salle est suggérée. Du modèle obtenu découlent les paramètres dynamiques d'un filtre égalisateur voulant compenser l'effet de la salle sur l'onde de son qui s'y propage.

1.2 Objectifs de recherche

Le présent travail a été rédigé dans le cadre d'un projet de recherche universitaire de deuxième cycle en génie informatique.

Les objectifs vont naturellement au-delà de la simple caractérisation d'une salle comme milieu de propagation acoustique. L'originalité de ce travail réside non seulement dans la technique développée pour en arriver à cette caractérisation, mais surtout dans les contraintes imposées dans la manière d'y arriver. La majorité des techniques documentées à ce jour pour la mesure de réponses impulsionales de salles ont une chose en commun. Elles font appel à des signaux d'excitation dédiés à la mesure ; des signaux dotés de caractéristiques favorables à la simplification du calcul de réponse impulsionnelle qui en découle. Pour ces techniques classiques, des signaux connus sont émis à travers un haut-parleur et la réponse de la salle à leur excitation est captée à l'aide d'un microphone à la position d'écoute. La figure 1.4 décrit l'exercice de mesure.

Quels que soient ces signaux et le calcul permettant d'extraire la réponse de la salle, l'exercice de mesure lui-même pose problème. Notamment, la réponse de la salle peut changer entre la prise de mesure, effectuée par exemple dans une salle vide, et l'écoute, lorsque par exemple la salle est occupée par un auditoire. La reconfiguration du décor ou des meubles dans la salle, le déplacement de haut-parleurs, ou même le simple déplacement des gens ; les facteurs pouvant altérer la réponse de la salle sont multiples. Dans le cas de techniques de modélisation utilisant des signaux dédiés, l'exercice de mesure doit recommencer chaque fois que la configuration de la salle se voit modifiée, ce qui pose problème principalement lorsqu'un auditoire est présent dans la salle. L'exercice de mesure n'a pas toujours sa place.

Un des objectifs de ce travail est le recensement des travaux effectués à ce jour dans lesquels des solutions aux problèmes de la modélisation expérimentale sont suggérées, notamment l'impossibilité d'utiliser des signaux dédiés à la mesure en présence d'un auditoire. Dans les ouvrages répertoriés, les paramètres extraits ne sont pas suffisants pour faire la conception d'un filtre égalisateur.

Le principal objectif est de développer une méthode innovatrice permettant de capturer la réponse impulsionnelle d'une salle à l'insu de l'auditoire. Pour ce faire, aucun signal dédié à la mesure ne doit être utilisé. La méthode développée permet la capture de la réponse

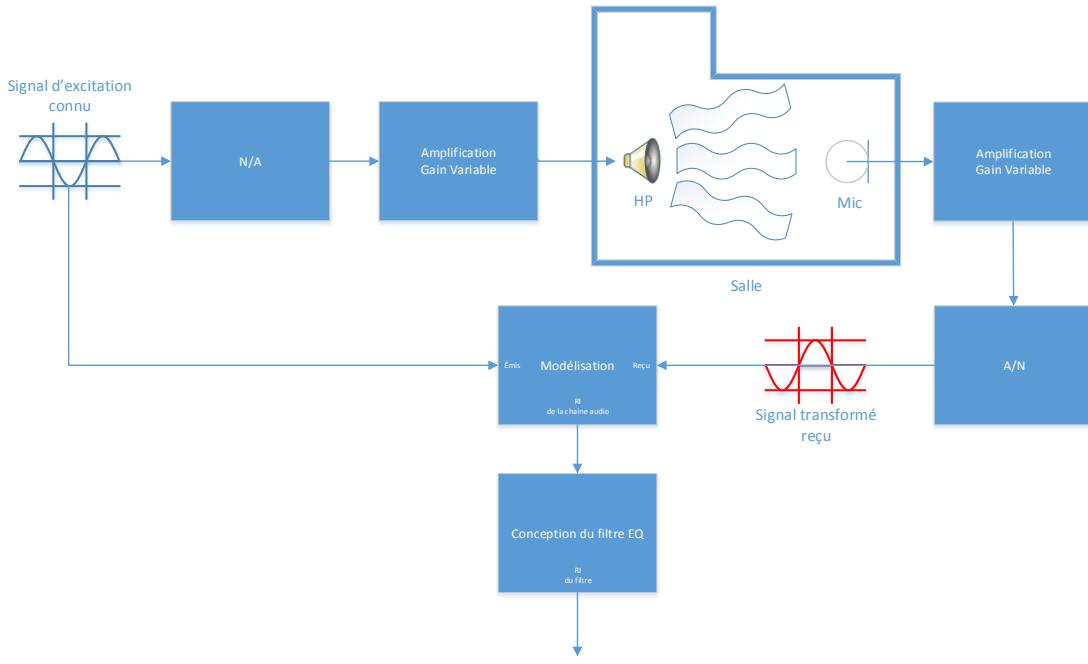


Figure 1.4 Mesure de la réponse du système

impulsionnelle acoustique en n'exploitant que les signaux musicaux lorsqu'il s'agit d'une salle de spectacle ou la régie d'un studio d'enregistrement, ou encore en utilisant la trame sonore d'un film lorsqu'il s'agit d'une salle de cinéma. En résultat, un algorithme permettant, à l'insu de l'auditoire, la modélisation dynamique et en continu de la réponse d'une salle, même lorsque celle-ci varie au fil du temps.

Finalement, on doit concevoir un filtre égalisateur numérique à réponse impulsionnelle finie, lui aussi capable de s'adapter dynamiquement au comportement de la salle pouvant évoluer au fil du temps. La plage de fréquences visée par un tel filtre égalisateur est limitée à la bande passante de l'oreille humaine (20 à 20 000 Hz).

La figure 1.5 montre comment un algorithme dynamique de modélisation ainsi qu'un filtre adaptatif peuvent être combinés afin de corriger automatiquement et dynamiquement la réponse d'une salle d'écoute.

Finalement, la familiarisation avec les concepts plus avancés de programmation orientée objet fait aussi partie des objectifs. Le C/C++ fut choisi et le standard Virtual Studio Technology (VST) de Steinberg fut adopté pour l'implémentation d'un module VST intégrant l'algorithme développé. Une technique permettant d'exploiter des signaux quelconques afin d'obtenir la réponse impulsionnelle et la réponse en fréquence de la salle est implémentée

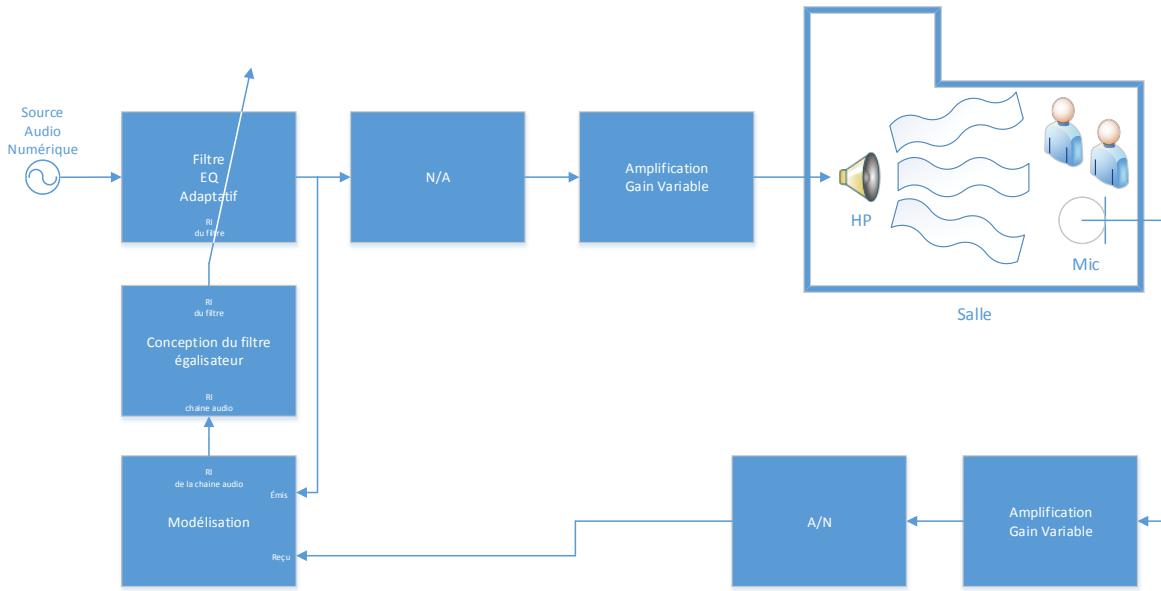


Figure 1.5 Système avec filtre égaliseur adaptatif

en C++ et testée expérimentalement à l'aide d'un logiciel hôte pour modules VST. Le logiciel hôte utilisé est VST Host 1.53 par Hermann Seib dont l'interface est montrée dans la figure 5.13. L'excitation est procurée par les signaux musicaux émis sur haut-parleurs durant l'écoute. Une moyenne pondérée est exploitée afin de construire statistiquement au fil du temps la réponse de la salle sur toute la plage de fréquence audible. Dans le but d'en quantifier la performance, la réponse en fréquence obtenue est comparée à celle obtenue par une méthode standard servant de référence, soit par balayage sinusoïdal de la plage de fréquence.

1.3 Plan du mémoire

Ce mémoire est constitué de quatre chapitres principaux.

Le chapitre 2 contient différentes définitions nécessaires à la compréhension de cet ouvrage. Des concepts de base ainsi que des rappels en lien avec le sujet y sont présentés.

Dans le chapitre 3, les différents ouvrages sur l'utilisation de méthodes issues du domaine du traitement numérique des signaux appliqués aux problèmes de la modélisation et de la correction (égalisation) adaptative et non invasive de l'acoustique de salles sont répertoriés. L'objectif principal est de trouver si, à ce jour, des méthodes ont été suggérées pour la modélisation non invasive de la réponse en fréquence d'une salle dans le but d'en arriver à l'implémentation d'un filtre égaliseur adaptatif.

Le chapitre 4 entre dans le détail de la solution suggérée. On reviendra sur le schéma bloc de la figure 1.5 en décrivant l'implémentation de chacun des blocs qui nous intéressent, soit 1.6, 1.7 et 1.8.

On y aborde en premier lieu la méthode novatrice suggérée pour modéliser la salle sans l'utilisation de signaux d'excitation dédiés à la mesure. On étudie la dynamique d'un tel système et comment, au fil du temps, la réponse en fréquence et la réponse impulsionale de la salle sont reconstruites seulement qu'à partir d'un signal musical qui défile.



Figure 1.6 Bloc de modélisation de la chaîne audio

Ensuite, on voit comment à partir de la réponse impulsionale obtenue on peut obtenir la réponse impulsionale d'un filtre égalisateur agréable à l'oreille.



Figure 1.7 Bloc de conception du filtre égalisateur



Figure 1.8 Bloc du filtre convolutif

Dans le chapitre 5, on voit comment un filtre convolutif peut être implémenté en bénéficiant de l'efficacité de calcul obtenue en exploitant la transformée rapide de Fourier ainsi que des propriétés liant la convolution linéaire et la convolution circulaire. Le filtre implémenté agit comme filtre égalisateur en convoluant le fil de donnée audio à la réponse impulsionale conçue à partir de celle de la chaîne audio.

Le chapitre 6 présente l'application de la méthode de modélisation développée dans le cadre de ce travail à un problème acoustique, mais cette fois issue du domaine du génie des mines plutôt que de l'acoustique des salles. Ce travail mené en collaboration avec des chercheurs de l'université McGill a mené à deux publications scientifiques.

Finalement, au chapitre 7, le mémoire se termine par une brève conclusion.

CHAPITRE 2

DÉFINITIONS, CONCEPTS DE BASE ET RAPPELS

Le présent chapitre contient différentes définitions, des concepts de base ainsi que des rappels visant à faciliter la compréhension de cet ouvrage.

2.1 Définitions

On utilise dans cet ouvrage le terme « modélisation » ou « modèle » pour faire référence à l'estimation de la réponse impulsionnelle de la chaîne audio à l'étude. Comme on le verra, l'obtention de cette réponse impulsionnelle découle d'un calcul statistique dont la convergence dépend du contenu spectral de la musique jouée à travers la chaîne audio à ce moment. Il s'agit donc d'un algorithme de « modélisation » et non d'une équation déterministe menant directement à la réponse impulsionnelle. Par exemple, dans cet ouvrage, on utilise « algorithme de modélisation de la chaîne audio » plutôt que « calcul de la réponse impulsionnelle de la chaîne audio » lorsqu'on fait référence au calcul de cette réponse impulsionnelle.

On utilise également le terme « filtre adaptatif » pour refléter la capacité du filtre égalisateur à modifier sa propre réponse impulsionnelle en fonction de l'estimation de la réponse de la salle lui étant fournie périodiquement. La méthode suggérée s'inspire de la littérature disponible et de la recherche faite à ce jour dans le domaine de l'égalisation de la réponse de salle, non pas de la littérature sur les filtres adaptatifs classiques comme les filtres MLS, tels que décrits entre autres par Ferrara (1980).

Dans cet ouvrage on utilise le terme « référence » pour décrire un modèle obtenu à l'aide d'une méthode standard connue et dont les performances sont prises pour acquises. On décrit dans la section 4.4 comment un modèle de référence est obtenu expérimentalement, modèle qu'on utilisera pour évaluer la performance de la nouvelle méthode suggérée dans la section 4.5. Le modèle obtenu à l'aide de la méthode suggérée sera comparé au modèle de référence en utilisant le critère de performance décrit dans la section 5.1. On utilise la notation « ref » en indice suivant le nom d'une variable lorsque celle-ci appartient au modèle de référence, et aucun indice lorsque la variable appartient au modèle obtenu à partir de la technique suggérée. Par exemple, $h_{ref}[n]$ est la réponse impulsionnelle du modèle de référence obtenue à l'aide d'une méthode standard, alors que $h[n]$ est la réponse impulsionnelle du modèle obtenue à l'aide de la nouvelle méthode suggérée.

L'unité de mesure dBFS (de l'anglais dB « full scale ») est utilisé comme mesure de

l'amplitude de signaux numérique. Cette unité de mesure est logarithmique tel que le décibel, mais il est en plus normalisé par rapport à la plage dynamique supportée par le format de données numériques utilisé. Une onde carrée remplissant la pleine plage dynamique a une amplitude moyenne quadratique de 0 dBFS.

La revue de littérature du chapitre 3 nous apprend que la réponse impulsionale d'une salle se décompose en trois parties, soit le signal direct, les premières réflexions et la réverbération tardive. Pour des raisons évoquées dans la section 3.4, seulement deux de ces trois parties font typiquement l'objet d'égalisation, soit le signal direct et les premières réflexions. Pour alléger l'écriture, nous utiliserons les termes « signal direct » et « champ direct » lorsqu'on fait référence à la portion de la réponse impulsionale de la salle que l'on veut égaliser. Par « signal direct » et « champ direct », on entend la combinaison du signal direct et des premières réflexions, excluant la réverbération tardive.

On utilise dans cette ouvrage la notion de temps de réverbération. La norme (BSI) expose les normes régissant les mesures des paramètres acoustiques des salles. Elle contient la définition du temps de réverbération correspondant au temps nécessaire pour que la pression acoustique dans une salle diminue à -60 dB (environ un millième) de sa valeur initiale suite à l'interruption d'une source sonore.

Dans la section 4.1, une analogie est faite entre le problème étudié dans le cadre du présent travail et celui de la commande d'un moteur à courant continu. Le contrôle d'un moteur électrique est un problème classique dans le domaine de l'analyse des systèmes dynamiques, du contrôle et des filtres adaptatifs en général. L'analogie présentée entre ce problème et celui de l'égalisation adaptative d'une salle par voie numérique sert à la fois à illustrer les similitudes et les différences qui existent entre ces deux problèmes. Notamment, dans le cas d'un moteur électrique, la notion d'erreur est définie dans le domaine temporel où le moteur a une réponse naturelle donnée qu'on tente d'altérer par l'insertion d'un filtre ou d'un contrôleur dans la chaîne de commande. Dans le cas de l'égalisation de la réponse de la chaîne audio, nous verrons dans la section 5.1 que c'est plutôt dans le domaine fréquentiel que l'erreur entre la réponse observée et la réponse désirée est définie.

2.2 Concepts de base

Le domaine de l'acoustique étant relativement large, l'étude présentée se concentre sur certains éléments spécifiques étudiés par cette science plutôt que sur l'ensemble des lois qu'elle régit. Le système analysé doit donc être bien défini et aucun élément sortant des critères préétablis ne sera considéré.

Le système est monovariable. On s'intéresse à la relation entre une grandeur d'entrée parti-

culière, dite grandeur de commande $u(t)$, et une grandeur de sortie, dite grandeur commandée $y(t)$; $y(t)$ est aussi appelée réponse du système à $u(t)$. Le but du système monovariable est d'imposer à la sortie une loi de variation donnée de $y(t)$ en fonction de l'entrée $u(t)$. La figure 4.4 de la section 4.2 illustre ce système.

Le système est considéré comme étant linéaire (Stewart et Sandler, 2007). Les salles d'écoute musicale sont communément modélisées comme des systèmes linéaires où la réponse à l'impulsion de la salle $h(t)$ décrit les caractéristiques de transmission entre une source et une position d'écoute donnée (Sarris et Cambourakis, 2003). Un système physique est dit linéaire si les sorties sont liées aux entrées par des équations différentielles linéaires à coefficients constants. En acoustique linéaire, le comportement microscopique de l'air est abandonné au profit de son comportement macroscopique, permettant de considérer le milieu comme continu, de sorte que les outils du calcul différentiel peuvent y être appliqués. En considérant des conditions de température et de pression normales, la plage de fréquences audibles (20 Hz à 20 kHz) et une amplitude inférieure au seuil de la douleur (130 dB SPL), l'air libre se comporte comme un fluide non visqueux dont seulement les composantes hydrostatiques de la pression se propagent. L'équation d'état des gaz parfaits est donc valide et l'énergie interne du fluide ne dépend que de la température. L'équation aux dérivées partielles décrivant la propagation est obtenue à partir des principes de conservation de la masse et de conservation de la quantité de mouvement :

$$\nabla^2 p - \frac{1}{c^2} \cdot \frac{\delta^2 p}{\delta t^2} = 0 \quad (2.1)$$

où ∇^2 est l'opérateur laplacien.

Cette hypothèse de linéarité reste valide si on considère qu'on a à faire à un système thermodynamique isolé. En thermodynamique classique, on appelle système thermodynamique isolé, une portion de l'univers que l'on isole par la pensée du reste de l'univers que l'on baptise alors milieu extérieur. Ce système est dit isolé s'il n'y a aucun échange avec le milieu extérieur. Dans ce cas, la paroi est qualifiée d'adiabatique et doit être indéformable. L'hypothèse de linéarité est adoptée dans le contexte du présent travail.

Le système est invariant (Stewart et Sandler, 2007). Les salles d'écoute musicale sont aussi communément modélisées comme des systèmes invariants (Sarris et Cambourakis, 2003). Un système invariant par décalage temporel est un système dont la sortie ne dépend pas explicitement du temps. Si le signal d'entrée $u(t)$ produit une sortie $y(t)$, alors quelle que soit l'entrée décalée temporellement $u(t + \delta)$, la sortie est elle aussi décalée $y(t + \delta)$. Cette approximation est commune pour la modélisation de l'acoustique d'une salle par sa réponse impulsionale (Sarris et Cambourakis, 2003). Dans le cadre de ce travail, on verra que le système étudié peut varier dans le temps. Dans ce cas, le modèle pondéré de la salle $h(t)$ (ou

$h[n]$ dans le domaine discret) est périodiquement mis à jour. Cependant, pour la durée d'une seule analyse menant au calcul de chaque modèle individuel $h_i(t)$, les comportements de la salle, du haut-parleur et de l'amplificateur sont considérés comme invariants.

Le système est déterministe. Un système déterministe est un système dans lequel la transition entre l'état présent et l'état suivant ne fait intervenir aucun aspect aléatoire. Le système déterministe réagit toujours de la même façon lorsque la même entrée lui est fournie.

Puisque le calcul de réponse impulsionnelle est au centre de l'approche suggérée, la salle comme milieu de propagation est considérée comme un système thermodynamique isolé, monovariable, linéaire, invariant et déterministe. Cette hypothèse est commune à un grand nombre d'ouvrages répertoriés et est au centre de l'approche expérimentale adoptée.

2.3 Rappels

Les opérateurs de déconvolution linéaire \div_{lin} et de déconvolution circulaire \div_{circ} sont utilisés aux sections 4.4 et 4.5 du chapitre 4. Puisque ces deux opérateurs de déconvolution sont l'inverse de ceux de convolution, une bonne compréhension de ce qui différencie la convolution circulaire de celle linéaire s'impose pour bien comprendre par la suite les règles à suivre lorsqu'on désire exploiter la déconvolution circulaire pour approximer la déconvolution linéaire de l'équation 4.10.

Un opérateur très similaire à celui de la convolution linéaire permet la simplification du calcul de la convolution, notamment par le passage du domaine temporel vers le domaine fréquentiel en utilisant la transformée de Fourier discrète. L'opérateur en question est celui de la convolution circulaire notée $*_{circ}$.

$$y = u *_{circ} h = \mathcal{T}\mathcal{F}\mathcal{D}^{-1}((\mathcal{T}\mathcal{F}\mathcal{D}(u) \cdot \mathcal{T}\mathcal{F}\mathcal{D}(h))) \quad (2.2)$$

Notons qu'on utilise ici le symbole « étoile » $*$ accompagné de « lin » ou de « circ » pour noter la convolution entre deux signaux dans le domaine temporel. À ne pas confondre avec la multiplication complexe proprement dite qu'on utilise pour calculer cette convolution dans le domaine fréquentiel.

La convolution linéaire $*$ et la convolution circulaire $*_{circ}$ sont des opérateurs très similaires. Dans le domaine discret, le calcul de la convolution de deux séquences, une première de M éléments et une seconde de N éléments, résulte en une séquence $y_{lin}[n]$ de $M+N-1$ éléments. Le calcul de la convolution circulaire de ces deux mêmes séquences résulte en une

séquence $y_{circ}[n]$ de M éléments.

$$h[n] = \{1; 2; 1\}; N = 3$$

$$u[n] = \{0; 1; 2; 3; 4; 5; 6; 7\}; M = 8$$

$$\begin{aligned} y_{lin}[n] &= \{0; 1; 2; 3; 4; 5; 6; 7\} *_{lin} \{1; 2; 1\} \\ &= \{0; 1; 4; 8; 12; 16; 20; 24; 20; 7\} \end{aligned}$$

$$y_{circ}[n] = \{0; 1; 2; 3; 4; 5; 6; 7; \} *_{circ} \{1; 2; 1; 0; 0; 0; 0; 0; \}$$

$$= \{20; 8; 4; 8; 12; 16; 20; 24\}$$

Dans le cas de la convolution circulaire, des zéros doivent être ajoutés pour que les deux séquences à convoluer soient de mêmes longueurs (M). On dit de cette convolution circulaire qu'elle est de modulo M . En comparant le résultat des deux types de convolution, on note tout d'abord la différence entre la longueur des séquences résultantes obtenues, $M + N - 1$ (10) et M (8), pour le résultat obtenu par convolution linéaire et celui obtenu par convolution circulaire respectivement. En portant une attention toute particulière au résultat, on note que les éléments N à M (3 à 8) des deux séquences résultantes sont identiques.

On note aussi que les $N - 1$ (2) premiers et les $N - 1$ (2) derniers éléments de y_{lin} sont sommés pour former les $N - 1$ (2) premiers éléments de $y_{circ}[n]$.

$$y_{circ}[1] = y_{lin}[1] + y_{lin}[M + N - 2]$$

$$y_{circ}[2] = y_{lin}[2] + y_{lin}[M + N - 1]$$

L'unique différence entre la convolution linéaire $*$ et la convolution circulaire $*_{circ}$ de modulo M entre deux séquences des longueurs M et N est le repliement des $N-1$ premiers et derniers échantillons dans le cas de convolution circulaire. Si $h[n]$ est la réponse impulsionnelle d'une chaîne audio, les premiers échantillons de cette dernière contiendront la majorité de l'énergie. Les derniers échantillons formant la queue de cette réponse impulsionnelle auront une amplitude plus faible, voire quasi nulle. Lorsque c'est le cas, l'effet du repliement des $N-1$ premiers et derniers observés dans la convolution circulaire s'atténue.

$$h[n] = \{2; 0,1; 0\}N = 3$$

$$u[n] = \{0; 1; 2; 3; 4; 5; 6; 7\}; M = 8$$

$$\begin{aligned} y_{lin}[n] &= \{0; 1; 2; 3; 4; 5; 6; 7\} *_{lin} \{2; 0,1; 0\} \\ &= \{0; 2; 4,1; 6,2; 8,3; 10,4; 12,5; 14,6; 0,7; 0\} \end{aligned}$$

$$\begin{aligned} y_{circ}[n] &= \{0; 1; 2; 3; 4; 5; 6; 7\} *_{circ} \{2; 0,1; 0; 0; 0; 0; 0; 0\} \\ &= \{0,7; 2; 4,1; 6,2; 8,3; 10,4; 12,5; 14,6\} \end{aligned}$$

On note encore une fois dans le cas de la convolution circulaire de modulo M que des zéros ont été ajoutés à la séquence $h[n]$ pour en faire une séquence de longueur M . En raison du peu d'énergie contenue dans la queue de $h[n]$, on note que les derniers éléments de la séquence $y_{lin}[n]$ résultante de la convolution linéaire sont quasi nuls. Résultat, l'erreur faite sur les $N - 1$ (2) premiers éléments en approximant la convolution circulaire comme une approximation de la convolution linéaire due au repliement des $N - 1$ derniers éléments reste minimale.

$$y_{lin}[n] \approx y_{circ}[n] \quad (2.3)$$

Dans le cas où très peu d'énergie se trouve dans les derniers $N - 1$ derniers échantillons de la réponse impulsionale $h[n]$, l'approximation de la convolution linéaire par $*$ par l'opérateur de convolution circulaire $*_{circ}$ est acceptable.

Le prochain chapitre répertorie quelque principaux travaux en acoustique des salles. On y fait également l'état de l'art en resséchant les travaux dans lesquels on s'est intéressé aux sujets abordés dans le présent projet de recherche.

CHAPITRE 3

REVUE DE LITTÉRATURE

Le but de la revue de littérature est de répertorier les divers travaux qui, à travers le temps, ont mené vers l'étude de l'acoustique des salles et des imperfections dans les chaînes audionumériques dans le but d'en faire la correction. Divers travaux furent répertoriés, allant des débuts de l'acoustique architecturale et de l'analyse en fréquence jusqu'aux récentes tentatives de modéliser expérimentalement l'acoustique d'une salle à l'insu de l'auditoire ainsi que différentes techniques suggérées pour l'égalisation par filtrage numérique.

3.1 Historique de l'acoustique comme science

Les premières constatations du phénomène physique de la propagation d'ondes proviennent d'observations de la surface de l'eau lorsque celle-ci est perturbée. À l'origine, l'onde se définit comme étant la propagation d'une perturbation produisant sur son passage une variation réversible des propriétés physiques locales et transportant une certaine quantité d'énergie sans déplacer avec elle la matière à travers laquelle elle se propage. Cette analogie au comportement des perturbations présentes à la surface de l'eau a été évoquée entre autre par le philosophe grec Chrysippe (240 av. J.-C.).

La science de l'acoustique a avancé, entre autres, grâce aux travaux de Galilée (1564-1642), Isaac Beeckman (1588-1637), René Descartes (1596-1650) et P. Marin Mersenne (1588-1648) qui incorporaient pour la première fois le concept de fréquence et de vibration dans l'étude du son. Une expérience valable réalisée par Mersenne dans Harmonie Universelle prouve qu'un timbre sonore est composé de fréquences multiples d'une fréquence nommée fondamentale.

Ce n'est qu'en 1686 qu'une théorie mathématique de la propagation du son est suggérée par Isaac Newton. Il avance une explication mécanique à la propagation en apportant l'explication voulant que le son soit transporté par la variation de la pression du fluide dans lequel il se propage.

3.2 Le domaine fréquentiel

Pour bien des acousticiens des temps modernes, une des arrivées les plus importantes du domaine fut celle apportée par Joseph Fourier (1768-1830) en 1822 grâce à la décomposition en série trigonométrique permettant de décomposer un son en une somme de composantes fréquentielles. Cette représentation est tout particulièrement intéressante compte tenu du fait

que l'oreille humaine, par sa physiologie, est sensible aux fréquences qui composent un son plutôt qu'à la variation temporelle de la pression de l'air entourant le tympan. La majorité des divers phénomènes observés en lien avec l'onde acoustique et le milieu dans lequel elle se propage sont caractérisés par leur comportement fréquentiel. C'est d'ailleurs le sujet de ce présent ouvrage où le comportement fréquentiel de la salle d'écoute musicale est à l'étude.

Plus récemment, des auteurs tels que Sreenivas *et al.* (1990) s'intéressent au comportement fréquentiel de l'oreille humaine et des outils mathématiques permettant de bien modéliser un signal ou un système en fréquence. On s'intéresse au comportement autant physiologique que psychologique de l'oreille humaine et de la perception du son par l'humain et on tente d'établir un lien entre ce comportement et ceux de différents outils mathématiques permettant d'analyser des signaux en fréquence, notamment des signaux de parole. On y modélise le comportement en fréquence de l'oreille humaine par la mise en cascade de filtres RIF et RII et on utilise la transformée de Fourier discrète pour calculer le spectre d'amplitude de différents signaux de voix passant à travers ce modèle. Ce genre de travaux laisse entrevoir qu'un tel modèle basé sur un filtre à réponse impulsionnelle finie permettrait de modéliser d'autres composants d'une chaîne audio, par exemple un haut-parleur ou une salle dans laquelle a lieu l'écoute, dans une approche qui d'un point de vue mathématique, concorde avec le fonctionnement de l'oreille humaine. Dans cet article, le concept de résolution spectrale de l'oreille est également introduit. L'auteur étudie la capacité de l'humain ainsi que celle de son modèle à différencier deux sinusoïdes distinctes lorsque les fréquences de ces dernières sont très proches. Depuis, c'est devenu la norme d'approcher le problème de la modélisation de l'acoustique d'une salle dans le domaine fréquentiel puisque ce genre de modèle est hautement corrélé au fonctionnement de l'oreille humaine.

Au-delà de la simple obtention de la réponse en fréquence d'une salle ou de la pleine chaîne audio obtenue en appliquant une transformée de Fourier à la réponse impulsionnelle, certains auteurs tels que Sarris et Cambourakis (2003) commencent très vite à s'intéresser à l'analyse temps-fréquence. On suggère notamment que plusieurs transformées de Fourier peuvent être appliquées à la réponse impulsionnelle de façon à décomposer par bandes de fréquences cette réponse impulsionnelle, ce qui permet d'appliquer un traitement différent pour différentes portions du spectre. La décomposition du spectre par bandes de fréquences notamment en octave est une pratique courante en audio et dans d'autres domaines. La littérature sur le sujet est riche. Dans son travail, Nelson *et al.* (1972) explique comment la méthode des résidus peut être utilisée pour la décomposition par bandes de fréquences et que les filtres utilisés doivent être choisis de façon à bien atténuer les lobes secondaires correspondant au chevauchement entre les bandes de fréquences et réduire au maximum l'effet de Gibbs.

3.3 Modélisation expérimentale

Schroeder (1979) et Farina (2000) décrivent et présentent des résultats expérimentaux pour deux méthodes standards permettant de modéliser le comportement d'une salle expérimentalement. La réponse de la salle à des signaux d'excitation émis sur haut-parleur est capturée à l'aide d'un microphone placé à la position d'écoute. Les signaux d'excitation utilisés sont des signaux dédiés à la mesure. Schroeder (1979) utilise un bruit blanc pseudo-aléatoire généré à l'aide de registres à décalage. Ce signal est connu sous le nom de MLS pour « Maximum Length Sequence », ou « séquence de longueur maximale » dans sa traduction française. Farina (2000) utilise une sinusoïde à fréquence variable comme signal d'excitation. Il fait varier la fréquence de cette sinusoïde de manière à couvrir l'ensemble de la plage de fréquence. Ce signal est connu sous le nom de « sine sweep » ou « balayage sinusoïdal ». Le terme français est utilisé dans le présent travail. Le signal capté est comparé à celui émis dans le but d'extraire la réponse impulsionale et la réponse en fréquence de la pleine chaîne audio (amplificateur, haut-parleur-salle). Une FFT est utilisée dans les deux cas pour assurer le passage dans le domaine de Fourier où sont effectués les calculs permettant d'extraire la réponse de la salle. Farina (2000) utilise la méthode de la division dans le domaine fréquentiel comme moyen efficace de calculer la déconvolution circulaire. Farina (2000) considère la déconvolution circulaire comme une bonne approximation de la déconvolution linéaire.

3.4 Réponse impulsionale et réponse en fréquence

Stewart et Sandler (2007) nous apprennent que la réponse impulsionale d'une salle est composée temporellement de trois différentes parties. On ne s'intéresse pas aux chiffres exactes en miliseconde ici puisque ceux-ci varient selon le volume de la salle. La première partie, constituée des premières millisecondes de la réponse impulsionale, est le signal direct. Le son direct correspond au premier front d'onde arrivant à la position d'écoute lorsqu'une impulsion est émise à travers le haut-parleur. Ce son a voyagé en ligne directe et n'a pas encore été réfléchi sur les différentes surfaces de la salle lorsqu'il arrive à la position d'écoute. La deuxième partie de la réponse impulsionale est constituée de premières réflexions. Reprenons l'exemple d'une impulsion transmise à un haut-parleur. Les premières réflexions arrivent à la position d'écoute après que l'impulsion émise sur le haut-parleur ait été réfléchie sur le plafond, le plancher, les murs latéraux, ainsi que le mur avant et arrière. Dans une réponse impulsionale, l'amplitude des différentes réflexions à leur arrivée à la position d'écoute décroît au fil du temps jusqu'à atteindre une amplitude suffisamment faible ne permettant pas à l'oreille de les différencier individuellement les unes des autres. Ce passage marque le début de la troisième partie de la réponse impulsionale, soit la réverbération tardive. Cette

troisième partie, constituée des dernières millisecondes de la réponse impulsionale, ne fait typiquement pas l'objet d'égalisation. L'auteur présente une équation permettant de calculer analytiquement le temps marquant, à l'intérieur d'une réponse impulsionale de salle, la fin des premières réflexions et le début de la réverbération tardive. Cette équation est basée sur le nombre de réflexions, le volume de la salle, la vitesse du son, et le délai entre l'émission d'une impulsion et son arrivée à la position d'écoute.

3.5 L'égalisation

Plomp (1964) explique que la résolution temporelle et spectrale de l'oreille humaine varie en fonction de la fréquence. Cette propriété est utilisée dans le cadre de ce travail pour la décomposition par bandes de fréquence de la réponse impulsionale de la salle menant à la conception du filtre égalisateur. Une approche multirésolution (temps-fréquence) est utilisée.

Miyoshi et Kaneda (1988) suggèrent une méthode d'inversion de la réponse impulsionale de la salle pour la conception de filtre égalisateur dans le cas d'un système multicanal. Cette technique peut par exemple être utilisée pour l'égalisation dans le cas où plusieurs haut-parleurs ou plusieurs points d'écoute sont considérés.

Hatziantoniou et Mourjopoulos (2003) et Hatziantoniou et Mourjopoulos (2004) nous présentent des méthodes expérimentales pour la modélisation de l'acoustique d'une salle. Les auteurs sont intéressés par l'obtention de la réponse impulsionale ainsi que la réponse en fréquence de la salle. Hatziantoniou et Mourjopoulos (2003), dans leur travail, utilisent un signal d'excitation dédié à la mesure. Il s'agit d'un bruit blanc pseudo-aléatoire généré à l'aide de registre à décalages. Hatziantoniou et Mourjopoulos (2004) utilisent plutôt le balayage sinusoïdal de la plage de fréquence. L'exercice de mesure a lieu et le modèle obtenu est ensuite utilisé pour concevoir un filtre égalisateur statique servant à corriger l'effet de la salle sur un signal qui s'y propage. L'exercice de mesure n'a pas lieu durant l'écoute musicale puisqu'il requiert l'émission sur haut-parleur de ce signal dédié à la mesure. Dans l'article répertorié, Hatziantoniou et Mourjopoulos (2003) nous expliquent que la simple inversion de la pleine réponse impulsionale de la salle dans le but d'obtenir la réponse impulsionale du filtre égalisateur n'est pas convenable comme approche. Le résultat est l'injection d'énergie par le filtre avant l'impulsion principale. Ceci a pour effet d'ajouter des « pré-échos » dans le signal lors de l'écoute. Par exemple, si on écoute la portion d'un enregistrement de percussion où un tambour est frappé, on entendra une partie du son de tambour avant qu'il ne soit frappé. Hatziantoniou et Mourjopoulos (2003) et Hatziantoniou et Mourjopoulos (2004) suggèrent tous deux l'utilisation d'une moyenne complexe pour réduire la longueur de la réponse de la salle avant son inversion menant à la réponse impulsionale du filtre égalisateur.

Les travaux de Mosayyebpour *et al.* (2013) indiquent que c'est la capacité du cerveau humain à différencier le champ direct et les premières réflexions du champ réverbéré qui devrait dicter la portion de la réponse impulsionale de la salle qu'on doit considérer lorsqu'il est question d'égalisation. Selon Mosayyebpour *et al.* (2013), le cerveau perçoit les 50 à 100 premières millisecondes de la réponse impulsionale d'une salle dite petite comme faisant partie du champ direct, le reste comme faisant partie du champ réverbéré. Ils expliquent aussi pourquoi seul le direct doit faire l'objet d'égalisation.

Pour contrer les problèmes causés par l'inversion de la pleine réponse en fréquence, Rubak et Johansen (2001) suggèrent plutôt une approche temps-fréquence, où la réponse de la salle est tout d'abord décomposée par bandes de fréquences avant d'être inversée. Une fois décomposée par bande de fréquences, on manipule la réponse de la salle pour cette bande. Par exemple, la longueur de la réponse du filtre, ainsi que la distribution temporelle de l'énergie dans celle-ci, peut être ajustée, et cet ajustement ne concerne que la bande dont il est question. Ensuite les différentes bandes de fréquences sont recombinées et la réponse obtenue est inversée et utilisée comme filtre égalisateur. Une telle méthode offre beaucoup de flexibilité et permet de contrer les problèmes de « pré-écho » présent lorsque la pleine réponse impulsionale est inversée. Trois bandes sont utilisées dans leurs travaux.

Zhang *et al.* (2008) montrent des résultats de test où ils font varier la longueur d'un filtre égalisateur utilisé pour corriger le champ direct ainsi que les premières réflexions de la réponse impulsionale d'une salle. Des résultats sont montrés pour des filtres numériques de 2048, 4096 et 8192 échantillons. Les filtres sont testés sur quatre différentes salles ayant des temps de réverbération de 200 ms, 400 ms, 1000 ms, et 1200 ms. Bien que le filtre plus court ne s'avère qu'à fonctionner que pour les salles avec de courts temps de réverbération typiquement de 500 ms ou moins, ils ne notent aucune perte de performance en utilisant un filtre plus long même sur une salle ayant un temps de réverbération court. Seul le filtre de 8192 échantillons mène à l'obtention de résultats satisfaisants pour l'ensemble des quatre salles.

Un filtre agréable à l'oreille assurant l'égalisation de la réponse de la salle est une chose, mais le simple fait que le filtre soit agréable à l'oreille et ne souffre pas du problème des « pré-échos » n'est pas suffisant pour valider que le filtre en question améliore bel et bien la réponse de la salle. Canclini *et al.* (2012) suggère l'utilisation d'un critère de performance, l'erreur quadratique moyenne, pour comparer la réponse en fréquence obtenue à la réponse en fréquence désirée. En mode modélisation, on utilisera dans le présent travail ce même critère pour mesurer la capacité de l'algorithme suggéré à estimer la réponse en fréquence de la salle à partir de signaux d'excitation musicaux. La réponse en fréquence désirée dans ce cas est celle obtenue par une méthode standard utilisant des signaux d'excitation dédiés

à la mesure. Nous utilisons le même critère en mode égalisation. Puisque dans notre cas, la réponse désirée est une réponse parfaitement plate, on utilise aussi l'erreur quadratique moyenne pour comparer la réponse en fréquence obtenue, après égalisation, à une réponse en fréquence parfaitement plate, dans le but de mesurer la performance de la méthode suggérée.

3.6 Égalisation adaptative à l'insu de l'auditoire

Nowlin *et al.* (2000) suggère une approche non invasive pour la modélisation de la réponse de la salle dans le but d'en faire la correction. L'algorithme de modélisation suggéré est inspiré d'une technique standard, le filtered-X LMS (FXLMS), utilisé dans divers filtres adaptatifs multientrées et multisorties. L'algorithme reconstruit au fil du temps, en comparant l'entrée du système à sa sortie, la réponse en fréquence de la salle. Les résultats montrent la capacité de l'algorithme à converger en quelques secondes seulement. Cependant, des résultats ne sont présentés que pour la plage de 50 à 1000 Hz et ne permettent pas de conclure que cette technique puisse être utilisée pour corriger la pleine réponse en fréquence (jusqu'à 20 kHz). La méthode de modélisation non invasive présentée n'exploite pas certaines propriétés intéressantes des signaux musicaux. Par exemple la propriété voulant que l'excitation procurée par les signaux musicaux finisse, au fil du temps, par couvrir l'ensemble du spectre. La méthode de modélisation suggérée ne reconstruit pas la réponse pour l'ensemble du spectre au fil du temps et se limite plutôt à la bande entre 50 à 1000 Hz en faisant l'hypothèse que celle-ci est excitée continuellement.

Rieta et Sanchis (2006) et Rocha *et al.* (2006) suggèrent l'utilisation des propriétés psychoacoustiques pour déterminer comment certains signaux d'excitation permettant de récupérer la réponse en fréquence d'un milieu de propagation peuvent être cachés à l'intérieur du signal musical de façon à ne pas être perçus par l'auditeur. L'algorithme suggéré consiste à insérer des tonalités à l'intérieur d'une pièce musicale tout en s'assurant que celles-ci soient imperceptibles pour l'oreille humaine. L'amplitude de ces tonalités à l'émission est comparée à celle mesurée par un microphone pour déduire l'atténuation de la salle pour ces tonalités. Cette méthode est cependant limitée pour ce qui est de la résolution spectrale au nombre de tonalités pouvant être ajoutées au signal sans que celles-ci ne deviennent perceptibles. De plus, la réponse en fréquence obtenue ne permet pas de déduire la phase du signal perçu par l'auditeur en fonction de la fréquence. Le filtre égalisateur implanté permet donc de ne corriger que l'atténuation et laisse malheureusement de côté la phase. Un temps de convergence important serait nécessaire si on voulait atteindre une résolution spectrale suffisante pour l'égalisation, par exemple quelques milliers de bandes de fréquence.

Le travail de Kendrick *et al.* (2006) tente d'utiliser le signal musical tel quel pour dé-

terminer certains paramètres acoustiques de la salle. Cependant, les paramètres calculés ne permettent pas la correction de la réponse en fréquence à l'aide de filtre égalisateur. Les techniques suggérées permettent entre autres de calculer le temps de réverbération, mais ce paramètre à lui seul ne permet pas de connaître la pleine réponse impulsionale ou même simplement l'atténuation pour toutes fréquences imposées à l'onde acoustique par son milieu de propagation. Cox *et al.* (2001) suggère l'utilisation de réseaux de neurones pour l'extraction de paramètres similaires qui à eux seuls ne suffisent pas pour la conception d'un filtre égalisateur.

Des chercheurs tels que Mosayyebpour *et al.* (2012) travaillent sur le problème inverse. Ils désirent filtrer le signal capté par un microphone dans l'environnement d'écoute plutôt que de filtrer un signal avant qu'il ne soit émis sur un haut-parleur. La motivation qui lie entre eux ces chercheurs est souvent en lien avec la déréverbération ou le débruitage du signal audio capté par des dispositifs ayant des capacités de reconnaissance vocale ou utilisés pour la transmission de la voix. Dans un système de vidéoconférence, par exemple, un microphone placé dans l'environnement d'écoute retransmet la conversation ayant lieu dans cette salle à des participants distants via Internet. Le signal capté, par exemple la voix, par le microphone souffre de déformation induite par l'environnement dans lequel le son se propage : la salle. Dans ce cas, il est impossible de corriger le son avant son émission puisqu'il s'agit de voix humaine et non d'un signal émis sur haut-parleur. Les approches utilisées dans ce genre de problème diffèrent beaucoup de celles utilisées dans le problème du filtre égalisateur adaptatif auquel on s'intéresse dans ce travail. Mosayyebpour *et al.* (2012) font une bonne revue de l'état de l'art dans ce domaine et comparent la performance de deux méthodes basées sur le « LP residual Skewnesss » et le « LP residual Kurtosis ».

L'état de l'art nous indique qu'il reste beaucoup de place aux nouvelles idées pour l'égalisation adaptative de la réponse de chaîne audionumérique. Dans le présent travail, une méthode novatrice est suggérée, implémentée et validée expérimentalement.

CHAPITRE 4

DÉTAILS DE LA SOLUTION

Dans ce chapitre est abordé le sujet de la modélisation d'une chaîne audionumérique. L'objectif est de développer une méthode innovatrice permettant de capturer la réponse impulsionale d'une salle à l'insu de l'auditoire. Pour ce faire, aucun signal dédié à la mesure ne doit être utilisé. La méthode développée permet la capture de la réponse impulsionale en n'exploitant que les signaux musicaux. La réponse impulsionale obtenue sert à la mise à jour périodique des coefficients d'un filtre égalisateur adaptatif, capable de s'adapter dynamiquement au comportement de la salle pouvant évoluer au fil du temps. Le rôle de ce filtre convolutif numérique est de réduire l'impact des différents composants d'une chaîne audio typiquement utilisée pour l'écoute musicale sur haut-parleur.

On veut notamment diminuer l'effet de la salle dans laquelle a lieu l'écoute sur l'onde acoustique qui s'y propage. En s'interrogeant sur les différences qui existent entre l'écoute musicale sur haut-parleur et l'utilisation du casque d'écoute, on comprend rapidement que le comportement acoustique de la salle dans laquelle a lieu l'écoute joue un rôle important dans le cas de l'écoute sur haut-parleur. Si on ignore les imperfections des haut-parleurs dans chacun des cas, le son est altéré par la salle entre le haut-parleur et le point d'écoute, alors qu'avec le casque d'écoute, le son perçu est tel qu'il figure sur l'enregistrement.

La conception d'un filtre égalisateur passe forcément par la modélisation de la salle devant être compensée. Les premières sections de ce chapitre, soit les sections 4.1 à 4.5, portent donc sur la technique de modélisation innovatrice suggérée.

Les sections 4.6 et 4.7 font un parallèle et illustrent les différences entre l'approche utilisée en acoustique des salles où il est question d'égalisation de la réponse en fréquence d'une salle comparativement à d'autres problèmes plus classiques tels que celui de la commande d'un moteur électrique où l'on s'intéresse plutôt au comportement temporel, comme par exemple la réponse à l'échelon. Dans chacun de ces deux cas, la notion d'erreur statique est définie.

Finalement, les sections 4.8 à 4.9 abordent le sujet du filtre égalisateur plus en détail, de l'obtention de sa réponse impulsionale découlant notamment de celle de la salle, jusqu'à son insertion dans la chaîne audio sous forme de filtre adaptatif tel qu'illustré dans la figure 1.5 du chapitre 1.

4.1 Analogie : modélisation d'un moteur

Le moteur électrique pour lequel on veut observer la vitesse de rotation en est un exemple classique. On alimente un moteur avec une certaine tension électrique $u(t)$ pouvant varier au fil du temps dans le but de le faire tourner à une vitesse de référence donnée, $r(t)$. Un tachymètre, dispositif fournissant une tension électrique proportionnelle à une vitesse de rotation, peut être rattaché à l'axe du moteur et utilisé pour mesurer la vitesse du rotor $y(t)$. Pour des fins de simplification, le gain en boucle ouverte du système peut être considéré comme unitaire, c'est le cas par exemple si le tachymètre est conçu pour générer 1 volt lorsque le moteur tourne en étant alimenté à 1 volt. Dans ce cas, la vitesse de rotation désirée $r(t)$ ainsi que celle mesurée $y(t)$ peuvent être exprimées directement en volt. Aussi, la vitesse désirée $r(t)$ ainsi que la commande $u(t)$ sont égales. En l'absence d'un contrôleur ou d'un régulateur dans la chaîne d'alimentation du moteur, on injecte directement au moteur une tension $u(t)$, correspondant à la vitesse désirée. L'erreur statique est observée dans le domaine du temps, soit la différence entre la vitesse du moteur et la vitesse de référence, et est dénotée $e(t)$.

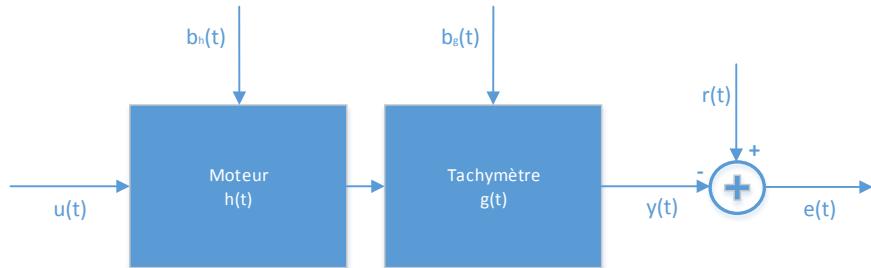


Figure 4.1 Moteur électrique et tachymètre

Des perturbations, notées $W_g(t)$ et $W_h(t)$, de nature souvent stochastique, telles que du bruit, peuvent venir altérer le comportement du système ou imputer notre capacité à bien l'observer. Le tachymètre à gain unitaire, ici représenté dans le domaine temporel par sa réponse impulsionale $g(t)$, est un instrument de mesure de haute qualité et est considéré idéal dans ce problème, c'est-à-dire qu'il fournit en sortie une tension égale à la vitesse de rotation du moteur en tout temps. On omet normalement de le représenter dans le schéma bloc.

Reste le moteur, ici représenté dans le domaine temporel par sa réponse impulsionale $h(t)$.

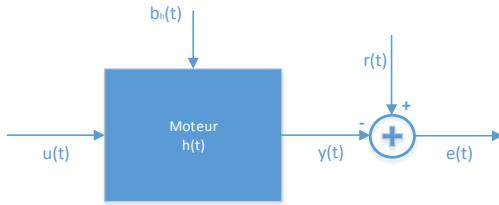


Figure 4.2 Moteur à courant continu représenté dans le domaine temporel par sa réponse impulsionnelle $h(t)$

4.2 La chaîne audio à modéliser

On pourrait directement transposer le problème du moteur électrique à celui de l'écoute musicale sur haut-parleur dans une salle. Le moteur est remplacé par une chaîne audio, notamment un amplificateur, un haut-parleur et la salle dans laquelle a lieu l'écoute. L'ensemble du système est excité par une tension électrique en entrée notée $u(t)$. Un microphone amplifié et placé dans la salle à la position d'écoute sert à observer la pression acoustique, remplaçant le tachymètre qui servait précédemment à observer la vitesse du moteur. Encore une fois, pour des fins de simplification, le gain en boucle ouverte du système peut être considéré unitaire, c'est le cas par exemple si un microphone placé dans la salle est conçu pour générer une sinusoïde d'un volt lorsque le haut-parleur est alimenté avec une sinusoïde d'un volt. Dans ce cas, le signal désiré, à la position d'écoute, peut être exprimé directement en volts, et $u(t) = r(t)$.

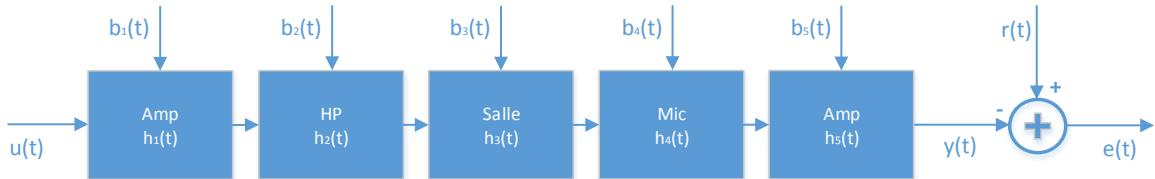


Figure 4.3 Chaîne audio et microphone de mesure

Comme c'est le cas pour le tachymètre, le microphone et son amplificateur, ici représentés dans le domaine temporel par leurs réponses impulsionnelles $h_4(t)$ et $h_5(t)$, sont des instruments de mesure de haute qualité et sont considérés idéaux dans ce problème, c'est-à-dire qu'ils fournissent en sortie une tension égale à la pression acoustique à la position d'écoute en tout temps. On omet normalement de les représenter dans le schéma bloc.

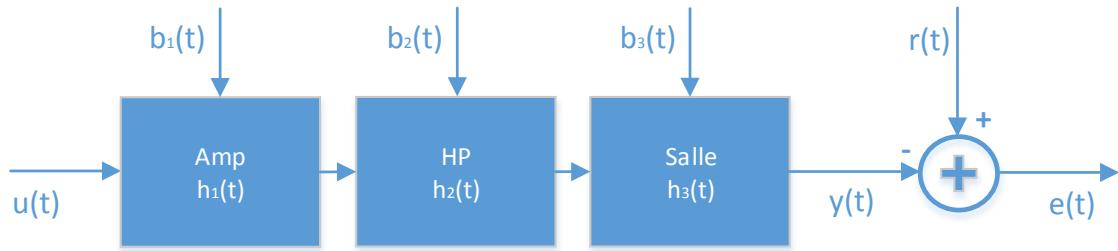


Figure 4.4 Chaine audio

Dans ce travail, on s'intéresse autant aux transformations induites sur le signal par le haut-parleur que celles induites par la salle. Le haut-parleur ainsi que son amplificateur sont vus comme des systèmes fermés et toute intervention dans le but d'améliorer la performance du système doit prendre place dans un contrôleur placé à l'entrée de l'amplificateur. Pour cette raison, l'ensemble amplificateur/haut-parleur/salle est considéré comme un seul système. Ici, $p(t)$ représente la somme des perturbations sur toute la chaîne audio, de la pollution électrique dans l'amplificateur du haut-parleur et celui du microphone, jusqu'au bruit ambiant dans la salle. La notation salle est simplement utilisée, et $h(n)$ est utilisé lorsqu'on fait référence à la réponse impulsionnelle du système complet. La nature de ce système complet est la principale cause de l'écart entre la commande $u(t)$ et la sortie observée $y(t)$.

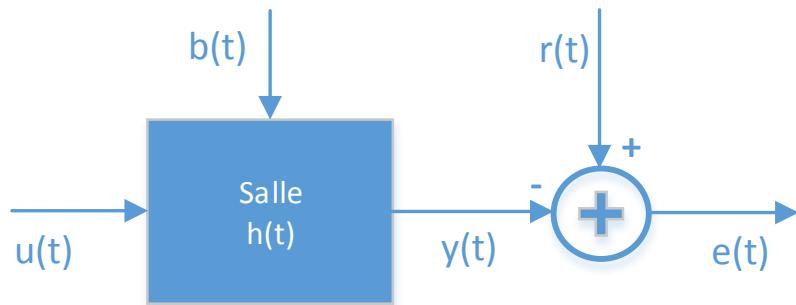


Figure 4.5 Système dans le domaine

Puisque la chaîne audio étudiée dans le cadre de ce travail est numérique, toute intervention dans le but d'améliorer la performance du système doit non seulement prendre place

dans un contrôleur placé à l'entrée de la chaîne audio mais ce contrôleur doit être numérique. Ceci implique que les signaux traités ont été préalablement échantillonnés et sont discrets plutôt que continus.

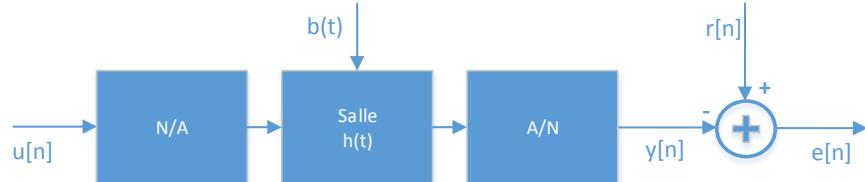


Figure 4.6 Passage du domaine continu au domaine discret

Les convertisseurs de la figure 4.6 sont eux aussi des instruments de mesure de haute qualité et sont considérés idéaux, c'est-à-dire qu'on considère qu'ils n'introduisent ni bruit, ni distorsion, ni transformations dans les signaux qui les traversent. On omet donc de les représenter explicitement dans le schéma bloc, nous traitons directement du système à temps continu $h(t)$ comme s'il s'agissait d'un système numérique, où $h[n]$ est la réponse impulsionnelle à temps discret tel que le montre la figure 4.7.

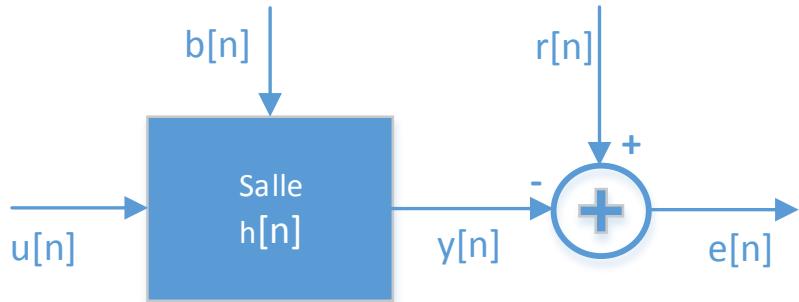


Figure 4.7 Système discret

Les sections 4.3, 4.4 et 4.5 portent sur la modélisation de cette réponse impulsionnelle $h[n]$.

4.3 La réponse impulsionale

La modélisation par excitation impulsionale est la première méthode de modélisation entrée-sortie considérée par cet exercice. Elle consiste à exciter le système à l'étude à l'aide d'un signal très puissant, mais de très courte durée, appelée l'impulsion de Dirac.

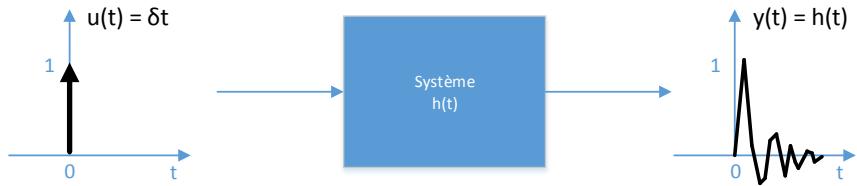


Figure 4.8 Réponse d'un système à l'impulsion

La sortie obtenue représente explicitement le comportement temporel du système et peut être utilisée pour calculer la réponse en fréquence du système.

La raison expliquant le grand intérêt envers l'impulsion comme signal d'excitation provient de la nature de ce signal. Tout d'abord, pour déterminer le contenu fréquentiel d'un signal quelconque, on utilise généralement la transformée de Fourier d'une fonction $f(t)$, définie comme suit :

$$\mathcal{F}(f) = F(j\omega) \Big|_{\omega=2\pi f} = \int_{-\infty}^{+\infty} f(t) \cdot e^{-j\omega t} dt \quad (4.1)$$

Pour bien comprendre l'intérêt de l'impulsion $\delta(t)$, le passage par la transformée de Fourier s'impose :

$$\mathcal{F}(\delta(t)) = \int_{-\infty}^{+\infty} \delta(t) \cdot e^{-j\omega t} dt = 1 \quad (4.2)$$

On remarque que le résultat de la transformée est égal à 1, ce qui indique que la représentation fréquentielle de l'impulsion est une constante. Ce résultat s'interprète de la manière suivante : puisque la composition spectrale de l'impulsion de Dirac est une constante, cette dernière se compose donc de chaque fréquence de la plage de fréquence $f \in (-\infty, \infty)$ où chacune de ces fréquences est présente avec une phase et une amplitude égales.

Dans une tentative de modéliser la réponse d'un système à l'aide d'un certain signal d'excitation, il est convenu que le modèle obtenu ne peut que représenter le comportement du système par rapport aux fréquences contenues dans le signal d'excitation exploité. Par exemple, un modèle obtenu à l'aide d'un signal d'excitation dont le spectre s'étend sur l'inter-

valle fermé $f \in [f_a, f_b]$ ne permet de déterminer le comportement du système à une fréquence d'excitation f_c que si elle appartient à l'intervalle $f_c \in [f_a, f_b]$. Ceci est dû au fait qu'une multiplication par zéro dans le domaine de Fourier ne permet plus de savoir quelle valeur a été multipliée. Le fait d'exploiter l'impulsion de Dirac en tant que signal d'excitation servant à la modélisation d'un système est en quelque sorte une garantie que le modèle obtenu caractérisera le comportement de ce dernier pour toutes fréquences.

Dans leurs tentatives de mesurer la réponse impulsionale acoustique de divers systèmes, les acousticiens ont développé différentes méthodes afin de générer une impulsion de Dirac acoustique la plus pure et la plus puissante possible. La première méthode est relativement triviale. Comme pour la génération d'un signal sonore de façon générale, le convertisseur numérique-analogique ainsi que le haut-parleur peuvent être utilisés pour générer une impulsion. Trois échantillons numériques suffisent pour construire une quasi-impulsion. Le signal est par la suite converti en signal analogique et amplifié pour être émis à l'aide d'un haut-parleur.

Le désavantage de la génération d'une impulsion sur haut-parleur est l'importance du bruit comparativement à la puissance du signal émis. Le bruit ambiant, ainsi que celui introduit par les appareils de mesure, est d'une puissance constante. Dans le but de conserver un rapport signal sur bruit (RSB) suffisant, la puissance transportée par l'impulsion émise doit être nettement supérieure à la puissance du bruit présent sur la mesure. La nature du haut-parleur ne permet cependant pas d'émettre une quasi-impulsion avec une puissance suffisamment grande pour que cette dernière puisse être réutilisée dans n'importe quelle circonstance. Théoriquement, l'amplitude d'une impulsion devrait être infinie, et la durée devrait tendre vers zéro. Ce n'est ni le cas pour une impulsion émise sur un haut-parleur ni pour celle émise par des méthodes alternatives telles que l'éclatement d'un ballon ou la détonation d'un fusil à blanc. Les techniques suggérant d'utiliser l'éclatement d'un ballon ou la détonation d'un fusil à blanc pour simuler l'émission d'une impulsion souffrent de l'impureté de la quasi-impulsion qu'elles génèrent. Pour calculer la réponse à l'impulsion d'un système à partir d'une telle excitation, il est nécessaire de déconvoluer le signal d'excitation de la réponse à cette excitation. Cependant, dans le cas de l'éclatement d'un ballon ou la détonation d'un fusil à blanc, le signal d'excitation n'est pas connu et peut varier de fois en fois.

L'impossibilité de générer une impulsion pure nous amène à nous pencher sur des techniques autres permettant d'obtenir expérimentalement la réponse impulsionale d'une salle ou d'une pleine chaîne audio (amplificateur/haut-parleur/salle). La section 4.4 couvre une des méthodes standards remédiant à ce problème. La section 4.5 couvre la méthode suggérée dans ce travail permettant à l'exercice de mesure d'avoir lieu à l'insu de l'auditoire durant l'écoute musicale.

4.4 Technique de modélisation de référence

La méthode vue dans cette section est la première de deux méthodes plus évoluées permettant d'obtenir la réponse impulsionnelle d'une salle. Cette méthode veut ajouter à la méthode précédente en augmentant considérablement la puissance du signal d'excitation ainsi qu'en facilitant son émission sur haut-parleur. Le signal généré, un signal sinusoïdal dont la fréquence évolue suivant une fonction linéaire, possède une puissance plus intéressante.

$$f_{1,2} = (f_2 - f_1)/T \quad (4.3)$$

$$u_{ref}[n] = \frac{1}{2} \cdot \sin(2 \cdot \pi \cdot (f_1 + n \cdot f_{1,2}/2) \cdot n/f_1); \quad (4.4)$$

Dans le cadre de cet exercice, les fréquences f_1 et f_2 sont choisies de manière à ce que le signal d'excitation généré couvre la plage de fréquence comprise entre 0 Hz et la fréquence de Nyquist, soit $f_1 = 0$ Hz et $f_2 = 22\,050$ Hz. La durée du signal de balayage a été fixée à 8192 échantillons, soit environ 0,186 seconde pour la fréquence d'échantillonage $f_{éch}$ utilisée de 44 100 éch/s, comme le montre l'équation 5.3 du chapitre suivant. Ce choix est justifié dans l'un des paragraphes suivants, dans la présente section, et découle des travaux de Zhang *et al.* (2008).

La représentation temporelle du signal généré n'est pas très intéressante, car il est difficile de percevoir la variation de la fréquence à cause de la durée importante du signal face à la période de l'oscillation. Cependant, une représentation spectrale à trois dimensions permet de bien visualiser le signal d'excitation généré. L'axe vertical se voit attribuer la fréquence, l'axe horizontal le temps, et l'échelle de couleur donnent l'amplitude de la transformée en dB. De tel spectrogrammes sont utilisés tout au long de cet ouvrage.

Afin de fermer la boucle de mesure montrée par la figure 1.5, la conversion analogique-numérique doit être lancée au moment marquant le début de l'émission du signal d'excitation afin de mettre en mémoire la réponse de la salle captée par le microphone de mesure. La figure 4.10 montre le spectrogramme du signal perçu par le microphone.

Pour obtenir l'amplitude exacte de ces atténuations, le passage par le calcul de la réponse impulsionnelle $h_{ref}[n]$ s'impose. Une fois la réponse $y_{ref}[n]$ de la salle au signal d'excitation $x_{ref}[n]$ enregistrée, celui-ci doit être déconvolué de $y_{ref}[n]$ pour obtenir la réponse impulsionnelle de la salle $h_{ref}[n]$. Farina (2000) utilise la méthode de la division dans le domaine fréquentiel comme moyen efficace de calculer la déconvolution circulaire. Dans la section 4.3, on voit plus en détail cette relation qui existe entre la convolution linéaire et circulaire et comment cette relation peut s'appliquer à la déconvolution linéaire et circulaire.

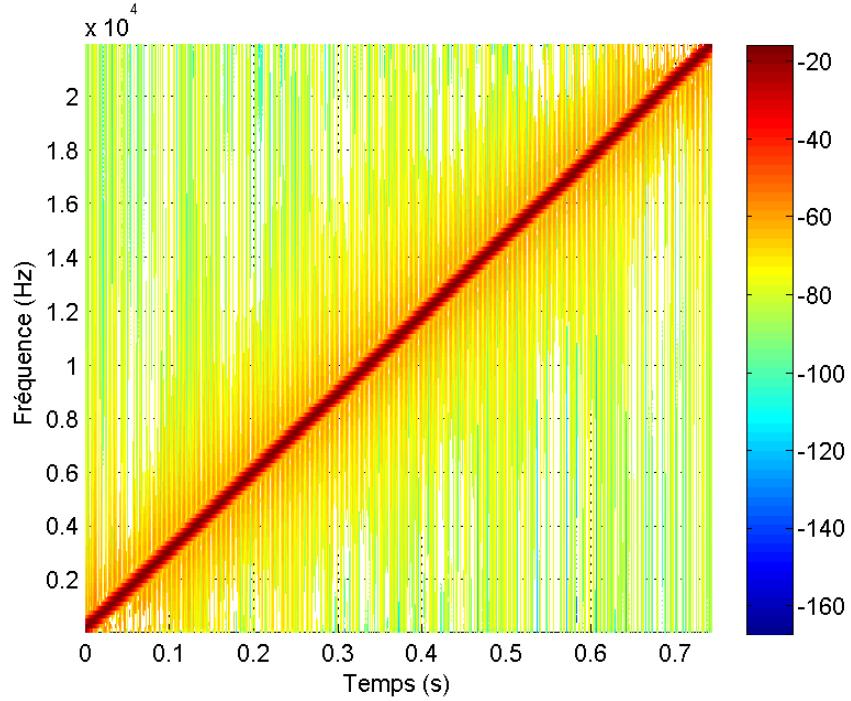


Figure 4.9 Balayage sinusoïdal de la plage de fréquence

$$h_{ref} \approx y_{ref} \div_{lin} u_{ref} \quad (4.5)$$

Bien que l'impulsion de Dirac en tant que signal d'excitation ne soit jamais intervenue dans la procédure de modélisation suivie, le vecteur $h_{ref}[n]$ correspond au signal mesuré supposant l'émission d'une impulsion pure.

Le résultat obtenu par calcul ne peut pas être exploité directement afin de modéliser le comportement de la salle. Le fait qu'il soit techniquement très difficile de lancer l'enregistrement de la réponse $y_{ref}[n]$ au moment exact où commence l'émission du signal $u_{ref}[n]$ doit être considéré. Afin d'assurer l'intégrité des mesures, le démarrage de l'enregistrement précède d'une fraction de seconde le lancement de la conversion numérique-analogique du signal d'excitation généré. La période entre laquelle les instructions lançant l'acquisition et celles lançant l'émission s'exécutent dépend souvent du nombre d'interruptions générées à ce moment par le processeur. Cette période est donc de durée variable dans le cas d'une carte de son ne supportant pas le standard ASIO. Dans le cadre de ce travail, le délai intrinsèque à la carte de son utilisée et décrite à la section 5.5 fut mesuré expérimentalement et l'alignement entre les blocs de données $y[n]$ et $u[n]$ servant au calcul de $h[n]$ est effectué en conséquence.

La résolution spectrale obtenue pour chaque bloc analysé est directement proportionnelle

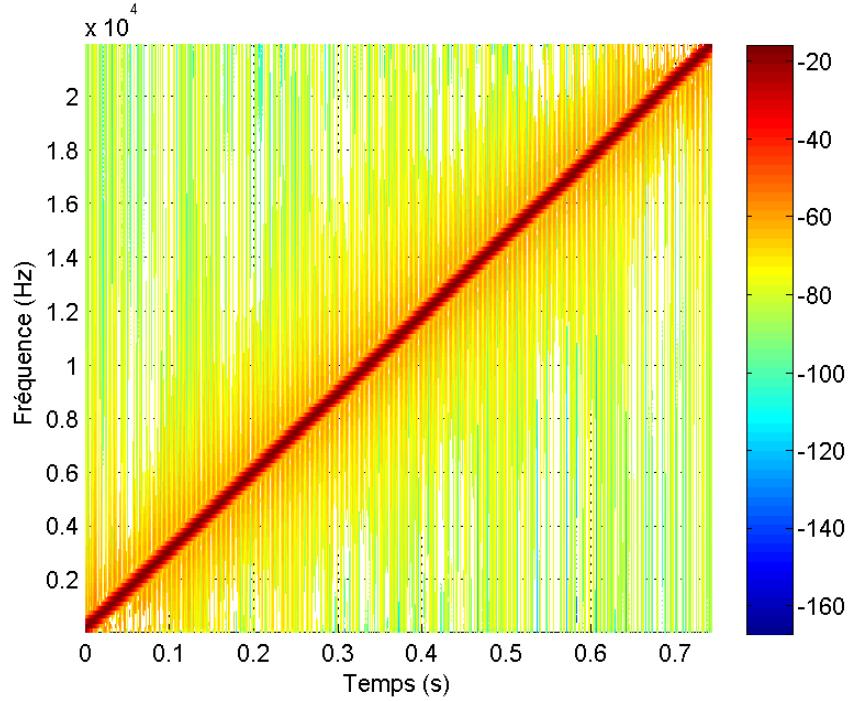


Figure 4.10 Réponse de la chaîne audio au balayage sinusoïdal de la plage de fréquence

à la taille N du bloc. Afin de bénéficier d'une bonne résolution ainsi que pour garantir une certaine richesse spectrale à l'intérieur du même bloc, un compromis doit être fait entre la dynamique et la performance du système lorsqu'on choisit la taille des blocs. Zhang *et al.* (2008) montre des résultats de test où le champ direct ainsi que les premières réflexions de la réponse d'une salle furent égalisés à l'aide de filtres numériques de 2048, 4096 et 8192 échantillons. Les filtres sont testés sur quatre différentes salles ayant des temps de réverbération de 200 ms, 400 ms, 1000 ms, et 1200 ms. Bien que le filtre plus court ne s'avère qu'à fonctionner que pour les salles avec de court temps de réverbération typiquement de 200 ms ou moins, ils ne notent aucune perte de performance en utilisant un filtre plus long même sur une salle ayant un temps de réverbération court. Seul le filtre de 8192 échantillons mène à l'obtention de résultats satisfaisants pour l'ensemble des quatre salles. Dans ce travail, la taille choisie est de $N = 8192$ échantillons, soit 186 ms au taux d'échantillonnage de 44,1 kHz. La résolution spectrale est donnée par

$$\delta_f = \frac{f_{\text{éch}}}{N} = 5,38 \text{ Hz} \quad (4.6)$$

4.5 Technique de modélisation suggérée

La présente section a pour but d'aborder le sujet de la modélisation d'une chaîne en audio en n'exploitant que des signaux musicaux. Plus précisément, il sera question de la méthode novatrice suggérée dans ce travail dans le but d'obtenir la réponse impulsionale d'une chaîne audio composée d'un amplificateur, d'un haut-parleur ainsi que de la salle dans laquelle a lieu l'écoute musicale à l'aide du signal capté par un microphone placé à la position d'écoute.

Le signal émis est noté ici $u[n]$, celui capté par le microphone de mesure est noté $y[n]$. Le spectre du signal émis dans la salle à travers le haut-parleur est donnée par sa transformée de Fourier discrète

$$U_k = \sum_{k=0}^{N-1} u_n \cdot \omega^{-kn} \quad (4.7)$$

et le spectre du signal reçu dans la salle par un microphone placé à la position d'écoute est donnée par sa transformée de Fourier discrète

$$Y_k = \sum_{k=0}^{N-1} y_n \cdot \omega^{-kn} \quad (4.8)$$

où $\omega = e^{2\pi \cdot i / N}$.

Le résultat est deux spectres de longueurs égales à celle du filtre égalisateur, celle du signal en entrée U et celle du signal en sortie Y . À partir de ces deux séquences discrètes, l'objectif est d'obtenir la réponse en fréquence H ainsi que la réponse impulsionale $h[n]$ de la chaîne audio. La définition même de la convolution linéaire discrète sert à établir la relation entre l'entrée $u[n]$ et la sortie $y[n]$ d'un système linéaire en fonction de la réponse impulsionale $h[n]$ de ce dernier. Ici l'opérateur de convolution est représenté par $*$.

$$y = u * h \quad (4.9)$$

Il est possible d'isoler la réponse du système $h[n]$ dans cette équation en passant par l'opérateur réciproque à celui de convolution linéaire, soit la déconvolution linéaire \div_{lin} . Notons qu'on utilise ici le symbole de division \div accompagné de « lin » ou de « circ » pour noter la déconvolution entre deux signaux dans le domaine temporel. À ne pas confondre avec la division complexe proprement dite qu'on utilise pour calculer cette déconvolution dans le domaine fréquentiel. La déconvolution du signal en entrée de la chaîne audio $u[n]$ à celui mesuré en sa sortie $y[n]$ permet d'obtenir directement la réponse impulsionale $h[n]$ de la chaîne.

$$h = y \div_{lin} u \quad (4.10)$$

Les propriétés de la convolution linéaire et de la convolution circulaire évoquées dans la section 2.3 du chapitre 2 s'appliquent également à la relation qui existe entre la déconvolution linéaire \div_{lin} et la déconvolution circulaire \div_{circ} . Dans le cas de la convolution, la méthode « overlap-add » couverte dans la section 4.9 permet de calculer efficacement une convolution linéaire, l'approximation de celle-ci par une convolution circulaire n'est pas nécessaire. Cependant, cette méthode ne s'applique pas à la déconvolution linéaire, le calcul d'une déconvolution linéaire est complexe, ce qui justifie l'approximation de celle-ci par la convolution circulaire. La déconvolution circulaire \div_{circ} est l'opération inverse de la convolution circulaire $*_{circ}$. L'intérêt de la déconvolution circulaire dans notre cas vient de l'efficacité avec laquelle, lorsque calculée dans le domaine fréquentiel, elle permet d'obtenir la réponse impulsionnelle de la chaîne audio $h[n]$ à partir de l'entrée $u[n]$ et la sortie $y[n]$.

$$h \approx y \div_{circ} u = \mathcal{T}\mathcal{F}\mathcal{D}\mathcal{I}\left(\frac{\mathcal{T}\mathcal{F}\mathcal{D}(y)}{\mathcal{T}\mathcal{F}\mathcal{D}(u)}\right) \quad (4.11)$$

Farina (2000) suggère aussi l'utilisation de la méthode reposant sur la division dans le domaine fréquentiel comme moyen efficace de calculer la déconvolution circulaire. La figure 4.11 montre une telle technique appliquée à un signal audio continu émis dans une salle $u[n]$ et un signal audio continu capté par un microphone dans l'environnement d'écoute $y[n]$. Elle montre que chacun de ces deux signaux est décomposé en blocs de L échantillons, et que la déconvolution circulaire est utilisée pour calculer une réponse impulsionnelle $h_i[n]$ de L échantillons pour chacun de ces blocs.

On note encore une fois qu'on utilise dans la figure 4.11 le symbole de division \div pour noter la déconvolution (circulaire dans ce cas) entre deux signaux dans le domaine temporel. Cette notation ne doit pas être confondue avec la division complexe proprement dite qu'on utilise pour calculer une déconvolution en fréquences.

Si on détient au préalable de l'information sur la façon avec laquelle l'énergie s'atténue au fil du temps dans la réponse impulsionnelle $h[n]$ de la chaîne audio, on peut choisir volontairement une longueur d'enregistrement N de l'entrée $u[n]$ et la sortie $y[n]$ pour que la queue de la réponse obtenue par déconvolution circulaire des deux séquences ne contienne que très peu d'énergie. La définition de « très peu d'énergie » est arbitraire. Nous savons cependant que dans la chaîne amplificateur/haut-parleur/salle, la salle est des trois systèmes celui qui emmagasine de l'énergie qui met une certaine quantité de temps à se dissiper. La définition du temps de réverbération donnée à la section 2.1 est utilisée. Le temps de réverbération est une mesure du temps mis par la salle à décroire son énergie acoustique

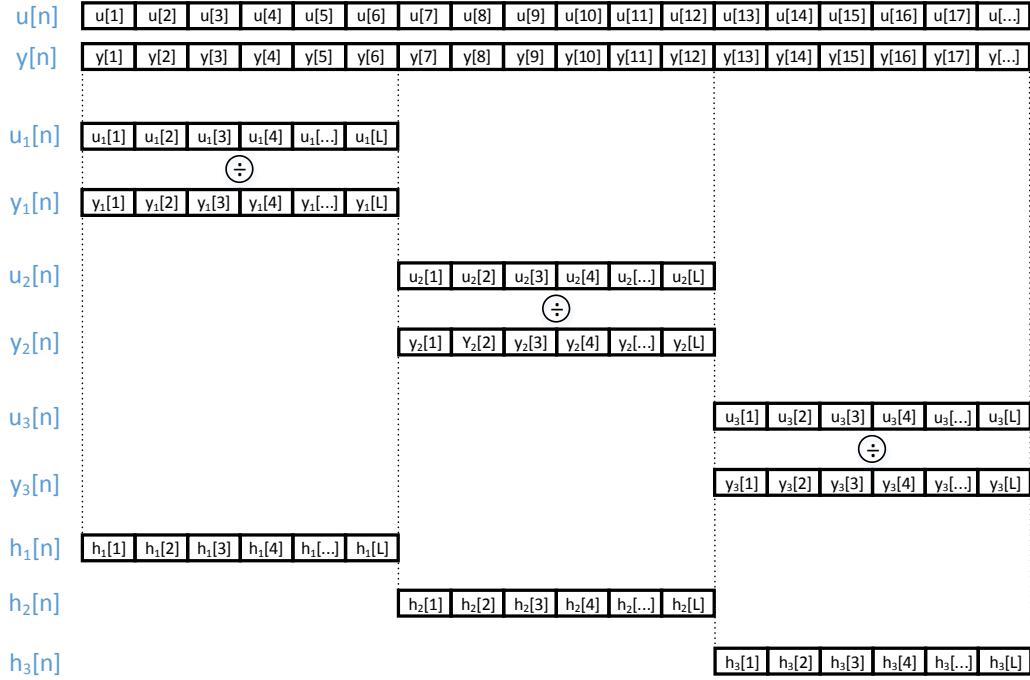


Figure 4.11 Approximation de la réponse d'un système discret par déconvolution circulaire

interne de 60 dB suite à l'interruption d'une source de buit constante à large bande. Puisque la salle est considérée comme linéaire, le taux de décroissance exponentielle de l'énergie dans la salle suite à une excitation constante et le taux de décroissance exponentielle de l'énergie dans la salle suite à une excitation impulsionale sont les mêmes. On établit donc que la partie repliée dans la réponse impulsionale $h[n]$ doit se situer après le temps de réverbération dans la réponse impulsionale pour que l'approximation $y \div_{circ} u \approx y \div u$ soit valide. Les séquences $u[n]$ et $y[n]$ étant de même longueur N , et la réponse $h[n]$ obtenue par déconvolution circulaire aussi, une valeur de N équivalent à doit être utilisée. Pour un taux d'échantillonnage $f_{éch}$ fixe, plus N est grand, plus grand peut être le temps de réverbération de la salle à l'étude, mais plus le temps de calcul de la déconvolution circulaire \div_{circ} grandit. Lorsqu'on effectue son calcul en temps réel à un flux d'échantillons continu qui défile, le temps alloué pour calculer $h[n]$ est fonction de N et du taux d'échantillonnage :

$$T = \frac{N}{f_{éch}} \quad (4.12)$$

La relation entre le temps de calcul alloué augmente proportionnellement à N , mais la complexité du calcul augmente plus rapidement, proportionnellement à $N \cdot \log(N)$, imposé par la complexité du calcul de la $\mathcal{T}\mathcal{F}\mathcal{D}$ et de la $\mathcal{T}\mathcal{F}\mathcal{D}\mathcal{I}$. Il existe donc une taille $N_{critique}$ à partir de

laquelle le calcul n'est plus réalisable en temps réel selon la puissance du processeur utilisé et l'implémentation logicielle du calcul. Cette valeur $N_{critique}$ fut trouvée expérimentalement et se situe autour de $N_{critique} = 200\,000$ selon l'implémentation actuelle et le processeur utilisé. Afin de ne pas monopoliser toutes les ressources disponibles, la valeur de N utilisée dans ce travail est de 8192. Cette valeur est bien en dessous de la valeur critique, mais ce critère n'est pas le seul à considérer. Dans la section 4.4, on a vu que les travaux de Zhang *et al.* (2008) suggèrent différents choix de N et présentent des résultats expérimentaux dans chaque cas. Le signal numérique envoyé en entrée de la chaîne audio $u[n]$ et celui en sortie $y[n]$ capturé à l'aide d'un microphone placé dans l'environnement d'écoute sont tous deux décomposés en bloc de 8192 échantillons. L'opération de déconvolution circulaire est appliquée à $u[n]$ et $y[n]$ ce qui permet de récupérer la réponse impulsionnelle de la chaîne audio, elle aussi sur 8192 échantillons. Pour que l'approximation par déconvolution circulaire soit correcte, le temps de réverbération de la salle ne doit pas dépasser

$$T = \frac{8192 \text{ éch}}{44\,100 \text{ éch/s}} = 0,186 \text{ s} \quad (4.13)$$

La salle à l'étude dans le cadre de ce travail a un temps de réverbération mesuré de 0,3 s. La mesure fut effectuée suite à l'interruption d'un bruit blanc constant émis sur haut-parleur dans la salle. Un microphone placé dans l'environnement d'écoute a permis d'observer le temps nécessaire pour que l'énergie acoustique décroisse de 30 dB dans la salle suite à l'interruption du bruit. Cette mesure sur 30 dB fut ensuite extrapolée sur 60 dB pour obtenir le résultat de 0,3 s. Cette technique reposant sur la déconvolution circulaire du signal numérique envoyé en entrée de la chaîne audio $u[n]$ à celui en sortie $y[n]$ capturé à l'aide d'un microphone placé dans l'environnement d'écoute permet l'estimation de la réponse impulsionnelle de la salle pour chaque bloc de 8192 échantillons d'un signal musical défilant.

Finalement, la technique de modélisation telle qu'appliquée dans le cadre de ce travail diffère légèrement de celle montrée dans la figure 4.11. La figure 4.12 illustre la technique de modélisation telle qu'elle est appliquée.

On verra dans les sections 5.2 et 5.3 du chapitre 5 que les échantillons en provenance du périphérique audio prennent la forme de nombres entier à virgule fixe et qu'une opération supplémentaire de conversion de type doit être effectuée pour que l'algorithme de déconvolution opérant sur des nombres à virgule flottante puisse fonctionner. Aussi, les échantillons du signal en provenance du périphérique audio se présentent par blocs de 1024 échantillons, alors que l'algorithme de déconvolution s'attend à recevoir des blocs de 8192 échantillons en entrées.

La période de temps sur laquelle s'effectue l'analyse du comportement de la salle dans le but d'en extraire la réponse impulsionnelle est arbitraire. La durée de l'analyse aura un

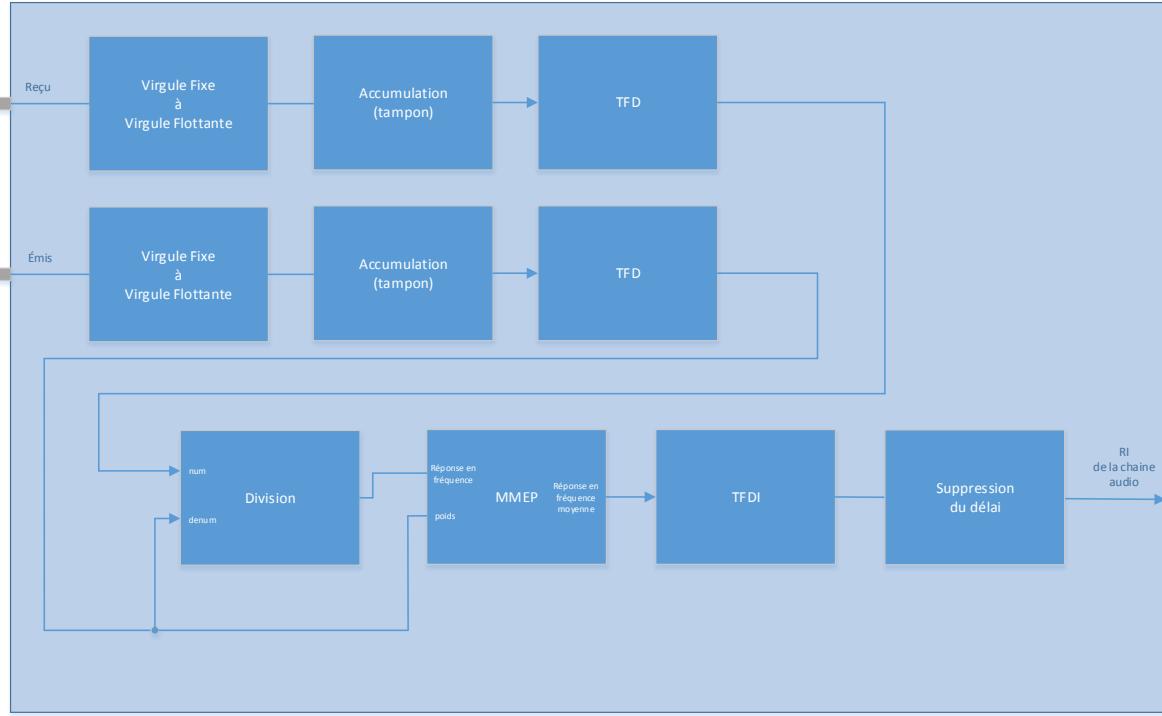


Figure 4.12 Implémentation de l'algorithme de modélisation de la chaîne audio

impact notamment sur la robustesse du système. De nature, le contenu fréquentiel des signaux musicaux change au fil du temps. C'est cette richesse spectrale, lorsqu'analysée sur une longue fenêtre de temps, qui nous permet à l'aide des signaux musicaux d'exciter la salle sur la pleine bande de fréquences pour y analyser son comportement. Cependant, lorsqu'analysé ponctuellement sur une très courte fenêtre de temps, le signal musical risque d'être peu riche. Dans un premier bloc d'analyse se trouveront certaines fréquences, dans un second d'autres. Le fait d'allonger la période d'analyse, donc d'augmenter la taille N des blocs sur lesquels sera appliquée la transformée de Fourier, assure un certain moyennage temporel du contenu fréquentiel. Les ordinateurs modernes bénéficient de librairies optimisées pour le calcul des transformées de Fourier rapides. Pour un signal audio échantilloné à 44,1 kHz, il en prend moins de temps pour calculer la transformée de Fourier rapide sur quelques secondes de données qu'il en prend à accumuler ces données en question.

Une accumulation de 64 blocs de 1024 échantillons chacun est nécessaire à la construction du bloc de $N = 8192$ échantillons utilisé pour la déconvolution. La conversion de type ainsi que l'accumulation sont montrées dans la figure 4.12.

Dans la figure 4.11, on obtient pour chaque bloc de L échantillons analysés, une estimation de la réponse impulsionnelle $h_i[n]$ de la chaîne audio. La précision de chacune de ces réponses

impulsionnelles dépend du contenu du signal d'excitation $u_i[n]$ ayant servi à son calcul. On verra dans les paragraphes qui suivent comment une moyenne mobile exponentielle pondérée peut être utilisée pour combiner chacune des réponses $h_i[n]$ obtenues en une seule réponse impulsionnelle pondérée $h[n]$ qui servira par la suite à la conception du filtre égalisateur adaptatif $g[n]$.

La moyenne mobile exponentielle pondérée notée MMEP dans la figure 4.12 a trois principaux objectifs. Le premier est de combiner la réponse en fréquence $H_i[j\omega]$ obtenue pour chaque bloc i de données analysées. L'aspect innovateur de l'algorithme développé dans ce travail repose sur le fait qu'on tente de modéliser un système expérimentalement à l'aide de signaux d'excitation qu'on peut observer, mais qu'on ne contrôle pas. C'est ce qui différencie ce travail de l'ensemble des travaux recensés tentant de modéliser l'acoustique d'une salle ou d'une chaîne audio complète expérimentalement. Tel que mentionné dans le chapitre, les techniques de modélisation expérimentales font typiquement appel à des signaux d'excitation dédiés à la mesure ; des signaux dotés de caractéristiques favorables à la simplification du calcul de réponse impulsionnelle qui en découle. Les paragraphes de la présente section (4.5) avaient justement pour but d'expliquer comment le calcul de la réponse impulsionnelle $h_i[n]$ ou de la réponse en fréquence $H_i[j\omega]$ peut être effectué pour un bloc de données analysé lorsque le signal d'excitation est, dans une certaine mesure, aléatoire. Cependant, une seconde propriété des signaux d'excitation typiquement utilisés dans les techniques standards de modélisation expérimentales est la couverture de la pleine plage de fréquence. Qu'il s'agisse d'utiliser une sinusoïde balayant la plage de fréquence ou un bruit blanc à large bande, les signaux d'excitation sont choisis afin de couvrir la pleine plage de fréquence. La figure 4.9 de la section 4.4 montre le spectrogramme de la sinusoïde balayant la plage de fréquence de 0 à 22 050 Hz utilisé dans ce travail pour l'obtention du modèle de référence. La figure 4.13 montre le spectrogramme d'un bruit blanc de même durée qui aurait aussi pu être utilisé pour l'obtention du modèle de référence.

Par opposition, la technique de modélisation développée utilise les signaux musicaux comme signaux d'excitation. Rien ne garantit dans ce cas que pour chaque bloc de données analysé de 8192 échantillons, soit 0,186 seconde, que la pleine plage de fréquence sera couverte comme le montre la figure 4.14.

La figure 4.14 montre le spectrogrammes d'une section de 0,186 seconde prise au hasard dans un extrait musical. On note que la totalité de la plage de fréquence allant de 0 à 22 050 Hz n'est pas forcément couverte par chaque bloc individuellement. Une combinaison des i réponses en fréquence $H_i[j\omega]$ consécutives obtenues est nécessaire pour que l'ensemble des fréquences soient couvertes par une réponse en fréquence combinée $H[j\omega]$. Une moyenne arithmétique mobile fut initialement considérée pour cette combinaison.

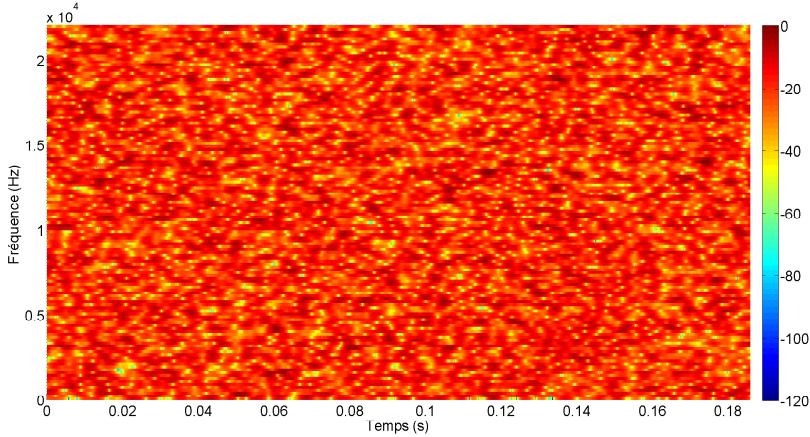


Figure 4.13 Bruit blanc gaussien de 0,186 ms (échelle de couleur en dBFS)

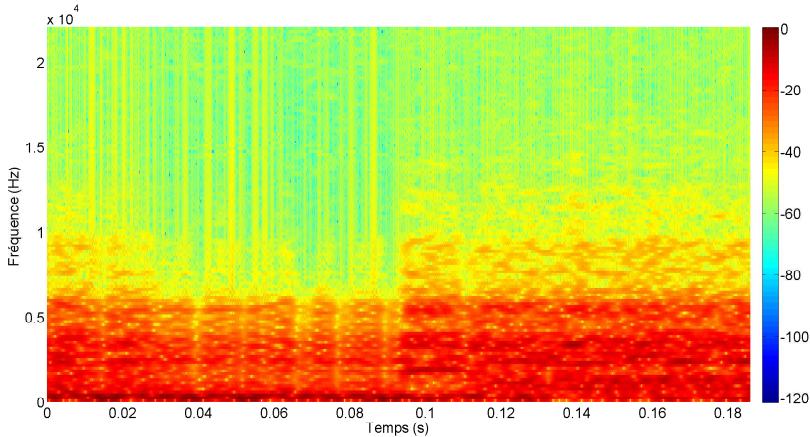


Figure 4.14 Spectrogramme d'une section de 0,186 seconde prise au hasard dans la chanson « Wickerman » (échelle de couleur en dBFS)

$$H[j\omega] = \frac{\sum_{k=1}^i H_k[j\omega]}{i} \quad (4.14)$$

Avec cette approche, la réponse en fréquence combinée après le défilement de i blocs $U_i[j\omega]$ de musique est une moyenne arithmétique de chacune de ces i réponses en fréquence $H_i[j\omega]$. L'hypothèse émise est qu'au fil du temps le contenu fréquentiel du signal musical varie et que cette combinaison des blocs mènera à l'obtention d'une réponse en fréquence $H[j\omega]$ couvrant mieux l'ensemble des fréquences que chacun de ces blocs $H_i[j\omega]$ analysés individuellement.

Un problème observé avec cette méthode de combinaison ne reposant que sur la moyenne arithmétique est que la pondération de la pleine plage de fréquences est la même pour chaque bloc de données, indépendamment de la présence marquée ou non de ces fréquences dans le

signal musical d'excitation $U_i[j\omega]$, pour ce bloc. Ceci cause un problème. Lorsque le signal d'excitation $U_i[j\omega]$ ne contient que très peu d'énergie pour une plage de fréquence $\omega_a < \omega < \omega_b$ donnée, la réponse en fréquence $H_i[j\omega]$ obtenue par la division de $Y_i[j\omega]$ par $U_i[j\omega]$ est erronée pour cette plage. La notion de « très peu d'énergie » fait référence ici au plancher de bruit du signal $Y_i[j\omega]$ capté par le microphone à la position d'écoute. Prenons le bloc de 8192 échantillons de la figure 4.14 comme exemple. Pour faciliter la compréhension, prenons comme exemple une chaîne audio idéale, n'entrant aucune transformation entre le signal y étant transmis et celui capté par un microphone à la position d'écoute.

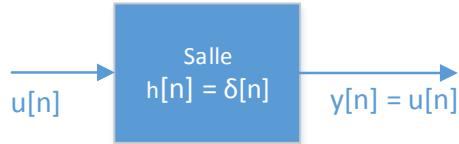


Figure 4.15 Chaîne audio idéale

De plus, supposons que le plancher de bruit du signal $Y_i[j\omega]$ capté par le microphone soit constant à -40 dB sur toute la plage de fréquences. Le spectre du signal émis $U_i[j\omega]$ pour ce bloc ainsi que celui capté par le microphone $Y_i[j\omega]$ sont montrés dans la figure 4.16. En observant le signal $Y_i[j\omega]$ capté par le microphone, on note que les portions du spectre où le signal $U_i[j\omega]$ est inférieur au plancher de bruit sont erronées ; rien ne peut être reçu en deçà du plancher de bruit. En traçant la réponse en fréquence $H_i[j\omega]$ obtenue pour ce bloc, on observe que cette réponse est elle aussi erronée pour les portions du spectre où le signal transmis est inférieur au plancher de bruit.

Le deuxième objectif de la moyenne mobile exponentielle pondérée notée MMEP dans la figure 4.12 est de contrer ce phénomène. La première avenue considérée est celle de la détection du plancher de bruit. Il aurait été possible, par exemple, d'exploiter les blocs dans lesquels certaines fréquences sont peu présentes dans le but de caractériser le plancher de bruit à mesure que la musique défile. Un plancher de bruit connu permettrait de ne pas tenir compte des fréquences sous ce plancher dans le calcul de la moyenne mobile de l'équation 4.14. Cette comparaison entre l'amplitude du signal d'excitation $U_i[j\omega]$ et le plancher de bruit $B[j\omega]$ doit être faite pour chaque fréquence $[j\omega]$.

$$H[j\omega] = \begin{cases} \frac{\sum_{k=1}^i H_k[j\omega]}{i} & U_i[j\omega] > B[j\omega] \\ H[j\omega] & \text{sinon} \end{cases} \quad (4.15)$$

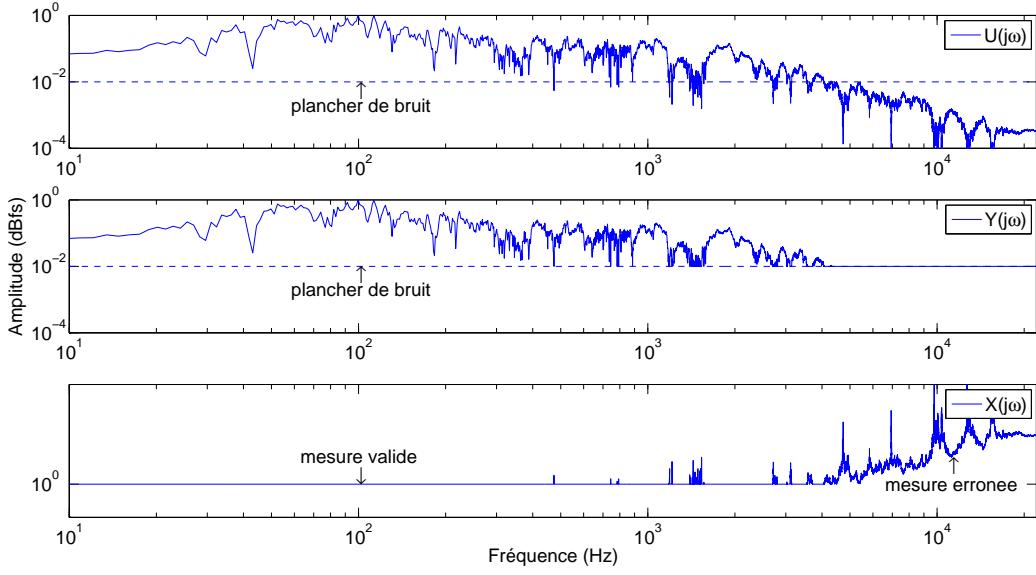


Figure 4.16 Impact du plancher de bruit sur la mesure

Ceci est équivalent à utiliser une moyenne pondérée afin de combiner les réponses en fréquence individuelles $H_i[j\omega]$ à une réponse pondérée $H[j\omega]$. On utilise un facteur de pondération nul pour les portions de $U_i[j\omega]$ sous le plancher de bruit et un facteur unitaire pour celles au-dessus. Le vecteur de poids $W_i[j\omega]$ variant pour chaque bloc de données traitées serait

$$W_i[j\omega] = \begin{cases} 1 & U_i[j\omega] > B[j\omega] \\ 0 & \text{sinon} \end{cases} \quad (4.16)$$

L'équation 4.15 devient

$$H[j\omega] = \frac{H[j\omega] \cdot W[j\omega] + H_i[j\omega] \cdot W_i[j\omega]}{W[j\omega] + W_i[j\omega]} \quad (4.17)$$

où

$$W[j\omega] = \sum_{k=1}^{i-1} W_i[j\omega] \quad (4.18)$$

Cette méthode souffre d'une lacune majeure. À l'initialisation de l'algorithme, si le premier bloc de musique à être joué ne contient pas de composantes à toutes les fréquences du spectre à une amplitude supérieure à celle du plancher de bruit, ces fréquences dans $H_1[j\omega]$ ne contribuent pas au premier calcul de la moyenne $H_i[j\omega]$. On se retrouve donc sans modèle

jusqu'à ce que les i premiers $U_i[j\omega]$ aient excité toutes les fréquences avec une amplitude supérieure au plancher de bruit au moins une fois, pour la durée d'un bloc de 8192 échantillons. La seconde avenue considérée, et celle choisie, repose encore sur l'utilisation de la moyenne mobile pondérée de l'équation 4.17. Cette fois, on utilise un poids continu plutôt qu'un poids n'ayant que 2 valeurs (0 ou 1) tel que c'était le cas pour l'équation 4.19. On définit le vecteur de poids $W_i[j\omega]$ pour un bloc de données i comme étant égal à l'amplitude de chacune des fréquences $[j\omega]$ dans le signal d'excitation musicale $U_i[j\omega]$.

$$W_i[j\omega] = |U_i[j\omega]| \quad (4.19)$$

L'équation 4.17 devient

$$H[j\omega] = \frac{H[j\omega] \cdot W[j\omega] + H_i[j\omega] \cdot U_i[j\omega]}{W[j\omega] + U_i[j\omega]} \quad (4.20)$$

où

$$W[j\omega] = \sum_{k=1}^{i-1} U_i[j\omega] \quad (4.21)$$

Ceci revient à dire que lorsqu'un nouveau bloc de données i est analysé, pour chaque fréquence $[j\omega]$, le poids de réponse en fréquence $H_i[j\omega]$ dans la réponse pondérée $H[j\omega]$ est proportionnel à l'amplitude de cette fréquence $[j\omega]$ dans le signal d'excitation $U_i[j\omega]$. Pour une fréquence $[j\omega]$ donnée, si cette fréquence était très présente dans le signal musical $U_i[j\omega]$ pour ce bloc i , alors cette fréquence de H_i contribuera beaucoup dans la réponse pondérée H . Par contre, si cette même fréquence $[j\omega]$ était peu présente dans le signal musical $U_i[j\omega]$ pour ce bloc i , elle contribuera peu dans la réponse pondérée H . On émet l'hypothèse que cette technique permet d'approximer, à mesure que les blocs de musique défilent, la réponse en fréquence réelle de la chaîne audio. Cette hypothèse est validée à la section 5.2 à l'aide d'une implémentation Matlab reposant sur les équations 4.19 et 4.17. Les résultats sont concluants.

Le troisième et dernier objectif de la MMEP de la figure 4.12 est de contrôler la dynamique du système. La pondération de la moyenne mobile telle que définie par l'équation 4.21 est grandissante. Chaque fois qu'un nouveau bloc de données i est traité, les poids du vecteur $H[j\omega]$ se voient augmenter. On peut dire que l'algorithme de modélisation a une mémoire. Il tient compte des événements du passé et n'accorde pas moins d'importance aux réponses en fréquence obtenues quelques itérations après son initialisation (exemple : H_1, H_2, H_3) qu'il en accorde à celles obtenues dans un passé rapproché (exemple : $H_{i-1}, H_{i-2}, H_{i-3}$). Ceci n'est pas un problème lorsque la réponse du système à l'étude ne varie pas. Cependant, tel qu'indiqué dans la section 1.2 du chapitre 1, l'objectif du présent travail est de concevoir un

filtre adaptatif, donc un algorithme de modélisation capable de s'adapter aux changements dans la chaîne audio tels qu'un changement de haut-parleur ou lorsque le décor dans la pièce où a lieu l'écoute est changé et que l'acoustique de la salle se voit altérée. L'algorithme de modélisation devrait donc accorder une plus grande importance au comportement de la salle tel qu'il a pu être observé dans les quelques secondes précédant l'instant présent.

La moyenne exponentielle pondérée peut facilement s'adapter pour remplir ce besoin. On commence par reprendre l'équation 4.21 en l'exprimant sous une forme récursive où à chaque itération, on ajoute le signal d'excitation U de l'itération précédente au vecteur de points $W[j\omega]$ afin de le mettre à jour.

$$W[j\omega] = \sum_{k=1}^{i-1} U_i[j\omega] = W[j\omega] + U_{i-1}[j\omega] \quad (4.22)$$

Ensuite, le concept de MMEP est introduit. Un facteur supplémentaire est ajouté au calcul des poids de l'équation 4.21. Un facteur multiplicatif fractionnaire α est appliqué au vecteur de poids chaque fois qu'un nouveau bloc de données vient d'être traité et que les poids sont mis à jour. Ce facteur α , nommé « constante de temps », assure une dégradation exponentielle de la contribution des blocs traités antérieurement. On incorpore cette constante de temps α à l'équation 4.22 pour obtenir le nouveau vecteur de poids $W[j\omega]$.

$$W[j\omega] = W[j\omega] \cdot \alpha + U_{i-1}[j\omega] \quad (4.23)$$

La constante α reste un paramètre de l'algorithme et peut être ajustée pour contrôler la dynamique du système. Sa plage de valeurs se situe entre 0 et 1. Plus α est petit, plus rapidement l'algorithme de modélisation oublie les résultats issus des blocs de données traités dans le passé au profit d'une dynamique plus rapide. Un tel algorithme saurait rapidement adapter la réponse pondérée $H[j\omega]$ à un changement dans l'acoustique de la salle, par exemple. Le désavantage est une moins grande immunité au bruit puisque, comme on l'a vu précédemment, la raison d'être de la MMEP est une meilleure immunité au bruit lorsque pour un bloc de données analysé, i certaines fréquences $j\omega$ du signal d'excitation $H_i[j\omega]$ se trouvent sous le plancher de bruit $B[j\omega]$. Un $\alpha = 0$ résulterait en un algorithme de modélisation qui n'accorde aucune importance au passé et ne se fie qu'à la réponse en fréquence $H_i[j\omega]$ courante pour caractériser la chaîne audio. Par opposition, une grande valeur d' α rendrait l'algorithme de modélisation plus sensible au passé et donc mieux immunisé contre le bruit, bénéficiant de la pondération d'une plus grande quantité de blocs de données menant forcément à la reconstruction d'une réponse pondérée $H[j\omega]$ plus riche spectralement. Le désavantage d'un α trop grand, c'est la lenteur de la dynamique du système. Lorsque la réponse de la salle change par exemple, un α ne permet pas une adaptation rapide de la réponse pondérée $H[j\omega]$. La

figure 4.17 montre l'exponentielle décroissante appliquée au poids résultant de la multiplication par la constante de temps α . On remarque que plus α augmente, plus le poids accordé à un bloc analysé à un certain moment précis dans le passé augmente. On remarque que la courbe n'est pas lisse mais plutôt en escalier ; chaque plateau correspond à la longueur du bloc de données de 8192 échantillons, soit 0,186 seconde au taux d'échantillonage de 44100 échantillons par seconde.

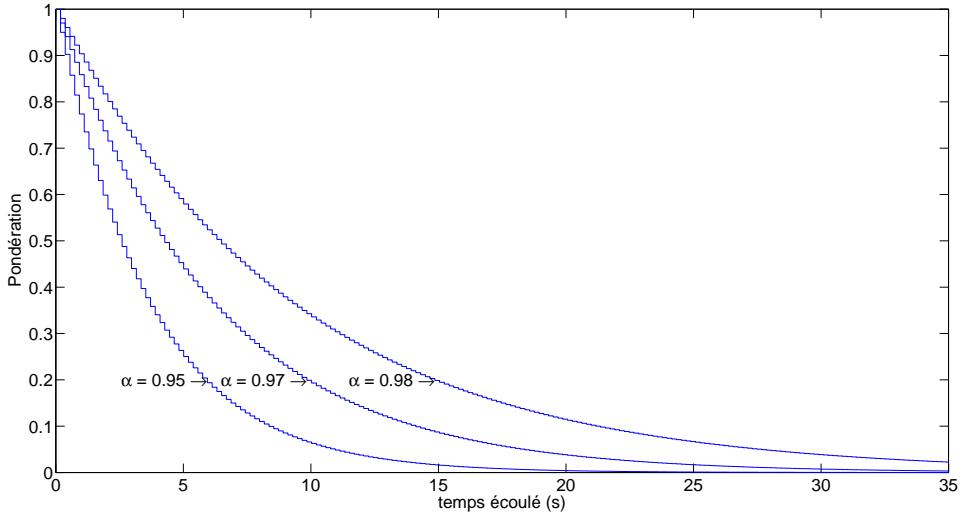


Figure 4.17 Pondérations décroissantes pour différentes valeurs d' α

Tel que vu précédemment, le choix de l'alpha a un impact direct sur la dynamique du système. On ne s'attend pas à ce que l'acoustique de la chaîne audio varie à un rythme particulier. C'est pourquoi alpha doit demeurer paramétrable. Dans le cadre de ce travail, on fixe arbitrairement à 20 secondes la période au-delà de laquelle on veut que l'algorithme de modélisation cesse de considérer les données analysées. On émet l'hypothèse que dans un contexte d'écoute musicale sur haut-parleur, le filtre adaptatif n'a nul besoin de réagir plus vite. Ce paramètre reste ajustable selon le contexte, mais on ne teste que la valeur de 20 secondes dans le présent travail. Comme le montre la figure 4.17, les fonctions exponentielles décroissantes ne permettent pas de couper abruptement le poids après 20 secondes, par exemple. On se donne plutôt un seuil, par exemple 95%, et on calcule la valeur d' α pour laquelle 95% du poids provient des 20 dernières secondes. Pour des blocs de 8192 échantillons, cette valeur d' α est 0,9637, tel qu'illustré par la figure 4.18.

La figure 4.19 montre l'exponentielle décroissante résultant du α choisi. C'est cette décroissance qui est appliquée aux anciens blocs de données dans notre cas.

La figure 4.20 illustre sommairement l'implémentation d'une telle moyenne mobile expo-

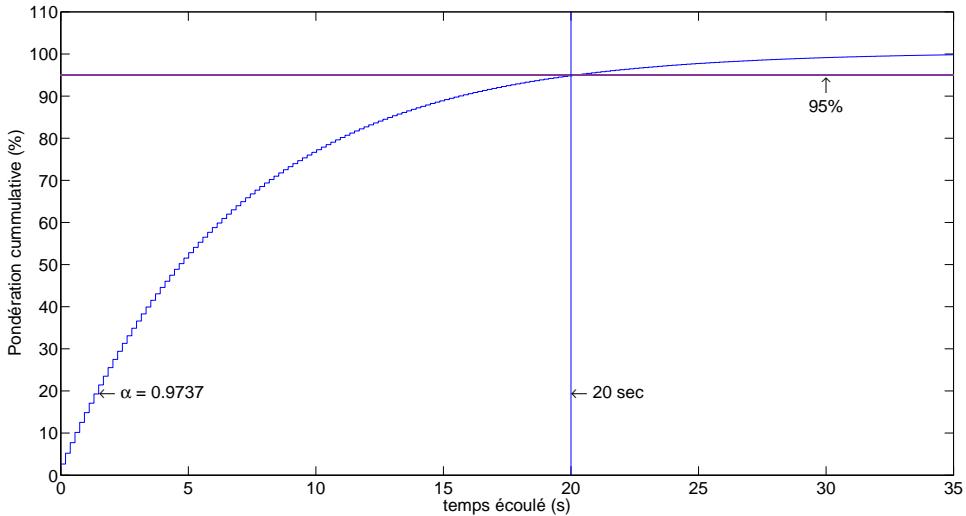


Figure 4.18 Choix d' α pour que 95% de la pondération provienne des 20 dernières secondes

nentielle pondérée.

Finalement, un certain délai existe dans la réponse impulsionale $h[n]$ obtenue à l'aide de l'algorithme de modélisation décrit précédemment. Ce délai de quelques millisecondes est dû au temps mis par l'onde de son à parcourir la distance de un mètre entre le haut-parleur et la position d'écoute. En utilisant $c = 330 \text{ m/s}$ comme vitesse du son, il est possible de calculer que ce délai est d'exactement 0,003 seconde pour les test effectués avec le microphone placé à 1 mètre de la source, et de 0,006 seconde pour les tests effectués avec le microphone placé à 2 mètres de la source. Pour que la méthode de modélisation suggérée fonctionne peu importe la distance entre le microphone et la source, un moyen de tronquer la réponse impulsionale mesurée doit donc être élaboré. Ce délai doit être éliminé avant d'entamer la conception du filtre égalisateur dont la réponse $g[n]$ découle de celle de la chaîne audio $h[n]$.

La norme (BSI) contient un algorithme d'alignement de la réponse impulsionale d'une salle. L'algorithme relativement simple consiste à parcourir à rebours le vecteur de la réponse calculée à partir du maximum absolu à la recherche d'un changement de signe de la première dérivée de la courbe. Le point correspondant est considéré comme le premier échantillon de la réponse impulsionale utile et les échantillons le précédent sont abandonnés. L'alignement utilisé dans le cadre de ce travail s'en inspire.

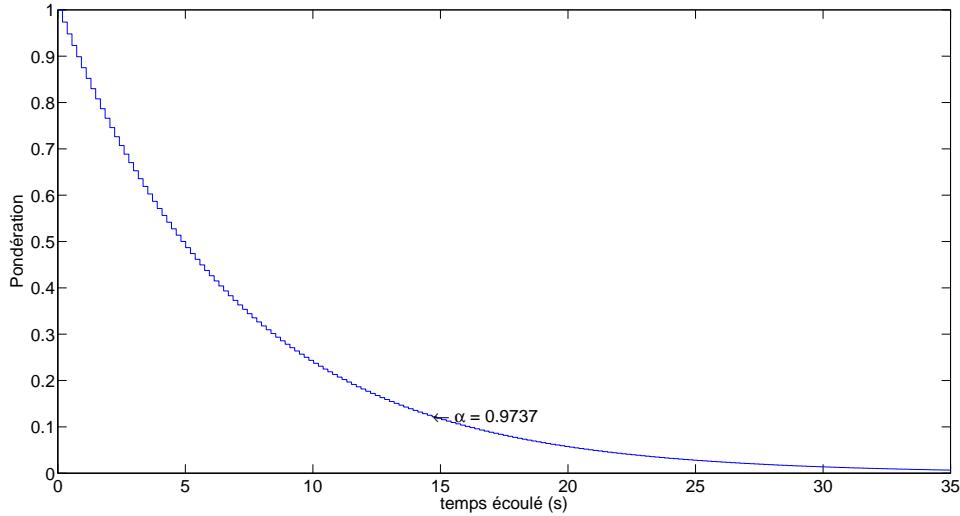


Figure 4.19 Pondérations décroissantes appliquées aux anciens blocs de données selon la valeur d' α choisie

4.6 Approche conventionnelle : domaine temporel

Qu'il s'agisse de système à temps continu ou à temps discret, il existe certaines conventions en analyse et en commande des systèmes dynamiques. Notamment, c'est souvent dans le domaine du temps qu'on décrit les différentes caractéristiques d'un système et qu'on mesure sa performance. Revenons à l'exemple du moteur électrique.

La nature du moteur est la principale cause de l'écart entre la commande $u(t)$ et la sortie observée $y(t)$, tout particulièrement lorsqu'on change subitement la vitesse désirée. Le but dans ce genre d'exercice est que la vitesse du moteur, exprimée en fonction du temps, s'apparente autant que possible à la vitesse commandée imposée par la tension électrique fournie au système variant elle aussi au fil du temps. Un problème souvent étudié pour le moteur à courant continu est le passage de l'état de repos à une vitesse constante en appliquant un échelon de tension à l'entrée du moteur. On s'intéresse à l'évolution de la vitesse de rotation du moteur au fil du temps durant cette phase transitoire résultant de son alimentation par l'échelon jusqu'à ce que la vitesse se stabilise et que le régime permanent soit atteint. Le sujet est analysé et traité dans le domaine temporel, le problème compte 4 paramètres, soit le temps de montée, le temps d'établissement, le dépassement et l'erreur statique.

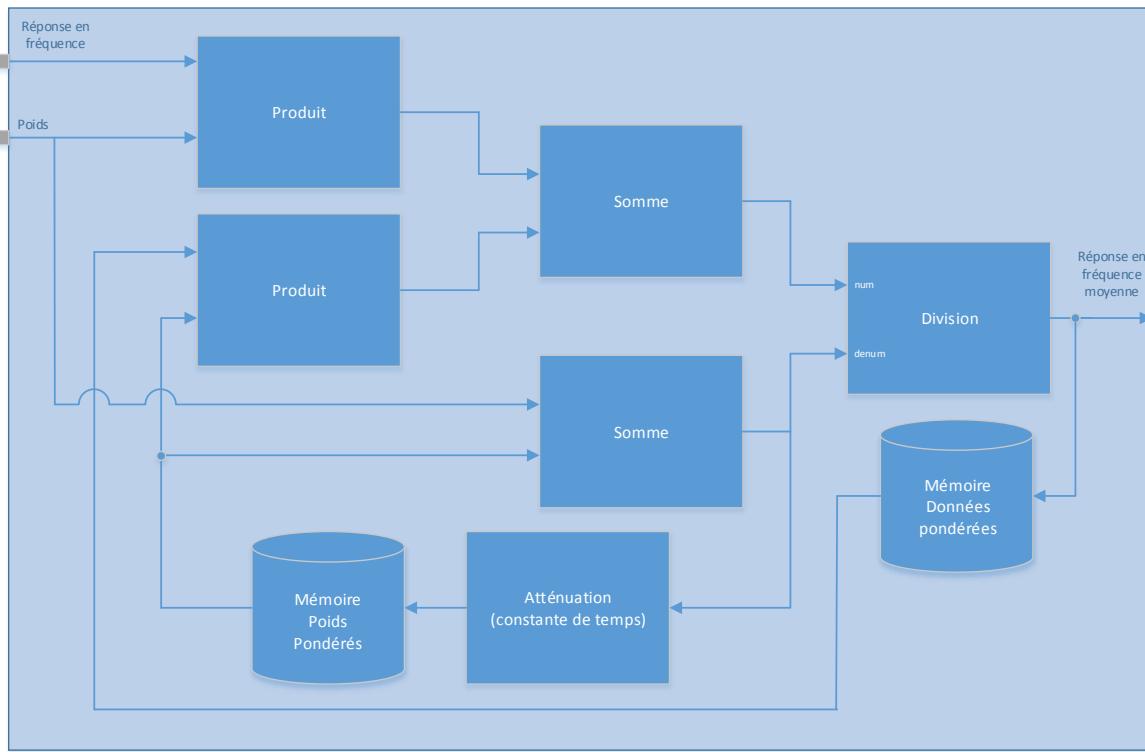


Figure 4.20 Implémentation de la moyenne mobile exponentielle pondérée

4.7 Approche alternative adaptée à la réponse en fréquence d'une salle

Revenons à l'exemple du moteur. Il est intéressant de noter que pour l'exemple du moteur, la vitesse de rotation est le phénomène physique de choix lorsqu'on veut établir une relation proportionnelle entre la tension électrique appliquée à l'entrée du moteur à courant continu et le phénomène mécanique d'entrainement du rotor qui en résulte. Lorsqu'alimenté avec une tension électrique constante, le moteur tourne à une vitesse constante. Il serait tout à fait possible d'observer un phénomène autre que la simple rotation, par exemple le déplacement d'une pièce mécaniquement entraînée par la rotation du rotor, un piston se déplaçant selon un mouvement de va-et-vient le long d'un axe fixe. Cet exercice plutôt intéressant permet d'observer que le déplacement d'une pièce entraînée par la rotation suit un patron sinusoïdal dont la fréquence est proportionnelle à la vitesse de rotation du moteur, et donc à la tension d'alimentation de ce dernier. Le concept de proportionnalité, dans ce cas, est transposé du domaine temporel au domaine fréquentiel. C'est la fréquence d'oscillation du piston qui est constante lorsque le moteur à courant continu opère en régime permanent. Le fait que certains phénomènes physiques sont plus faciles à observer dans le domaine fréquentiel qu'ils

le sont dans le domaine temporel ne se limite pas qu'au mouvement d'un piston sur un axe. C'est le cas pour bien des phénomènes physiques, notamment le son qui, en plus d'être bien adapté à l'analyse fréquentielle de par sa nature ondulatoire, se propage dans l'oreille interne le long de la cochlée où chaque cellule répond préférentiellement à une certaine fréquence, pour permettre à l'humain de percevoir les sons en termes de fréquences. Dans le cas où le phénomène physique étudié est l'acoustique, et tout particulièrement dans le cas où il est question de la modélisation de l'acoustique d'une salle, le passage par la représentation en fréquence est de mise.

L'erreur telle que définie dans les figures 4.3 à 4.7 s'applique mal au problème de l'égalisation de l'acoustique d'une salle. En fait, le terme *égalisation* sous-entend que c'est de la réponse en fréquence dont il est question. La métrique d'erreur utilisée doit être choisie conséquemment et doit mesurer la déformation introduite par la réponse en fréquence de la salle, donc mesurer la différence entre le spectre du signal de référence $r[n]$ et celui du signal observé en sortie $y[n]$. La notion de spectre est définie comme la transformée de Fourier discrète de ces deux signaux, soit $R[j\omega]$ et $Y[j\omega]$. La nouvelle notion d'erreur $E[K]$ est montrée dans la figure 4.21, où $K = j \cdot \omega$. La référence $r[n]$ est égale au signal source $u[n]$.

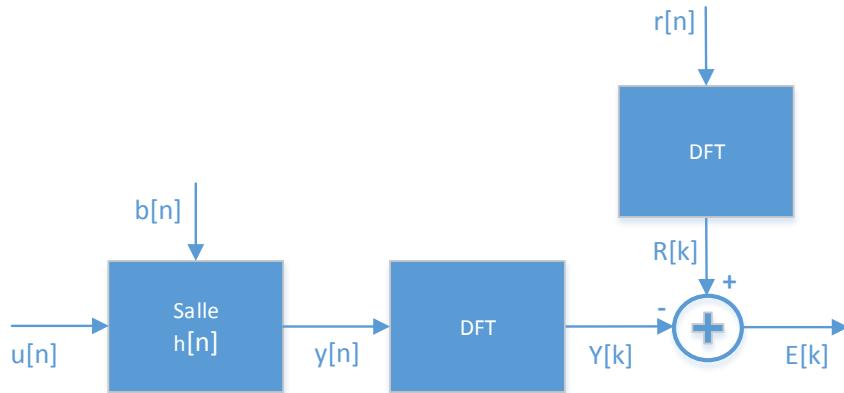


Figure 4.21 Erreur sur le spectre d'amplitude

L'objectif est de diminuer cette erreur $E[K]$ (ou $E[j\omega]$) en introduisant dans la chaîne audio le filtre égalisateur adaptatif conçu. Les sections à venir portent sur la conception de ce filtre égalisateur, soit 4.8 et 4.9.

Nous verrons dans la section 5.1 du chapitre 5 suivant comment cette notion d'erreur $E[k]$ se concrétise sous forme d'une métrique d'erreur, en l'occurrence l'erreur quadratique moyenne et peut être utilisée pour quantifier la performance de l'algorithme d'égalisation

dans un contexte expérimental tel que suggéré par Canclini *et al.* (2012).

4.8 Conception du filtre égalisateur

Dans cette section, nous verrons comment la réponse à l'impulsion de la chaîne audio $h[n]$ obtenue à l'aide de l'algorithme de modélisation présenté précédemment peut être utilisée afin d'obtenir la réponse impulsionnelle d'un filtre $g[n]$ qui servira de filtre égalisateur. Le but ici est de concevoir un filtre $g[n]$ qui égalisera les fréquences et sera agréable à l'oreille et non une pure inversion de la pleine réponse impulsionnelle de la chaîne audio. Une pure inversion de la pleine réponse impulsionnelle engendre la création d'énergie dans la réponse impulsionnelle du filtre avant l'impulsion principale, ce qui prend la forme de « pré-échos » lors de l'écoute. On a utilisé l'exemple du tambour dans la section 3.5 pour illustrer l'effet « pré-échos ». Ce phénomène est indésirable car il est désagréable durant l'écoute comme l'explique Hatziantoniou et Mourjopoulos (2003).

La réponse impulsionnelle d'une salle est constituée du son direct, des premières réflexions des murs et différentes surfaces ainsi que d'une réverbération tardive (Stewart et Sandler, 2007). Les premières réflexions sont un ensemble de réflexions discrètes de grande densité. Alors que le son direct est un événement unique qui peut être facilement identifié, il correspond dans notre cas à l'arrivée au microphone du premier front d'onde lorsqu'une impulsion est émise à travers le haut-parleur.

Les premières réflexions et la réverbération tardive sont plus difficiles à identifier ou à séparer. Dans une réponse impulsionnelle, l'amplitude des différentes réflexions à leur arrivée à la position d'écoute décroît au fil du temps jusqu'à atteindre une amplitude suffisamment faible ne permettant pas à l'oreille de les différencier individuellement les unes des autres. Ce passage marque le début de la troisième partie de la réponse impulsionnelle, soit la réverbération tardive. Cette troisième partie ne fait typiquement pas l'objet d'égalisation.

La distinction entre les premières réflexions et la réverbération tardive est aussi fonction de la fréquence. Les basses fréquences dont la longueur d'onde est du même ordre de grandeur que les obstacles constituant l'environnement dans lequel elles se propagent ont tendance à avoir un comportement modal, même stationnaire. Pour cette raison, en basses fréquences, la distinction entre les ondes de son dont la provenance est directe versus celles réfléchies fait peu de sens. La salle a plutôt un comportement modal en basses fréquences. Les concepts de champs direct et de champs réverbéré perdent leur sens, le désir de ne travailler que sur les quelques premières millisecondes de la réponse impulsionnelle de la salle aussi. De manière générale, pour les basses fréquences, on veut qu'un filtre égalisateur puisse corriger ce qui est perçu par l'oreille humaine comme faisant partie du comportement modal de la salle, donc

la pleine réponse impulsionales, soit 92 ms dans notre cas. Pour les hautes fréquences, on veut raccourcir cette durée. Certains auteurs tels que Stewart et Sandler (2007) suggèrent une approche analytique pour faire la distinction entre les premières réflexions, faisant l'objet d'égalisation, et la réverbération tardive ne faisant pas l'objet d'égalisation. Cependant, dans le présent travail, nous utilisons plutôt une approche par bande de fréquences, permettant de faire varier en fonction de la fréquence le temps critique dans la réponse impulsionale à partir duquel on ne désire plus appliquer de correction.

Dans le domaine temporel, la réponse impulsionale de la chaîne audio révèle les temps d'arrivée du signal direct et des réflexions sur différentes surfaces dans la salle ainsi que le taux de décroissance de celle-ci. Dans le domaine des fréquences, la réponse en fréquence de la chaîne audio présente la réponse de fréquence globale du système à l'étude. Cependant, les différentes propriétés, par exemple le taux de décroissance, des fréquences différentes ne sont pas observables à partir de ces représentations. L'analyse purement fréquentielle est utilisée pour donner un aperçu sur le comportement de la chaîne audio en fréquence (amplitude et phase). Les méthodes temps-fréquence permettent d'analyser les deux domaines conjointement et ainsi donner la possibilité de décrire en fonction du temps et de la fréquence les caractéristiques de signaux non stationnaires. La réponse impulsionale d'une salle est non stationnaire et a différentes propriétés d'atténuation et de déphasage à des fréquences différentes et est donc bien adaptée à l'analyse temps-fréquence (Sarris et Cambourakis, 2003). Les travaux recensés montrent que l'utilisation de plusieurs transformées de Fourier de différentes longueurs sur la même réponse impulsionale s'avère une méthode valide pour la décomposition par bandes de fréquences.

Ce concept ne se limite pas nécessairement à la séparation entre les basses et les hautes fréquences. Pour contrer les problèmes causés par l'inversion de la pleine réponse en fréquence, Rubak et Johansen (2001) suggèrent une approche temps-fréquence, où la réponse de la salle est tout d'abord décomposée par bande de fréquences avant d'être inversée. Une fois décomposée par bande de fréquence, on manipule la réponse de la salle pour cette bande. Par exemple, la longueur de la réponse du filtre ainsi que la distribution temporelle de l'énergie dans celle-ci peuvent être ajustées, et cet ajustement ne concerne que la bande dont il est question. Ensuite les différentes bandes de fréquences sont recombinées et la réponse obtenue est inversée et utilisée comme filtre égalisateur. Une telle méthode offre beaucoup de flexibilité et permet de contrer les problèmes de « pré-écho » présent lorsque la pleine réponse impulsionale est inversée. De manière plus générale, la méthode utilisée pour concevoir le filtre égalisateur à partir de la réponse impulsionale de la chaîne audio doit permettre une décomposition par bandes de fréquences, la durée de l'apport du filtre pour chacune de ces bandes de fréquences doit être contrôlable. Plus on monte en fréquence, plus la durée de

l'apport du filtre est courte, cet apport sera concentré dans les premières millisecondes de la réponse impulsionale du filtre. Conséquemment, plus on descend en fréquence, plus la durée de l'apport du filtre est longue, cet apport sera réparti sur la pleine réponse impulsionale du filtre. La fenêtre permettant cette répartition temporelle de l'énergie selon la bande de fréquences est montrée par la figure 4.24.

Le terme *multirésolution* est employé pour décrire une telle approche dans laquelle un signal de la réponse du filtre est décomposé en bandes de fréquences comprenant chacune une portion de longueur variable de la réponse impulsionale totale du filtre.

Dans le cadre de ce travail, c'est la méthode des résidus et les filtres Kaiser qui sont utilisés pour décomposer par bandes la réponse de la chaîne audio. La fenêtre de Kaiser est typiquement utilisée pour la décomposition du spectre par bandes pour sa capacité à bien atténuer les lobes secondaires et réduire au maximum l'effet de Gibbs (Nelson *et al.*, 1972). La méthode des résidus utilisée est documentée dans ce même ouvrage.

On appelle banc de filtres en octave l'ensemble des filtres Kaiser utilisés puisque le facteur par lequel est réduit la fréquence de coupure de deux filtres de bandes voisines est 2. Kaiser est une famille de fenêtres pour laquelle un paramètre nommé *alpha* permet d'ajuster le plancher des lobes secondaires pour bien atténuer les fréquences supérieures à la fréquence de coupure. En revanche, plus les lobes secondaires sont atténués, plus le lobe primaire est large et donc moins la coupure est abrupte. Un plancher de -50 dB est utilisé dans ce travail. Comme on le constate dans la figure 4.23, le point de transition entre deux filtres de bandes consécutives se trouve à -6 dB. On dit qu'une telle décomposition par bandes est une décomposition en octave. Une telle analyse temps-fréquence de la réponse impulsionale $h[n]$ de la chaîne audio permet une meilleure interprétation du comportement de la chaîne audio mais génère en revanche une importante quantité de données qui sont difficiles à traiter en temps réel.

Le premier filtre construit est celui de la bande de fréquences supérieure, c'est un simple passe-haut donné par l'équation de Kaiser paramétrisée de façon à ce que la contrainte de -50 dB pour le plancher des lobes secondaires soit respectée. La réponse impulsionale du filtre obtenue est tronquée. Pour ce premier filtre passe-haut, la coupure a lieu après le 31e échantillon. Ensuite, pour construire le filtre passe-bande de la bande inférieure voisine, on déconvolute le filtre passe-haut précédemment conçu d'une pure impulsion. Le résultat est un filtre passe-bas que l'on nommera *reste*. Ce *reste* est en fait le filtre complémentaire. À *reste* on convolute un second filtre passe-haut ayant maintenant une fréquence de coupure réduite d'un facteur 2 par rapport au passe-haut utilisé pour la bande voisine supérieure. Le résultat de la combinaison du passe-bas et du passe-haut est un filtre passe-bande. L'exercice est répété plusieurs fois en déconvoluant le filtre passe-bande obtenu de ce qu'il *reste* afin d'obtenir un passe-bas qu'on combinera à un passe-haut pour obtenir le prochain passe-bande. À une

itération donnée, si on calcule la convolution de l'ensemble des filtre générés au cours des itérations précédentes avec *reste*, on obtient une parfaite impulsion. Dans le cadre de ce travail, on a choisi de s'arrêter à 10 bandes, la dernière étant tout simplement un filtre passe-bas, soit le *reste* de l'avant dernière itération. En utilisant à chaque fois le *reste* pour concevoir le filtre de la bande voisine inférieure, on s'assure que l'erreur introduite lorsqu'on tronque l'équation de Kaiser ne s'accumule pas de bande en bande. À chaque fois, le passe-haut ayant une fréquence de coupure réduite par un facteur 2 comparativement à son voisin est conçu à l'aide de l'équation de Kaiser. Pour réduire la charge de calcul, on exploite la propriété voulant que plus la fréquence augmente, plus le filtre utilisé pour isoler cette bande peut être court (Sarris et Cambourakis, 2003). Cette propriété découle du fait que la résolution temporelle et spectrale de l'oreille humaine varie en fonction de la fréquence (Plomp, 1964).

Le tableau 5.3 montre la longueur des 10 filtres utilisés dans ce travail ainsi que les points de transition entre bandes consécutives.

Tableau 4.1 Longueur et point de transition de chaque bande

	Type	Longueur (éch)	Longueur effective (s)	f_{bas}	f_{haut}
J	Passe-bas	8081	0,0916		41,25
I	Passe-bande	8081	0,0916	41,25	77,12
H	Passe-bande	4041	0,0458	77,12	149,0
G	Passe-bande	2021	0,0229	149,0	292,7
F	Passe-bande	1011	0,0115	292,7	580,6
E	Passe-bande	505	0,0057	580,6	1158
D	Passe-bande	253	0,0029	1158	2318
C	Passe-bande	127	0,0014	2318	4663
B	Passe-bande	63	0,0007	4663	9454
A	Passe-haut	31	0,0004	9454	

On remarque à chaque fois que le nombre d'échantillons est impair. On tronque volontairement la réponse impulsionnelle des filtres de Kaiser à des nombres d'échantillons impairs afin de conserver la symétrie de celle-ci. C'est ce qui garantit que chaque filtre a une phase nulle et qu'un signal traversant ces 10 filtres peut être parfaitement reconstruit en sommant la sortie de chacun. La figure 4.23 montre la réponse impulsionnelle de chacun des 10 filtres. Les figures 4.22 et 4.23 montrent la réponse impulsionnelle et la réponse en fréquence de chacun des 10 filtres respectivement.

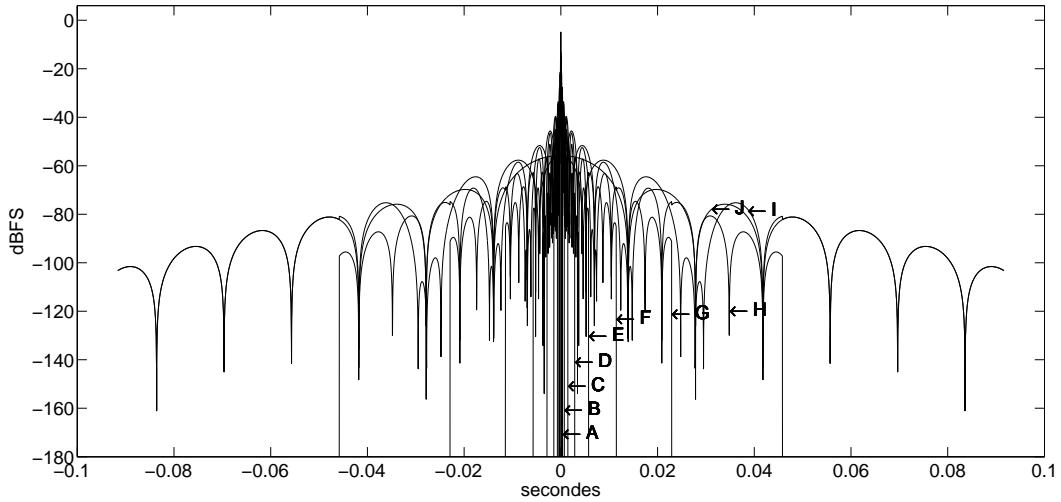


Figure 4.22 Réponse impulsionnelle de la banque de filtres en octave

La conception du filtre égalisateur passe par la convolution de la réponse impulsionnelle de la chaîne audio $h[n]$ par la réponse impulsionnelle de chacun des filtres de la banque montrée par la figure 4.22. On utilise cette fois la convolution linéaire directement. Au résultat de chacune de ces convolutions, on applique la fenêtre temporelle assurant la répartition temporelle de l'énergie selon la bande de fréquences montrée par la figure 4.24. On multiplie point par point cette fenêtre à la réponse impulsionnelle pour une bande donnée.

En regardant attentivement les colonnes « Longueur (éch) » et « Longueur effective (s) » du tableau 4.1, on se rend compte qu'un facteur 2 existe entre la longueur du filtre permettant de décomposer le signal par bandes mesurées en échantillons et la longueur effective en secondes qu'aura cette bande. La figure 4.25 facilite la compréhension de ce phénomène. Pour la bande G montrée par la figure couvrant les fréquences de 149,0 à 292,7 Hz, par exemple, la réponse du filtre par laquelle est convolue la réponse impulsionnelle de la salle pour l'obtention de la réponse impulsionnelle de la salle pour cette bande de fréquences est d'une longueur de 2021 échantillons, soit 0,0458 seconde. Mais la réponse de ce filtre est symétrique. Cependant, la réponse de la salle ne l'est pas. Elle ressemble plutôt à une première impulsion représentant le signal direct en provenance du haut-parleur et captée par le microphone à la position d'écoute, suivie de plusieurs images de cette première impulsion représentant le champ réverbéré résultant de réflexions de diverses réflexions du signal sur les murs de la salle avant d'atteindre la position d'écoute. Une fois la réponse de la salle convolue à celle du filtre de la bande de G, la première impulsion de la réponse de la salle, celle du champ direct, est décalée de 1010 échantillons, soit 0,0229 s. Ce décalage de 1010

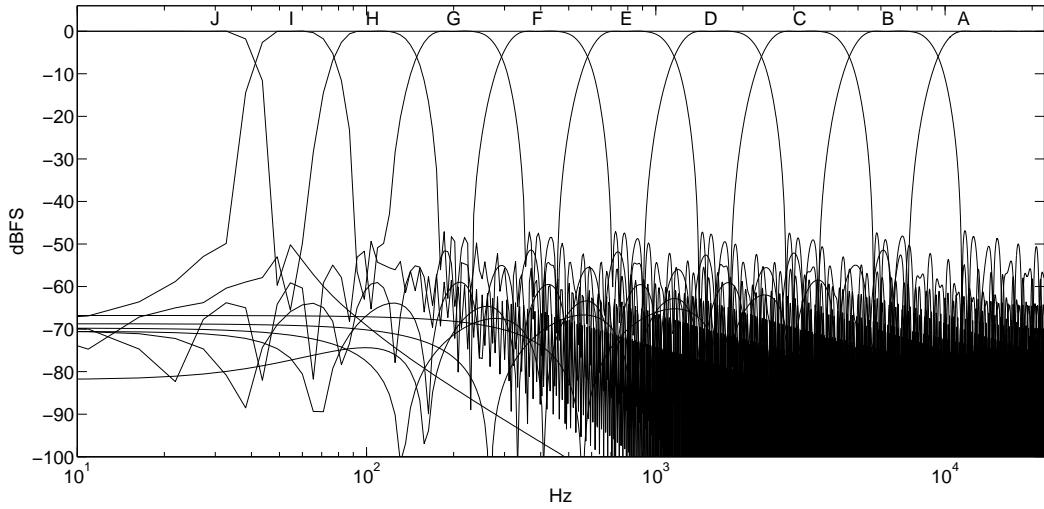


Figure 4.23 Réponse en fréquence de la banque de filtres en octave

échantillons est introduit par la convolution de la réponse impulsionnelle symétrique du filtre de la bande de G centrée à l'échantillon 1011. Pour connaître le décalage introduit par la convolution d'un filtre à réponse symétrique, on soustrait 1 de la position correspondant au centre de cette symétrique. Dans le cas d'un filtre symétrique de longueur 3 avec symétrie par rapport au deuxième échantillon, le décalage du résultat de la convolution serait de 1. Dans le cas d'un filtre de 2021 échantillons avec symétrie par rapport à l'échantillon 1011, le décalage observé est de 1010.

Pour la partie qui nous intéresse dans le cas de la bande G, soit les 1010 premiers échantillons de la réponse de la salle, les échantillons se retrouvent eux aussi décalés, on les retrouve entre les échantillons 1011 et 2021. À l'échantillon 2021 se termine ce qu'on considère comme étant le champ direct et débute ce qu'on considère comme étant le champ réverbéré. Pour la bande G, cet échantillon est celui où la fenêtre appliquée entame sa coupure. Ce temps de coupure est donné en nombre d'échantillons dans la colonne « Longueur » du tableau 4.1, et noté T_c (en ms) dans la figure 4.26. La longueur efficace pour cette bande G est donc de 0,0229 s, soit la moitié de la longueur du filtre pour cette bande. C'est d'ailleurs le cas pour toutes les autres bandes. Cette longueur efficace est comprise entre $T_c/2$ et T_c . Pour la bande J, celle la plus longue et couvrant les fréquences les plus basses, cette longueur efficace est de 91,6 millisecondes, soit presque 100 ms, la longueur maximale sur laquelle on doit tenter d'égaliser la réponse impulsionnelle de la salle (Mosayyebpour *et al.*, 2013). Selon Mosayyebpour *et al.* (2013), le cerveau perçoit les 50 à 100 premières millisecondes de la réponse impulsionnelle d'une salle dite petite comme faisant partie du champ direct, le reste

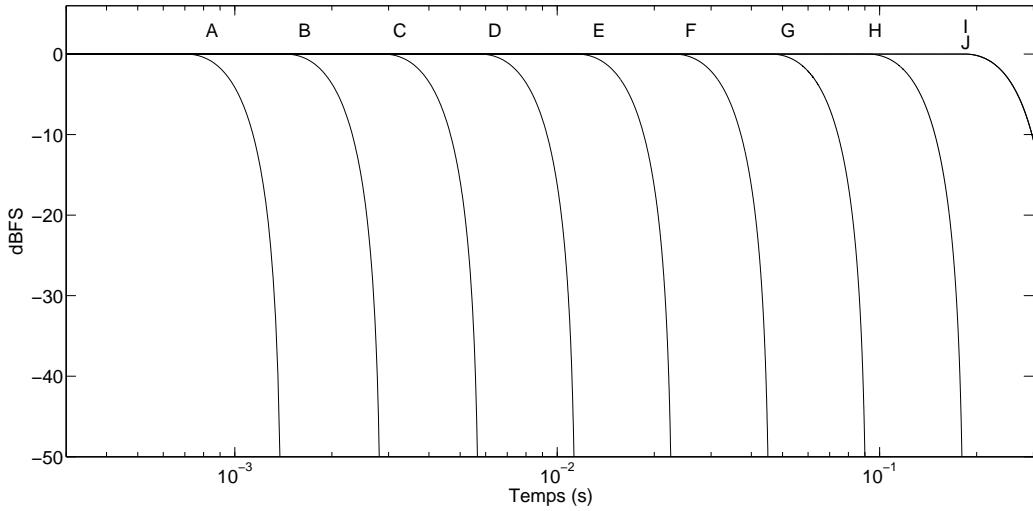


Figure 4.24 Fenêtres temporelles selon la bande de fréquences

comme faisant partie du champs réverbéré. Ils expliquent aussi pourquoi seul le direct doit faire l'objet d'égalisation.

Les convolutions fenêtrées résultantes sont réalignées et sommées en une seule réponse impulsionnelle $g[n]$. C'est cette réponse impulsionnelle $g[n]$ qu'on va inverser dans le but d'obtenir la réponse du filtre égalisateur. Un passage vers le domaine des fréquences est de mise. Par la suite, chaque échantillon complexe de la réponse en fréquence est inversé, ce qui permet d'obtenir la réponse en fréquence du filtre égalisateur.

C'est à cette étape qu'on normalise l'amplitude absolue de la réponse en fréquence du filtre pour s'assurer qu'on ne tente pas de corriger le gain en boucle ouverte du système. Par exemple, si le contrôle de l'amplificateur de la chaîne audio montré dans la figure 1.5 est altéré par l'utilisateur du système durant l'écoute, on ne veut pas que le filtre adaptatif compense pour cette atténuation en augmentant toutes les fréquences. On peut laisser la liberté à l'utilisateur d'augmenter ou de diminuer le gain en transmission de la chaîne sans que le filtre adaptatif tente de corriger ce changement de gain. Pour ce faire, la réponse en fréquence du filtre égalisateur est normalisée par rapport à un paramètre A_{max} . En attribuant une valeur de 1 à A_{max} , le filtre fonctionne donc strictement en atténuant les fréquences favorisées par la salle. Le résultat est un signal audio filtré de faible puissance. S'il y a un écart de 30 dB entre le pic le plus important de la réponse en fréquence du filtre et sa moyenne, alors l'ensemble de la réponse en fréquence se voit retrancher 30 dB à cause de ce pic. Par conséquent, une valeur de A_{max} supérieure à 1 peut être sélectionnée, de manière à gagner de la puissance, au risque d'avoir de la saturation dans le signal si la valeur d'échantillons mise en sortie sort

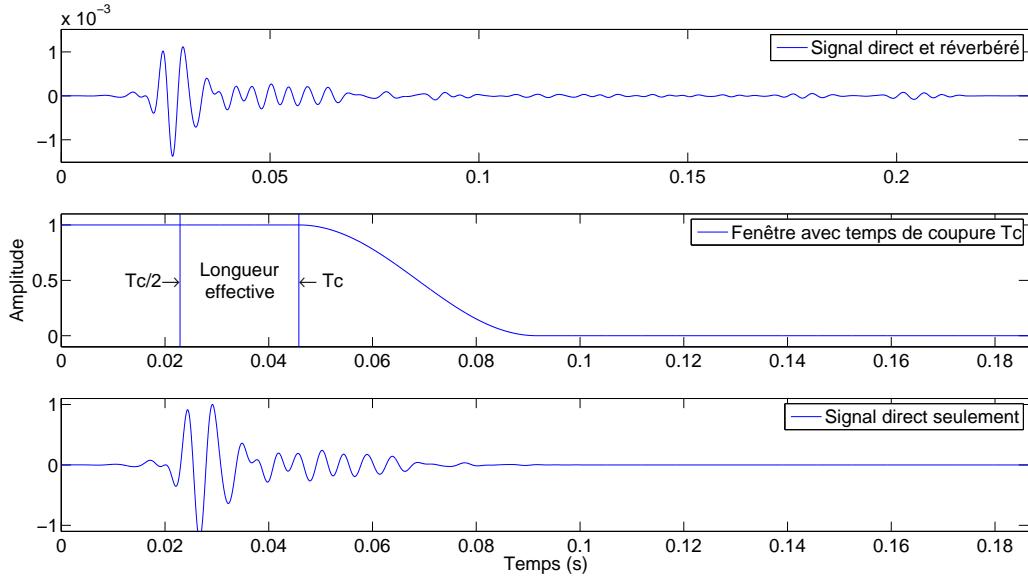


Figure 4.25 Temps de coupure de la fenêtre appliqué à une bande et longueur effective de cette bande

de la plage $[-1, 1]$.

Finalement, une transformée de Fourier discrète inverse est appliquée pour obtenir la réponse impulsionnelle du filtre égalisateur. Le filtre convolutif décrit dans la section 4.9 utilise cette réponse impulsionnelle $g[n]$ pour la convolution avec le signal audionumérique pour le filtrer avant de l'acheminer vers le convertisseur numérique-analogique de la chaîne audio.

La figure 4.26 résume l'ensemble des opérations décrites dans cette section permettant d'obtenir la réponse impulsionnelle du filtre égalisateur $g[n]$ à partir de la réponse impulsionnelle de la chaîne audio $h[n]$.

4.9 Fonctionnement du filtre convolutif

Le filtre égalisateur implémenté doit être capable de traiter un flux constant d'échantillons de musique formant un signal numérique continu. Le schéma bloc suivant représente le fonctionnement interne du filtre égalisateur adaptatif de la figure 1.5. Cette section contient le détail de l'implémentation d'un tel filtre égalisateur reposant notamment sur le théorème de la convolution.

Tout d'abord, le filtre doit donc être numérique, le choix entre l'une des deux grandes familles de filtre numérique s'impose : Filtre à réponse impulsionnelle infinie (RII) ou Filtre à réponse impulsionnelle finie (RIF).

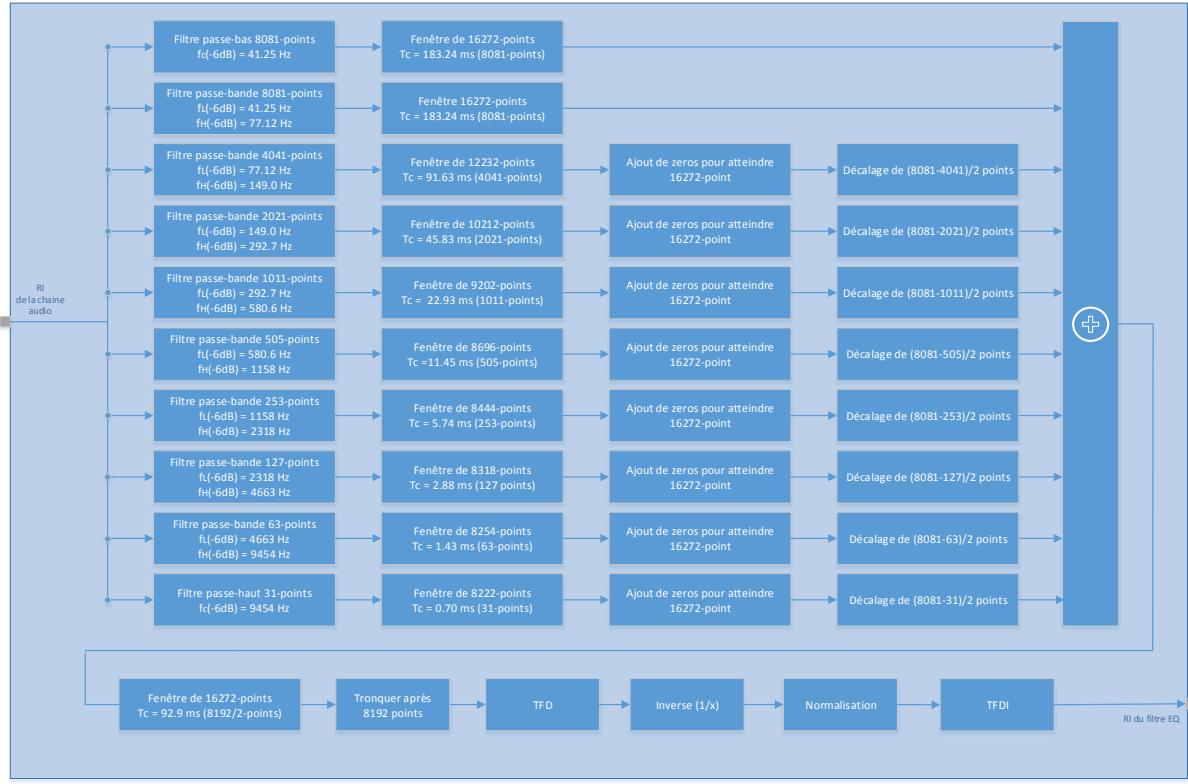


Figure 4.26 Implémentation de la conception du filtre égalisateur

Le principal avantage des filtres à réponse impulsionnelle infinie est leur capacité à répondre à des contraintes sévères en termes de facteur de qualité (facteur-Q), de l'amplitude des lobes secondaires et de la pente du filtre, c'est-à-dire que la transition entre la bande passante et la bande rejetée est aussi rapide que dans le cas du filtre RIF. Le filtre RII peut rencontrer les contraintes de conception du filtre tout en alliant une longueur N inférieure à celle qui serait nécessaire au filtre RIF pour répondre aux mêmes exigences. Un filtre plus court permet d'économiser en temps de calcul et minimise le délai entre l'entrée du filtre et sa sortie. Lorsqu'on parle de taille, on fait référence au nombre de coefficients a et b dans les figures ci-dessous, où z^{-1} représente un délai d'une période d'échantillonnage. Dans le cas d'un filtre RII, la réponse impulsionnelle $g[n]$ est une somme infinie d'exponentielles ou d'exponentielles pondérées. C'est pour cette raison qu'on les appelle filtre à réponse impulsionnelle infinie.

Par opposition, la réponse impulsionnelle d'un RIF fini est d'une durée finie. On les appelle aussi filtre non récursif ou encore filtre tout-zéro. Dans la fonction de transfert $G(z)$ se présentant sous forme d'une fraction, on nomme zéros les racines du numérateur et celles

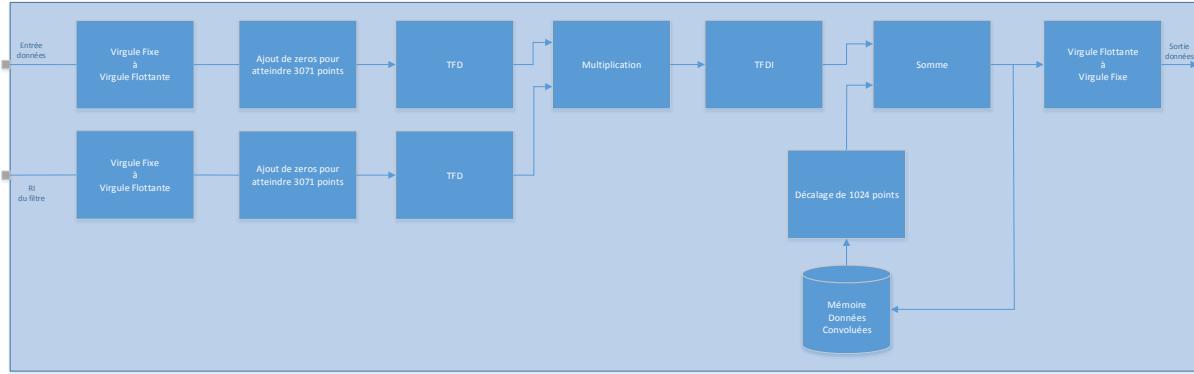


Figure 4.27 Implémentation du filtre convolutif

du dénominateur. On les appelle les filtres tout-zéro puisque la fonction de transfert $G(z)$ d'un filtre RIF n'a que des termes au numérateur, à l'exception de multiples pôles à l'origine lorsqu'on trace ceux-ci dans un plan complexe. Les filtres à réponse impulsionnelle infinie sont typiquement plus faciles à concevoir. La réponse d'un RIF est souvent une troncature de la réponse d'un RII. Cela est particulièrement vrai lorsque l'obligation n'est pas l'un des cas habituels (passe-haut, passe-bas, entaille, etc.) qui ont été étudiés et optimisés pour les filtres analogiques.

Aussi, les filtres RIF peuvent être facilement conçus pour avoir une phase linéaire, une propriété qui n'est pas facilement satisfaite en utilisant des filtres RII. Dans notre cas, on profite d'une des propriétés des RIF permettant de les concevoir directement à partir d'une réponse en fréquence donnée. Pour une réponse en fréquence donnée dans le domaine de Fourier discret, soit une séquence de nombres complexes d'une longueur finie et ayant une symétrie complexe conjuguée par rapport à son centre, la transformée de Fourier discrète inverse mène directement à la réponse impulsionnelle d'un filtre RIF. La figure 4.28 montre le chemin emprunté par le signal entre l'entrée et la sortie dans un filtre RIF.

Dans l'équation liant la sortie $u[n]$ d'un RIF à son entrée $r[n]$, chacun des coefficients B_n du filtre est multiplié par les N derniers échantillons du signal en entrée et les N résultats sont sommés pour obtenir la valeur d'un échantillon en sortie :

$$u[n] = b_0 \cdot r[n] + b_1 \cdot r[n - 1] + b_2 \cdot r[n - 2] + \dots + b_N \cdot r[n - N] \quad (4.24)$$

$$u[n] = \sum_{i=0}^N b_i \cdot r[n - i] \quad (4.25)$$

Cette équation s'apparente à celle de la convolution discrète sous sa forme générale :

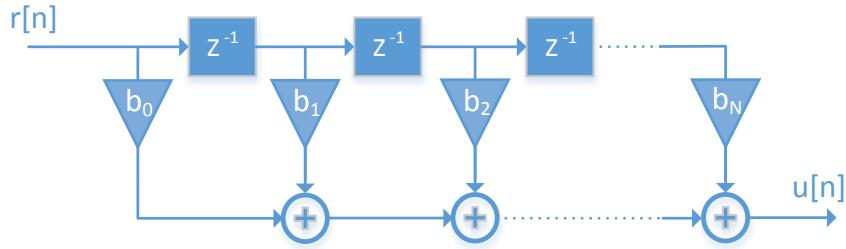


Figure 4.28 Filtre à réponse impulsionnelle finie

$$(r * g)[n] \stackrel{\text{def}}{=} \sum_{i=-\infty}^{\infty} h_i \cdot r[n-i] \quad (4.26)$$

Le fait qu'on s'intéresse, dans le cas d'un RIF, à des séquences de longueurs finies N permet de limiter la portée de la sommation.

$$(r * g)[n] = \sum_{i=0}^N g_i \cdot r[n-i] \quad (4.27)$$

Cette équation est équivalente à celle de la convolution d'une entrée $r[n]$ de longueur finie N à la réponse impulsionnelle $g[n]$ du filtre, elle aussi de longueur finie N . Les coefficients b_n correspondent directement à la réponse impulsionnelle du filtre.

Il est donc possible d'appliquer une transformée de Fourier inverse à la réponse en fréquence désirée pour obtenir la réponse à l'impulsion correspondante, et par le fait même les coefficients b_n . Cette propriété des RIF permettant l'obtention directe de leur coefficient à partir de leur réponse en fréquence est un critère important dans le cadre de ce travail, car le passage par le domaine des fréquences s'impose, le RIF est donc le type de filtre retenu.

Puisque le calcul de la sortie pour un RIF consiste à calculer la convolution entre le signal $r[n]$ en entrée et la réponse impulsionnelle $g[n]$, N multiplication et $N - 1$ additions sont nécessaires pour le calcul de chaque échantillon en sortie pour un filtre dont la réponse à l'impulsion est de longueur N . Avec la disponibilité grandissante de librairies permettant d'effectuer optimalement des transformées de Fourier rapides sur des processeurs multiusages tels que ceux se retrouvant dans nos ordinateurs personnels, de nouvelles techniques ont vu le jour, permettant de calculer plus efficacement la convolution linéaire en exploitant habilement les propriétés liant la convolution circulaire dans la domaine temporel à la multiplication dans le domaine fréquentiel. La technique « overlap-add » illustrée par la figure 4.29 est très

répandue. Dans le cadre de ce travail, on utilise une implémentation directe de la technique « overlap-add » telle que décrite dans Stockham (1966).

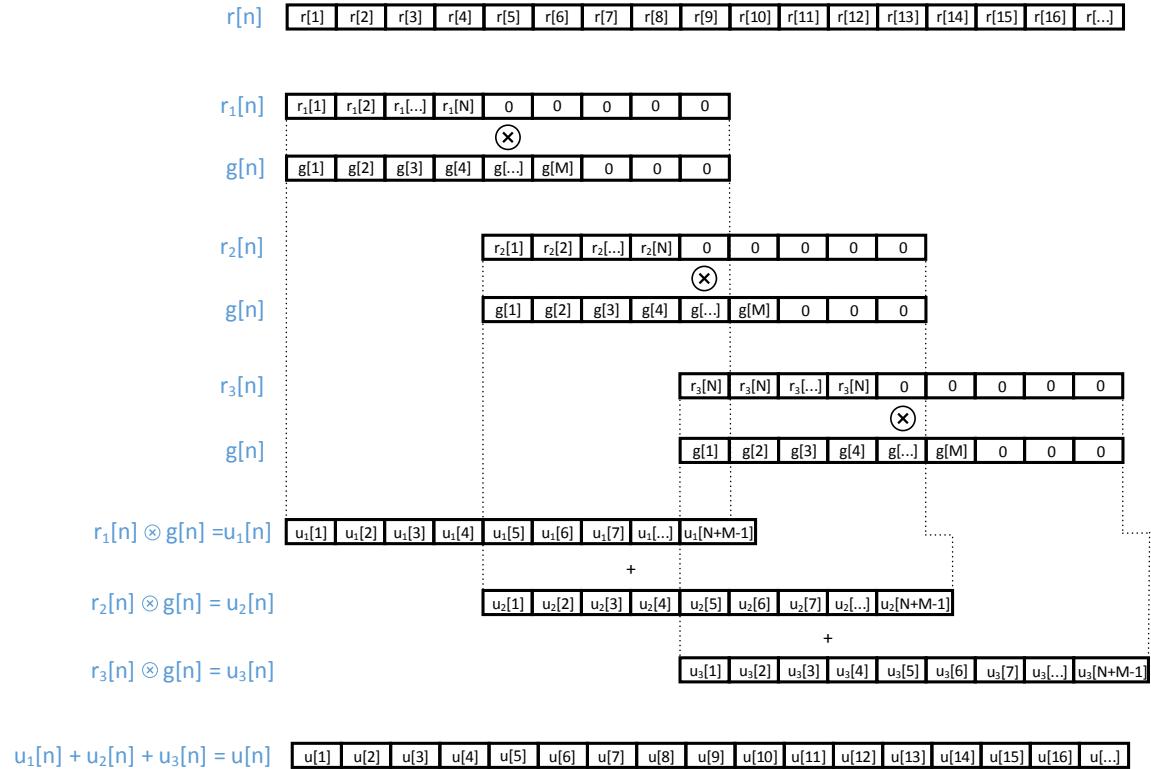


Figure 4.29 Méthode « Overlap-Add » pour la convolution linéaire

La méthode de overlap-add (OA) est un moyen efficace pour évaluer la convolution discrète d'un long signal $x[n]$, voire d'une longueur inconnue, avec la réponse impulsionnelle $g[n]$ d'un filtre de longueur finie (RIF).

$$(r * g)[n] \stackrel{\text{def}}{=} \sum_{i=-\infty}^{\infty} g_i \cdot r[n - i] = \sum_{i=0}^N g_i \cdot r[n - i] \quad (4.28)$$

Le concept repose sur la décomposition de la séquence $r[n]$ en plusieurs petites sous-séquences $r_i[n]$ de 1024 échantillons de longueur dans notre cas :

$$r_i[n] \stackrel{\text{def}}{=} \begin{cases} r[n + k \cdot L] & n = 1, 2, \dots, L \\ 0 & \text{sinon} \end{cases} \quad (4.29)$$

C'est cette technique qui est utilisée dans le cadre de ce travail pour convoluer en temps réel le flux de données audio décomposé en blocs de 1024 échantillons à la réponse impul-

sionnelle du filtre égalisateur de 8192 échantillons de longueur. Le flux de données résultant ne contient pas de discontinuité et est équivalent au produit d'un RIF tel que montré dans la figure 4.28. La figure 4.30 montre l'insertion du filtre $g[n]$ dans la chaîne audio. La partie de droite de cette figure montre que le signal de référence $r[n]$ ainsi que le signal capté à la position d'écoute $y[n]$ sont passés dans le domaine fréquentiel à l'aide de la transformée de Fourier discrète. C'est dans le domaine fréquentiel que l'erreur entre ces deux signaux est calculée. La métrique d'erreur utilisée est décrite plus en détail dans la section 5.1 mais commence par la soustraction du spectre fréquentiel des deux signaux comme le suggère la figure 4.30.

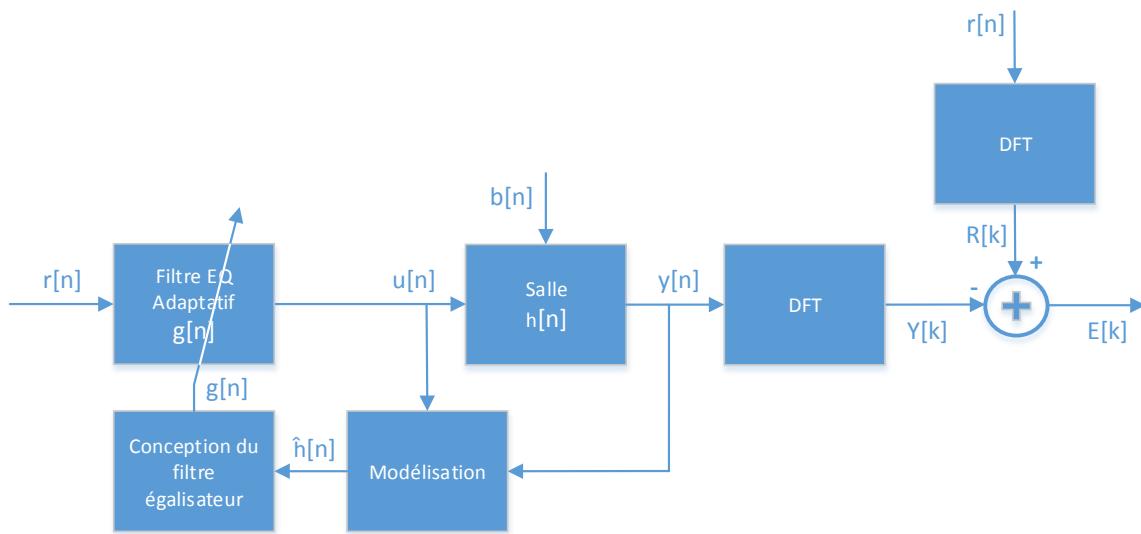


Figure 4.30 Chaîne audio avec filtre adaptatif pour l'égalisation des fréquences

Ce chapitre couvrait en détail l'algorithme suggéré pour la modélisation et la correction l'acoustique d'une salle à l'insu de l'auditoire. Le prochain chapitre présente l'implémentation logicielle de cet algorithme ainsi que les résultats expérimentaux obtenus dans le but de quantifier la performance.

CHAPITRE 5

IMPLÉMENTATION LOGICIELLE ET RÉSULTATS EXPÉRIMENTAUX

Dans le but de générer des résultats expérimentaux et de caractériser la performance de la solution suggérée, un montage expérimental est réalisé. Chaque élément de la chaîne audio de la figure 1.5 doit y être présent, notamment un haut-parleur amplifié, la salle dans laquelle a lieu l'écoute, le microphone servant à la mesure, et un périphérique audio assurant la conversion analogique-numérique et numérique-analogique ainsi que l'interface avec l'ordinateur contenant l'algorithme implémenté.

La section 5.1 du présent chapitre contient tout d'abord une description du critère de performance utilisé pour quantifier la performance de l'implémentation logicielle de l'algorithme de modélisation et la correction de la réponse en fréquence d'une chaîne audio présentée dans le chapitre 4. Naturellement, cette implémentation logicielle sera couverte en détail. On commence, dans la section 5.2, par l'implémentation Matlab partielle de l'algorithme ayant permis de valider l'algorithme en début de projet. On verra ensuite dans la section 5.3 l'implémentation finale en langage C++ de l'ensemble des composants de l'algorithme. L'algorithme final fut encapsulé dans un plugin VST, un standard de l'industrie permettant son utilisation dans une vaste sélection de logiciels voués à l'audionumérique. L'encapsulation du code est présentée dans la section 5.3. Ce chapitre décrit également le montage expérimental réalisé ainsi que les spécifications du matériel utilisé. Dans la section 5.8, on voit comment la méthode standard du balayage sinusoïdal de la plage de fréquence est utilisée pour générer la réponse de la salle qu'on utilisera par la suite comme référence. On y compare, dans la section 5.9, la réponse en fréquence obtenue en 5.8 à l'aide du nouvel algorithme suggéré en utilisant le critère de performance pré-établi. On y voit comment l'erreur décroît à mesure que la musique défile. La section 5.10 montre comment la réponse en fréquence obtenue par le nouvel algorithme suggéré est utilisée pour mettre périodiquement à jour les coefficients du filtre égaliseur adaptatif. Finalement, on y voit comment ce filtre améliore la réponse en fréquence de la chaîne audio. Le même critère de performance est utilisé pour quantifier cette amélioration.

5.1 Définition du critère de performance

Aucune procédure de modélisation ne serait complète sans la définition de son critère de performance. Ces indices, souvent calculés en comparant des résultats numériques issus d'un

modèle discret aux résultats réciproques retournés par le système réel, servent, entre autres, à évaluer la performance du modèle et à comparer les diverses techniques de modélisation. L'absence de travaux montrant explicitement des mesures quantitatives de la performance d'un modèle acoustique entrée-sortie limite l'utilisation des critères de performance à la comparaison des méthodes étudiées dans ce travail.

Les critères de performance seront calculés en comparant la réponse du modèle à celle de la salle à certains signaux d'excitation couvrant la plage de fréquences considérée. Ces signaux d'excitation sont de même nature que ceux exploités par deux des techniques de modélisation entrée-sortie présentées. Le critère de performance choisi est souvent utilisé pour quantifier la performance de la réponse en fréquence de modèles entrée-sortie. Ces quatre critères serviront par la suite à comparer la performance de différentes méthodes de modélisation pour parvenir à l'adoption du modèle le plus performant. Canclini *et al.* (2012) suggèrent l'utilisation de l'erreur quadratique moyenne (EQM) comme métrique.

L'erreur quadratique moyenne sert de métrique d'erreur et montre que plus la musique défile, plus la réponse en fréquence obtenue s'apparente à la référence pour un même point d'écoute. Les résultats obtenus montrent qu'il est bel et bien possible d'obtenir la réponse impulsionale et la réponse en fréquence d'une salle en pondérant la réponse de celle-ci à un signal musical qui défile. L'équation 5.1 définit l'erreur quadratique moyenne. À noter que l'index i est utilisé pour itérer sur l'ensembles des fréquences $j\omega$ comprises entre 0 et $f_{éch}/2$, discrétisées en $N/2 + 1$ points.

$$\text{EQM} = \frac{1}{N/2 + 1} \sum_{i=1}^{N/2+1} ((H_{ref}[i] - H[i])^2) \cdot W[i] \quad (5.1)$$

où $N/2 + 1$ est le nombre d'échantillons dans le domaine fréquentiel et $W[i]$ est un vecteur de poids décroissant exponentiellement en fonction de la fréquence discrète i de façon à ce que la contribution à l'erreur diminue à mesure que la fréquence augmente. Si une telle pondération n'était pas utilisée, la moitié de l'erreur proviendrait des fréquences comprises entre $f_{éch}/4$ (11 025 Hz) et $f_{éch}/2$ (22 050 Hz), et serait mal adaptée au comportement de l'oreil humaine. Ce choix est justifié par les mêmes raisons évoquées à la section 4.8 pour justifier la décomposition de la réponse impulsionale de la salle en bandes d'octave, bandes dont la largeur varie en fonction de la fréquence. Le taux de décroissance exponentiel en fonction de la fréquence pour ce vecteur de poids W est choisi de façon à ce que la somme des poids soit égale à 1.

Puisque le but de l'ajout du filtre est d'égaliser la réponse en fréquence, la performance de ce dernier peut être obtenue par la mesure de la similarité entre la réponse en fréquence corrigée et une réponse en fréquence dite idéale. La réponse en fréquence idéale a un gain

constant pour toute la plage de fréquences considérée. Le critère de performance utilisé est encore une fois l'erreur quadratique moyenne donnée par l'équation 5.2.

$$\text{EQM} = \frac{1}{N/2 + 1} \sum_{i=1}^{N/2+1} ((H_{idéale}[i] - H[i])^2) \cdot W[i] \quad (5.2)$$

où la réponse idéale parfaitement plate $H_{idéale}[i]$ est égale à la moyenne arithmétique des échantillons de $H[i]$, et donc parfaitement plate.

5.2 Prototypage logiciel de l'algorithme

L'un des objectifs du projet étant la familiarisation avec les concepts plus avancés en programmation orientée objet, une implémentation logicielle de l'algorithme de filtrage adaptatif suggéré est de mise.

En début de projet, l'implémentation logicielle avait principalement pour but de valider le fonctionnement de l'algorithme suggéré. La méthode suggérée dans ce travail se divise en deux parties. Premièrement, l'algorithme doit avoir la capacité d'évaluer la réponse impulsionale et la réponse en fréquence de la chaîne audio à partir de signaux musicaux. Deuxièmement, l'algorithme doit assurer le filtrage du signal audio avant qu'il n'atteigne la chaîne audio dans le but de compenser l'effet indésirable de cette dernière sur l'écoute à une position donnée. L'apport en innovation dans ce travail se situe principalement dans la première de ces deux parties. La méthode suggérée pour obtenir la réponse de la salle strictement à partir de signaux musicaux est nouvelle et devait être validée en début de projet. La portion filtrage s'inspire de techniques bien connues et maintes fois implémentées et comportait beaucoup moins de risque.

L'objectif était en début de projet de prototyper rapidement l'algorithme de modélisation suggéré pour en valider la capacité à estimer la réponse de la chaîne audio aussi bien que les méthodes classiques sont capable de le faire. Tel qu'expliqué dans la section 4.4, la réponse impulsionale ainsi que la réponse en fréquence de la chaîne audio à l'étude pouvaient être préalablement obtenues en utilisant une méthode expérimentale standard telle que le balayage sinusoïdal de la plage de fréquence. Les résultats obtenus à l'aide de la méthode suggérée pouvaient être comparés à ceux de références obtenus à l'aide de la méthode standard. Cet exercice permet une validation de la méthode suggérée sans pour autant demander une implémentation finale et performante de l'algorithme.

Le langage de calcul scientifique Matlab fut utilisé afin d'implémenter le concept de base de l'algorithme suggéré pour des fins de validation en début de projet. On voulait notamment valider qu'il était bel et bien possible de reconstruire par moyenne mobile pondérée la réponse en fréquence de la chaîne audio pour un signal musical capté par un microphone placé à la

position d'écoute que l'on décompose en bloc et qu'on analyse individuellement à mesure que la musique défile.

Il était nécessaire pour cet exercice d'avoir un moyen de communiquer, à partir de l'environnement de développement Matlab, avec un périphérique audio pour PC assurant notamment la conversion analogique-numérique d'un signal provenant d'un microphone ainsi que la conversion numérique-analogique d'un signal destiné au haut-parleur. Les classes audioRecorder et audioPlayer de la librairie de base de Matlab permettent l'accès au périphérique audio de l'ordinateur. La classe audioRecorder crée un objet permettant l'accès au périphérique en lecture via l'interface de son du système d'exploitation Windows, en l'occurrence MME, DirectSound, ou WASAPI. La classe audioPlayer permet l'accès en écriture au périphérique audio de la même façon. Les classes dsp.AudioPlayer et dsp.AudioRecorder permettaient respectivement l'écriture directe et la lecture directement dans les mémoires tampons d'un périphérique audio pour PC et d'utiliser un pilote ASIO permettant de ne passer par l'interface de son du système d'exploitation Windows. Le but d'une telle approche est de pouvoir opérer avec une latence plus faible. La latence n'étant pas critique à cette étape de validation, les classes audioPlayer et audioRecorder de la librairie de base de Matlab furent utilisées pour communiquer avec un périphérique audio. La figure 5.1 montre le canevas adapté à la modélisation d'une chaîne audio que procure l'utilisation de ces 2 classes, l'audio joué à travers le haut-parleur et celui capté par le microphone placé dans l'environnement d'écoute sont manipulables en temps réel dans l'environnement de développement Matlab.

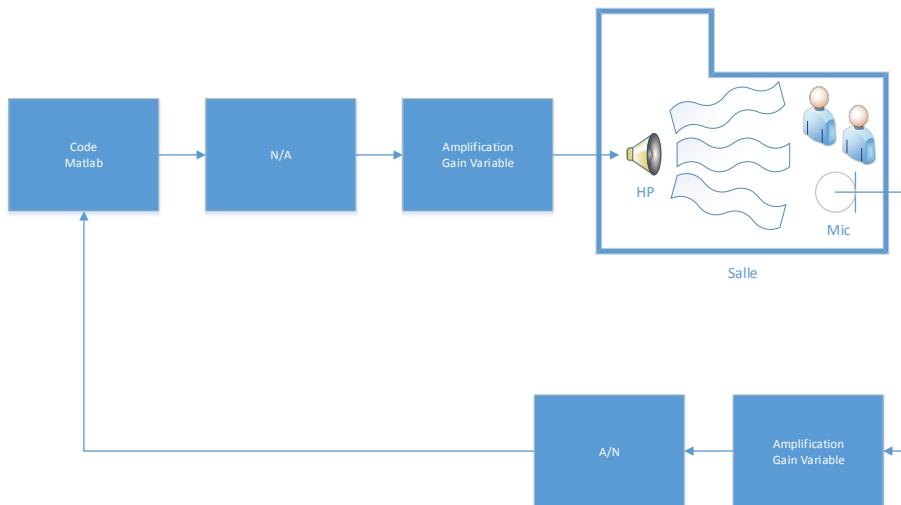


Figure 5.1 Signaux audio dans l'environnement Matlab

Avec l'accès aux signaux audio dans l'environnement Matlab, la réalisation d'une implé-

mentation logicielle prototype permettant de valider la méthode suggérée se résume à une implémentation de l'algorithme de modélisation montré dans la figure 4.12 de la section 4.5.

Selon l'algorithme tel qu'il est décrit dans la section 4.5, le signal audio envoyé dans la chaîne audio et celui en provenance du microphone sont tous deux décomposés en blocs de 1024 échantillons, ce sont ces vecteurs de 1024 échantillons que l'algorithme prend comme entrées sur les ports « émis » et « reçu » de la figure 4.12. Matlab est très bien adapté au traitement de données numériques sous formes de vecteurs ou de matrice.

Les classes audioRecorder et audioPlayer permettent de recevoir directement les échantillons audio en virgule flottante sur une plage normalisée entre -1 et 1. La conversion de virgule fixe à virgule flottante de la figure 4.12 n'a donc pas besoin d'être programmée. L'accumulation et la division sont assurées par des opérateurs de base en Matlab. L'implémentation de la moyenne mobile exponentielle pondérée notée « MMEP » dans la figure 4.12 et décrite plus en détail dans la figure 4.20 est assurée elle aussi par des opérateurs de base en Matlab, notamment la division, la multiplication, la somme appliquée à des vecteurs de nombres à virgules flottantes.

De l'ensemble des blocs des figures 4.12 et 4.20 décrivant l'implémentation de l'algorithme de modélisation suggéré, les fonctions qui attirent tout particulièrement notre attention sont celles de la TFD et de la TFDI. Ces deux fonctions sont disponibles dans la librairie de base, soit FFT et IFFT respectivement. Les fonctions FFT (FFT, fft2, ifft, ifft2, ifftn) de Matlab sont basées sur une bibliothèque appelée FFTW. Pour calculer une TFD de N points, lorsque N n'est pas un nombre premier ($N = N_1 N_2$), la bibliothèque FFTW décompose le problème en utilisant l'algorithme de Cooley-Tukey. Lorsque N est un nombre premier, la bibliothèque FFTW le décompose en une série de problèmes pour lesquels N n'est pas premier à l'aide de l'algorithme de Rader. Il utilise ensuite la décomposition Cooley-Tukey décrite ci-dessus pour calculer la TFD.

Le temps d'exécution de FFT dépend de la longueur de la transformée. Il est plus rapide pour les puissances de deux. Dans notre cas, 64 blocs de 1024 échantillons sont accumulés et la fonction Matlab FFT est appellée pour calculer la TFD d'un vecteur de 8192 échantillons réels. La fonction Matlab IFFT est utilisée afin de calculer la TFDI d'un vecteur symétrique de 4097 nombres complexes qui donnent un vecteur de 8192 nombres réels en sortie. Les tailles utilisées sont donc toutes des puissances de deux.

On a mentionné plus tôt que l'utilisation de Matlab était rendue possible en raison du fait que la latence et la performance n'étaient pas des enjeux critiques pour cette implémentation prototype de l'algorithme de modélisation. Pourtant, la déconvolution circulaire montrée par la figure 4.11 laisse sous-entendre que le temps de calcul est critique puisque les files de données $u[n]$ et $y[n]$ y sont traitées de manière continue, c'est-à-dire que le calcul de $h_i[n]$

à partir de $u_i[n]$ et $y_i[n]$ doit être complet avant que les données $u_{i+1}[n]$ et $y_{i+1}[n]$ soient disponibles pour le calcul suivant, soit celui de $h_{i+1}[n]$. On peut altérer légèrement cette figure pour qu'elle reflète le séquencement des opérations telles qu'implémentées en Matlab de façon à ce que le temps de calcul ne soit pas critique. Pour ce faire, deux conditions doivent être rencontrées. Premièrement, la continuité n'est pas essentielle entre les résultats individuels $h_{i+1}[n]$ générés par le calcul. C'est le cas pour l'algorithme de modélisation puisque chaque $h_{i+1}[n]$ représente une réponse impulsionale valide et distincte calculée à partir d'un bloc de données. Deuxièmement, on doit admettre que des échantillons de données des $u[n]$ et $y[n]$ ne soient pas traités. C'est également le cas pour notre algorithme de modélisation. La figure 5.2 montre le séquencement des opérations dans l'implémentation Matlab. C'est le temps de calcul qui est maître, les blocs d'échantillons traités ainsi que la fréquence à laquelle de nouvelles réponses $h_{i+1}[n]$ sont générées s'adaptent.

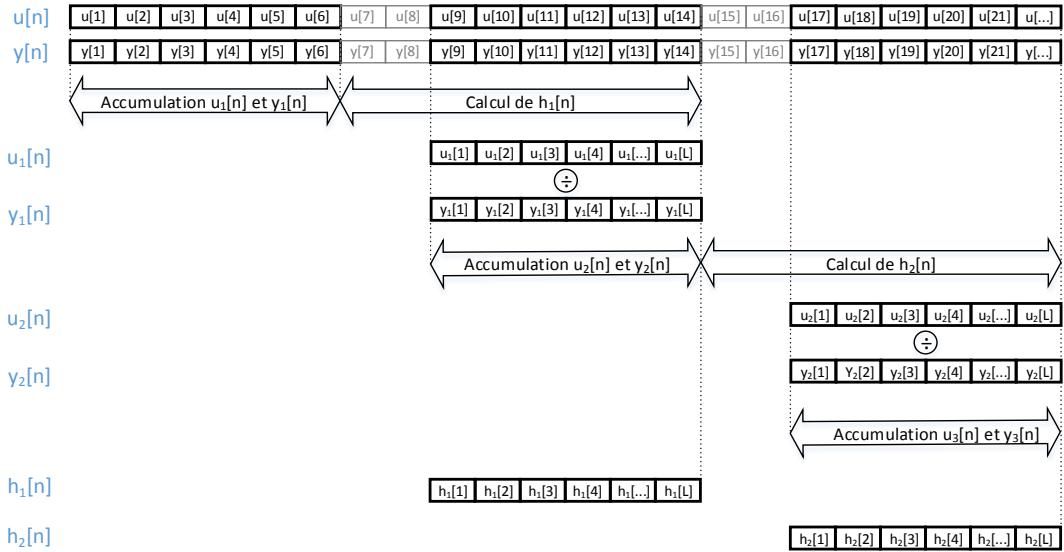


Figure 5.2 Séquencement où le temps de calcul est maître

On note que les échantillons 7 et 8 ainsi que 15 et 16 des signaux $u[n]$ et $y[n]$ ne participent pas aux calculs de réponses impulsionales. L'accumulation des séquences $u_i[n]$ et $y_i[n]$ est faite à l'aide de mémoires tampons circulaires. Ainsi, au moment où le calcul de $h_i[n]$ est terminé, le calcul de $h_{i+1}[n]$ est immédiatement lancé en utilisant les L derniers échantillons de $u[n]$ et $y[n]$ contenus dans les mémoires tampons circulaires. Le désavantage d'une telle approche est un plus grand temps de convergence pour l'algorithme de modélisation qui ne bénéficie pas de l'utilisation de l'entièreté des données disponibles pour reconstruire au fil du temps son modèle. Ceci n'empêche pas par contre la validation de la méthode, ce qui était

le but de ce prototype Matlab en premier lieu. Tel que vu dans les derniers paragraphes de la section 4.5 et illustré par la figure 4.12, une moyenne mobile exponentielle pondérée est utilisée pour reconstruire la réponse en fréquence $H[j\omega]$ à partir des réponses individuelles $H_i[j\omega]$. La constante de temps de la portion exponentielle de cette moyenne fut fixée à 1 pour la totalité des essais effectués à l'aide de cette implémentation Matlab pour maximiser le taux de convergence et démontrer la validité de la méthode.

Le prototype d'implémentation résultant a permis, tel que désiré, de valider la capacité de l'algorithme suggéré à obtenir la réponse en fréquence de la chaîne audio. La mesure de la réponse en fréquence de la chaîne audio par balayage sinusoïdal de la plage de fréquence fut aussi implémentée telle que décrite dans la section 4.4. La figure 5.4 montre une réponse en fréquence obtenue avec le prototype d'implémentation Matlab.

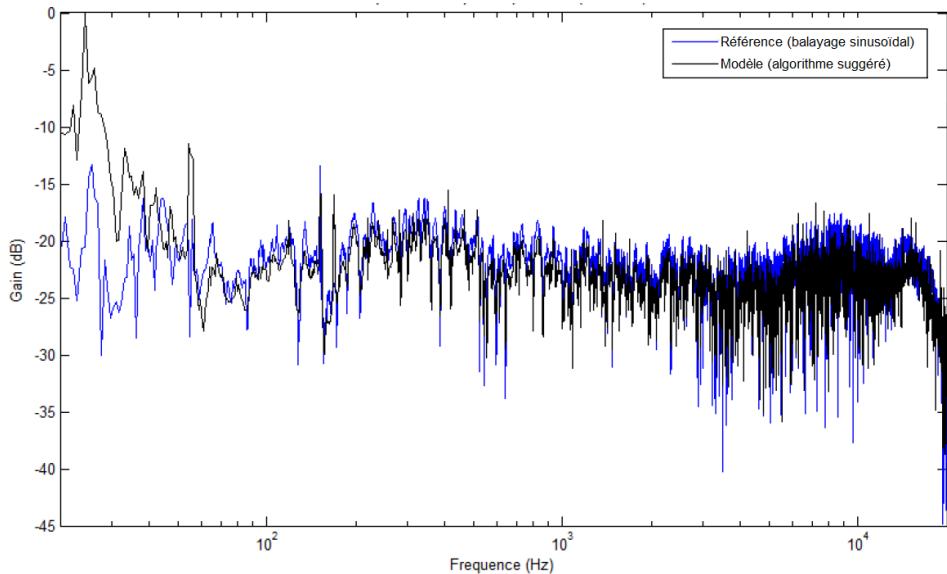


Figure 5.3 Réponse en fréquence obtenue avec le prototype d'implémentation Matlab

Cette méthode standard sert à obtenir la réponse en fréquence servant de référence et avec laquelle on compare celle obtenue à l'aide de l'algorithme suggéré afin de mesurer la performance de ce dernier. La métrique utilisée pour en évaluer la performance est l'erreur quadratique moyenne décrite dans la section 5.1. La figure 5.4 montre l'erreur quadratique moyenne qui décroît à mesure que la musique défile et que des blocs de 8192 échantillons sont analysés.

Un tel résultat en début de projet permet de valider la méthode suggérée et de procéder avec une implémentation plus performante qui non seulement caractérisera la chaîne audio à l'étude mais inclura aussi un filtre égalisateur adaptatif dont la réponse découle de cette chaîne.

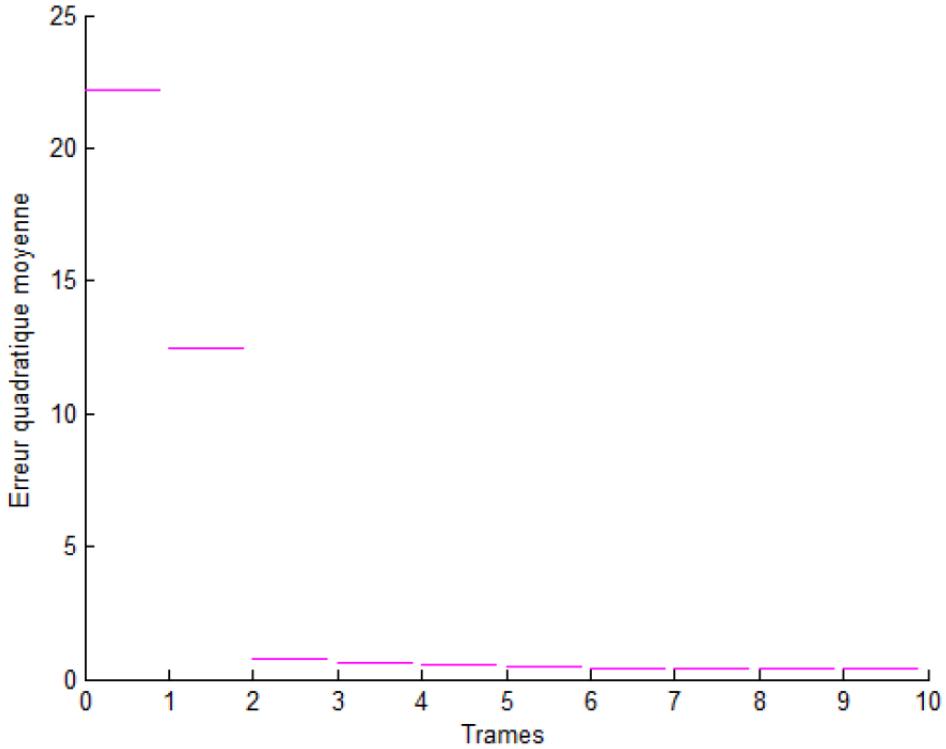


Figure 5.4 EQM la référence obtenue par balayage sinusoïdal et la méthode de modélisation suggérée dans le prototype d'implémentation Matlab

5.3 Implémentation logicielle finale de l'algorithme

La section précédente (5.2) portait sur l'implémentation partielle de l'algorithme, notamment la portion novatrice nécessitant tests et validation en début de projet. Pour de tels tests, la rapidité et la facilité avec laquelle l'algorithme pouvait être implémenté primaient sur la performance de cette implémentation ou sur la latence observée lors de l'exécution. Pour ce faire, Matlab fut utilisé. La présente section traite de l'implémentation finale de l'algorithme développé. Le but est non seulement d'implémenter plus efficacement la portion de la méthode suggérée implémentée en Matlab précédemment, mais le reste de la méthode suggérée, notamment la portion du filtrage adaptatif, doit aussi être implémentée.

Un langage de programmation offrant plus de performance que Matlab était nécessaire. Contrairement à l'approche utilisée dans l'implémentation prototype Matlab de l'algorithme de modélisation présenté dans la section 5.2, il est question cette fois pour la version finale de l'algorithme d'une implémentation temps-réel. Récapitulons. L'objectif premier de l'algorithme de modélisation est d'obtenir la réponse impulsionnelle $h[n]$ de la chaîne audio en comparant le signal $u[n]$ y étant transmis à celui capté en sortie $y[n]$ capté par un microphone placé dans l'environnement d'écoute. La figure 4.11 montre le concept de base derrière cet

algorithme de modélisation reposant sur la déconvolution. La figure 4.12 montre la variante développée dans le cadre de ce travail permettant notamment de faire une moyenne pondérée sur les réponses en fréquence $H_i[j\omega]$. La figure 5.2 montre le séquencement des opérations utilisées dans l'implémentation Matlab. Cette implémentation n'était pas temps-réel dans le sens où toutes les données n'étaient pas traitées dans le cas où il en prend plus de temps à les traiter qu'il n'en prend à les accumuler. L'approche utilisée pour l'implémentation finale diffère du prototypage Matlab. Cette fois-ci, l'implémentation doit être suffisamment performante pour que toutes les données sans exception soient traitées afin de maximiser le taux de convergence de l'algorithme de modélisation. On définit ici le taux de convergence comme étant la rapidité avec laquelle l'erreur quadratique moyenne sur la réponse en fréquence décroît au défilement de la musique. La figure 5.5 montre le séquencement des opérations utilisées.

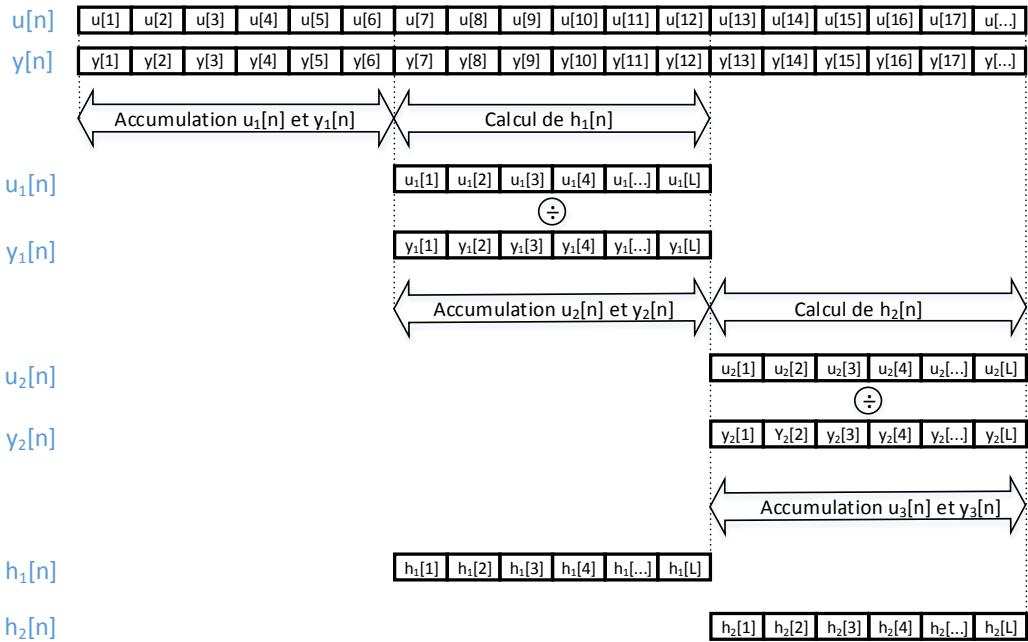


Figure 5.5 Séquencement où le calcul doit être terminé dans le temps alloué

Seule la période durant laquelle les données $u_{i+1}[n]$ et $y_{i+1}[n]$ sont accumulées en vue de l'itération suivante est allouée pour compléter le calcul de $h_i[n]$. Ce temps de calcul T peut être calculé à partir du nombre d'échantillons accumulé L et de la fréquence d'échantillonnage $f_{éch}$.

$$T = \frac{T_{éch}}{L} = \frac{8192 \text{ éch}}{44\,100 \text{ éch/s}} = 0,186 \text{ s} \quad (5.3)$$

Du calcul de $h[n]$ découle le calcul de la réponse du filtre égalisateur $g[n]$. Tel que vu dans la section 4.7, l'analyse temps-fréquence et la décomposition par bandes sont utilisées pour le calcul périodique de $g[n]$ à partir de $h[n]$. Une telle analyse temps-fréquence de la réponse impulsionale $h[n]$ de la chaîne audio permet une meilleure interprétation du comportement de la chaîne audio, mais génère en revanche une importante quantité de données qui sont difficiles à traiter en temps réel. Cette partie de l'algorithme justifie aussi l'utilisation d'un langage de programmation plus performant. De plus, le filtre servant à faire le traitement sur chacune des 10 bandes individuellement fut décomposé de façon à alléger la charge sur le processeur. Chaque fois que la méthode de traitement des données est appelée à faire le traitement des 1024 échantillons mis dans le tampon d'entrée et à calculer la valeur des 1024 échantillons qui seront mis dans le tampon de sortie $u_i[n]$, une seule des 10 bandes est traitée. Le tableau 5.1 montre le séquencement des opérations. Chaque itération i correspond à l'arrivée en entrée de 1024 nouveaux échantillons, il s'agit donc d'itérations de 23 ms chacune.

Tableau 5.1 Séquencement des opérations de l'implémentation logicielle

Itér.	Calcul effectué
1	Calcul de $u_i[n]$ et $h_{bande1}[n]$
2	Calcul de $u_{i+1}[n]$ et $h_{bande2}[n]$
3	Calcul de $u_{i+2}[n]$ et $h_{bande3}[n]$
4	Calcul de $u_{i+3}[n]$ et $h_{bande4}[n]$
5	Calcul de $u_{i+4}[n]$ et $h_{bande5}[n]$
6	Calcul de $u_{i+5}[n]$ et $h_{bande6}[n]$
7	Calcul de $u_{i+6}[n]$ et $h_{bande7}[n]$
8	Calcul de $u_{i+7}[n]$ et $h_{bande8}[n]$
9	Calcul de $u_{i+8}[n]$ et $h_{bande9}[n]$
10	Calcul de $u_{i+9}[n]$ et $h_{bande10}[n]$
11	Calcul de $u_{i+10}[n]$ et recombinaison des 10 bandes
12	Calcul de $u_{i+11}[n]$ et $g[n]$ par inversion
13	Calcul de $u_{i+12}[n]$ et alignement et fenêtrage $g[n]$

De plus, comme la réponse pondérée $h[n]$ est calculée sur des blocs de 8192 échantillons, donc toutes les 8 itérations, un mécanisme repousse à l'itération suivante le traitement des données par bande de fréquences si un calcul de $h[n]$ a eu lieu durant la présente itération. Ce mécanisme permet d'éviter de faire trop d'opérations durant une même itération. La même réponse pondérée $h[n]$ est utilisée pour le calcul sur chacune des 10 bandes de fréquence afin

d'éviter un mauvais alignement au moment de la recombinaison des 10 bandes à l'itération 11.

La notion de temps réel dans l'implémentation s'étend au-delà des blocs de « Modélisation » et de « Conception du filtre égalisateur » de la figure 1.5. La portion du filtrage adaptatif représenté par le bloc « Filtre EQ Adaptatif » requiert aussi que le traitement soit fait en temps réel. Par « temps réel », dans ce cas, on entend que le résultat d'un calcul effectué sur un bloc de 1024 échantillons en entrée $r_i[n]$ doit générer 1024 nouveaux échantillons en sortie ($u_i[n]$) avant qu'un nouveau bloc $r_{i+1}[n]$ se présente en entrée. La notation r et u utilisée pour nommer l'entrée et la sortie du filtre égalisateur est celle utilisée dans l'ensemble du chapitre 4. Dans le cas bien précis de ce filtre, pour connaître la quantité de temps alloué au calcul, on utilise l'équation suivante :

$$T = \frac{N}{f_{\text{éch}}} = \frac{1024 \text{ éch}}{44\,100 \text{ éch/s}} = 0,0232 \text{ s} \quad (5.4)$$

L'implémentation du « Filtre EQ Adaptatif » doit être suffisamment performante pour ne pas que la musique traitée soit interrompue. Dans le cas bien précis de ce filtre, on peut calculer la quantité de temps alloué au calcul de la convolution linéaire pour un bloc de N échantillons tel qu'illustré par la figure 4.29 en utilisant la même équation.

Vu l'impossibilité de rencontrer ces requis pour ce qui est du temps de calcul avec Matlab, on s'est tourné vers le C++ pour l'implémentation finale. Matlab étant à la base un langage fait pour opérer sur des matrices ou des vecteurs de données, le passage vers le C++ créait le besoin pour le même genre de fonctionnalité puisque dans l'algorithme suggéré, l'ensemble des données se présente sous forme de blocs (vecteurs). Addition de vecteurs, soustraction de vecteurs, multiplication de vecteurs et division de vecteurs ; des fonctions mathématiques de base capables d'opérer sur des vecteurs devaient être implémentées. Aussi, l'implémentation de l'algorithme suggéré requiert des opérateurs permettant de manipuler les éléments d'un vecteur. On compte parmi ces opérateurs le redimensionnement, l'ajout d'éléments nuls, la translation de l'ensemble des éléments, la rotation de l'ensemble des éléments, le seuillage, la génération de données linéairement ou logarithmiquement espacées et l'inversion ($1/x$).

5.4 Environnement de test

Il a été préétabli que dans ce travail, on s'intéresse aux salles d'écoute dites petites. Cependant, à des fins de test, il était important d'avoir accès à un environnement de test afin de mesurer la performance de l'algorithme développé dans un réel contexte d'écoute. Pour ce faire, un partenariat fut formé avec un studio d'enregistrement de la région de Montréal, en l'occurrence La Coopérative du Studio CME. Ce studio comporte quatre salles principales.

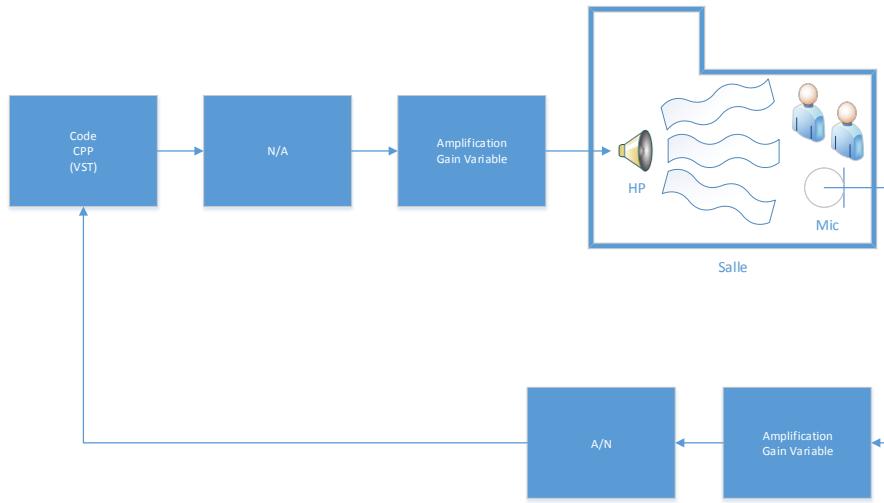


Figure 5.6 Signaux audio dans l'environnement VST

La salle choisie pour la réalisation de ce travail est la salle de régie puisque sa fonctionnalité primaire est l'écoute sur haut-parleur. La régie du Studio CME est utilisée pour le montage et le mixage de bandes sonores et de pièces musicales.

Tableau 5.2 Caractéristiques de la salle de régie

Superficie de 41,75 mètres carrés
Plancher avec tapis
Plafond suspendu installé en accordéon
Antichambre munie de 2 portes patio pour accéder à la salle de prise de son
Fenêtre thermos double de 1,25 mètre par 2,5 mètres avec vue sur la prise de son

Le volume relativement restreint, la géométrie quasi cubique et l'absence d'absorbant, de diffuseur ou de résonateur acoustiques, font de cette pièce un environnement propice à l'étude de l'acoustique des petites salles. L'image suivante montre l'apparence globale de la salle traitée.

La distance entre la position d'écoute et la source y varie normalement entre 1 et 2 mètres. L'ensemble des tests expérimentaux réalisés et pour lesquels les résultats sont présentés dans les sections 5.9 et 5.10, le microphone de mesure est placé à un mètre de distance devant la source (haut-parleur). Les tout derniers tests expérimentaux réalisés et présentés à la toute fin des sections 5.9 et 5.10 ont été réalisés avec une distance de deux mètres entre la source et le microphone dans le but d'évaluer l'effet de la distance entre la source et la position



Figure 5.7 Salle de régie du Studio CME

d'écoute sur les méthodes de modélisation et de filtrage suggérées.

5.5 Matériel utilisé

Un composant de la chaîne audio nécessaire à l'égalisation adaptative telle que montrée dans la figure 1.5 est le microphone. Dans le cadre de ce travail, le microphone est placé aussi près que possible de la position d'écoute sans obstruer l'auditeur. Tel que décrit dans la section 4.2, le microphone et son amplificateur, dans la figure 4.3 représentés dans le domaine temporel par leurs réponses impulsionales $h_4(t)$, est un instrument de mesure de haute précision et est considéré idéal. C'est-à-dire qu'il fournit en sortie une tension égale à la pression acoustique à la position d'écoute.

Typiquement, un microphone conçu spécialement pour la mesure est utilisé dans ce cas. Les microphones de mesure partagent des propriétés qui les distinguent des micros typiquement utilisés pour enregistrer la voix ou des instruments. Le microphone à condensateur utilisé ici est un microphone de mesure Behringer, le ECM8000 montré dans la figure 5.8.

Les microphones de mesure sont normalement omnidirectionnels, c'est-à-dire que leur courbe de directivité reste inchangée peu importe l'angle d'incidence de la source. La figure 5.9 montre le patron polaire omnidirectionnel du ECM8000.

La réponse en fréquences de microphones de mesure est normalement très plate, à l'intérieur d'une tolérance de $\pm 0,1$ dB sur la plage de fréquence de 20 Hz à 20 kHz. Cette donnée



Figure 5.8 Microphone de mesure ECM8000 de Behringer, image tirée de la documentation du fabricant (ecm, 2000)

n'est pas explicitement fournie par le fabricant du ECM8000. Cependant, la courbe de la figure 5.10 extraite de la fiche technique du produit (ecm, 2000) montre que la réponse en fréquence du ECM8000 est contenue à l'intérieur d'une tolérance de ± 2 dB. Le microphone ECM8000 est utilisé même s'il n'a pas une réponse en fréquence exemplaire en raison de son accessibilité et de son prix de 60 dollars.

Les microphones de mesure sont normalement calibrés, c'est-à-dire qu'on connaît avec précision leur sensibilité en volts par pascal. Cette sensibilité est normalement plus grande que dans le cas des micros utilisés pour enregistrer la voix, permettant d'enregistrer des sources sonores de très faible amplitude lorsque nécessaire. Dans le cadre de ce travail, la calibration du microphone est peu pertinente puisque l'algorithme exploitant le signal en provenance du microphone fait une normalisation de l'amplitude tel que vu dans la section 4.8. On ne s'intéresse pas au gain du système en boucle ouverte, par conséquent il n'est pas nécessaire de connaître précisément la sensibilité du microphone en volts par pascal. À titre de référence, cette sensibilité fut établie à 8,20 mV/Pa lorsqu'alimenté avec une tension de polarisation de 48 V.

L'ensemble de ces propriétés sont plus facilement atteignables chez une famille de microphones toute particulière : les microphones à condensateurs. Le ECM8000 fait partie de cette famille.

Les convertisseurs numériques-analogiques et analogiques-numériques de la figure 1.5 sont présentés comme deux entités distinctes. Dans une chaîne audio classique, il n'est pas rare que la CNA et la CAN soient assurées par un seul système. Certains nomment un tel système carte de son, d'autres interface de son. La première appellation est plus commune et est utilisée dans cet ouvrage. En plus d'assurer le passage du domaine temporel continu au domaine temporel discret, une carte de son comporte normalement un pré-amplificateur permettant d'amener le signal électrique en provenance d'un microphone à une amplitude à celle de la plage dynamique des convertisseurs. Normalement, le gain de cet amplificateur est variable tel que dans la figure 1.5 et donc contrôlable soit par logiciel ou à l'aide d'un bouton de réglage comme c'est le cas sur la carte de son Focusrite Scarlett 2i2 utilisée dans le cadre de ce travail. La carte Scarlett 2i2 comporte deux entrées micros pré-amplifiées et donc deux boutons de réglage de gain étiquetés « GAIN 01 » et « GAIN 02 » dans la figure 5.11 montrant l'avant

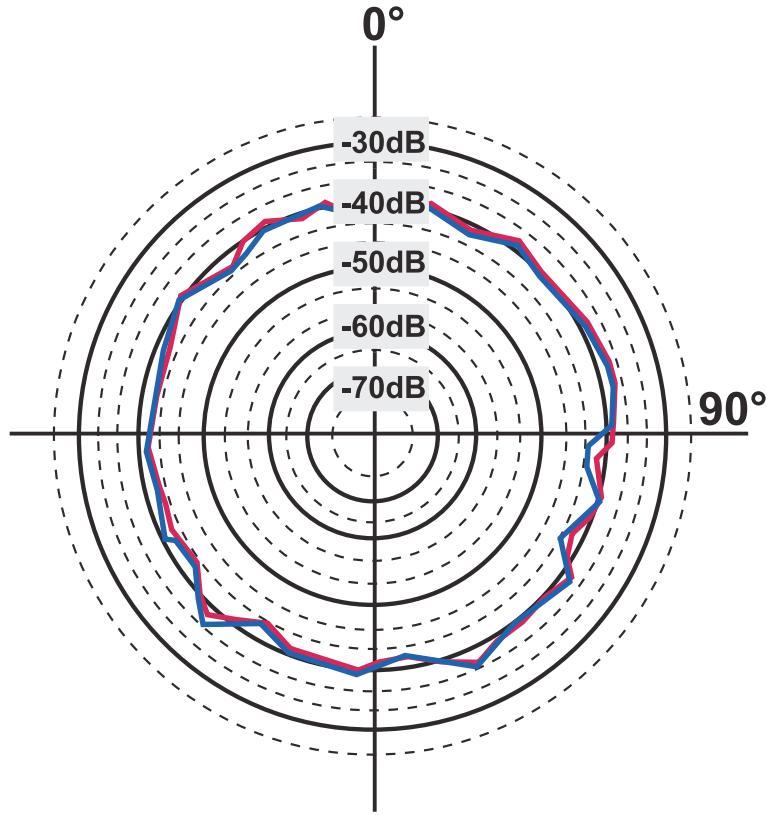


Figure 5.9 Patron polaire omnidirectionnel du ECM8000, image tirée de la documentation du fabricant (ecm, 2000)

de la carte.

Dans le cas d'un système intégré tel que la carte de son Scarlett 2i2 (2i2, 2011), les fabricants omettent souvent d'indiquer la plage en volts des convertisseur analogique-numérique utilisé dans leur design.

Le haut-parleur utilisé dans le cadre de ce travail est un moniteur de studio B2031A de Behringer. La plage de fréquence couverte par ce haut-parleur est de 50 à 21000 Hz selon la documentation. On verra dans la section 5.10 que ce paramètre est essentiel au fonctionnement du filtre égalisateur implémenté. Ce haut-parleur est du type actif. Contrairement au type passif qui requiert la présence d'un amplificateur externe, le B2031A a son propre amplificateur à l'intérieur.

Bien que le haut-parleur fasse partie de la chaîne audio faisant l'objet de modélisation et d'égalisation dans le cadre de ce travail, la revue des spécifications techniques approfondies telles que des courbes de réponses en fréquence n'était pas nécessaire pour la réalisation de ce travail. Comme le montre la figure 4.4, dans l'approche utilisée, le haut-parleur est vu comme une boîte noire au même titre que les autres composantes de la chaîne audio telles

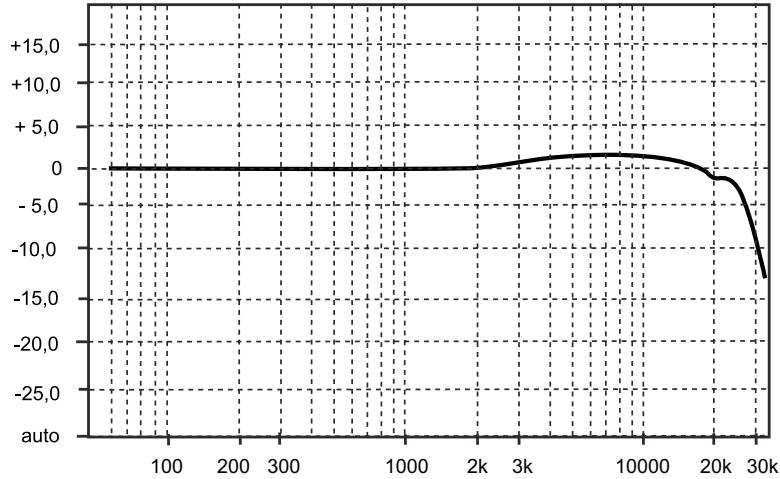


Figure 5.10 Réponse en fréquence du ECM8000, image tirée de la documentation du fabricant (ecm, 2000)

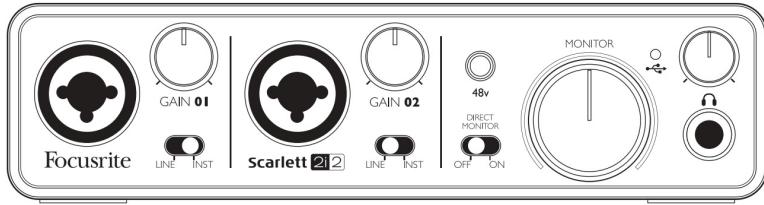


Figure 5.11 Vue avant de la carte de son Focusrite Scarlett 2i2, image tirée de la documentation du fabricant (2i2, 2011)

que l'amplificateur. La figure 4.5 montre que nous choisissons de considérer l'ensemble de la chaîne comme partie intégrante de la salle. Nous ne nous attardons pas aux spécificités et aux imperfections propres aux haut-parleurs, pas plus que nous le faisons pour l'amplificateur dans le cas où il serait externe (il est intégré dans le cas du B2031A de Behringer). On s'attarde plutôt sur la salle, sur les propriétés qui lui sont propres, et sur la manière de contrer les transformations qu'elle induit sur l'onde acoustique qui s'y déplace.

5.6 Fonctionnement du module VST dans l'hôte « VST-Host »

Telle qu'évoquée dans la section 5.3, l'encapsulation de l'implémentation C++ du filtre égalisateur adaptatif dans un « plugin » suivant le standard VST de Steinberg rend le filtre plus versatile. Le standard VST implémente une méthode appelée « process » lorsqu'un tampon en entrée d'une taille prédéfinie est rempli d'échantillons audio prêts à être lus. Libre au développeur de manipuler ces échantillons selon ce qu'il choisit de mettre comme code

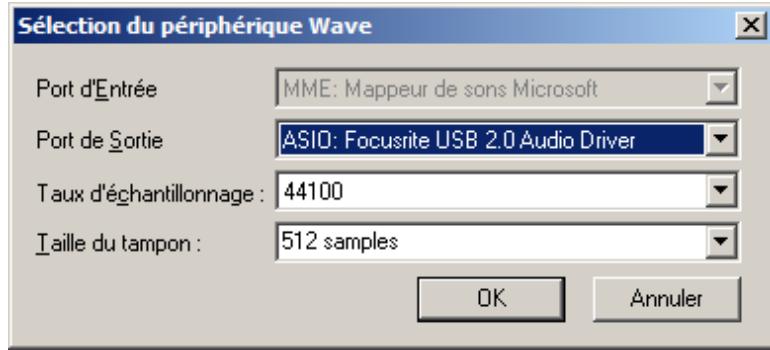


Figure 5.12 Configuration du périphérique audio « Focusrite »

dans la méthode « process ». Dans une application temps-réel, avec une entrée et une sortie, tel que c'est le cas dans le présent travail, le code va lire les échantillons en entrée, leur appliquer une transformation quelconque, et écrire les échantillons dans le tampon de sortie. Les modules VST peuvent être cascadés de façon à ce que la sortie d'un premier soit l'entrée d'un deuxième. Dans une application combinant logiciel et matériel, le premier module de la chaîne reçoit l'audio en provenance des entrées du périphérique audio. Les sorties du dernier module de la chaîne sont envoyées à la sortie du périphérique audio. Les modules VST n'ont pas à gérer l'interface avec le périphérique audio, c'est le logiciel hôte dans lequel les modules sont chargés qui assume ce rôle. Le fait que les données soient traitées par blocs de taille configurable rend le standard attrayant pour un algorithme tel que celui implémenté dans le cadre de ce travail où à chaque itération, un bloc de 1024 échantillons en provenance d'une source audionumérique entre dans le filtre égalisateur et un bloc de 1024 échantillons en sort pour être envoyé vers la sortie du périphérique audio.

En plus de procurer une couche d'abstraction entre l'implémentation logicielle de l'algorithme et le matériel tel qu'un périphérique audio, il favorise la réutilisation du code. Le standard permet de s'assurer que les différents modules développés par différentes personnes et ayant tous des fonctionnalités différentes puissent être chargés dans un même logiciel hôte et connectés les uns aux autres pour définir le chemin qu'empruntera le signal audionumérique et les transformations qui y seront appliquées en cours de route. Plusieurs logiciels hôtes supportent le standard, ce qui permet à chacun d'utiliser les modules qui leur conviennent dans l'environnement qui leur convient. Le logiciel hôte utilisé dans le cadre de ce travail est VST-Host version 1.53 (Seib, 2013), disponible en téléchargement gratuit et développé par Hermann Seib. Dans notre cas, c'est VST-Host qui gère l'interface avec le périphérique audio Focusrite. L'interface graphique du VST-Host est montré dans la figure 5.13.

Le bloc « Entrée du moteur » situé du côté gauche dans la figure 5.13 est le bloc qui donne accès au signal audio en provenance de l'entrée du périphérique audio. Le bloc de droite

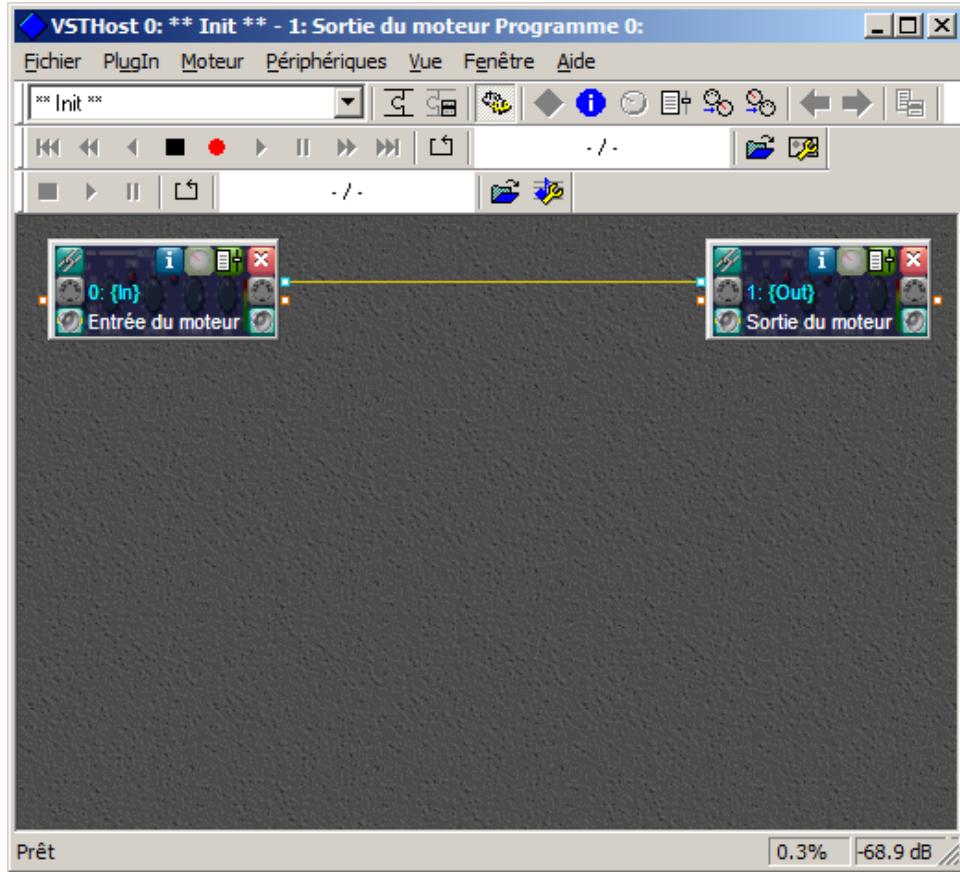


Figure 5.13 Interface graphique VST-Host

« Sortie » permet d'envoyer un signal en sortie au périphérique audio. Dans le branchement montré par la figure, l'entrée est mise directement en sortie. Le terme « moteur » est la traction mot à mot de l'anglais « engine » automatiquement effectuée par le logiciel VST-Host lorsqu'il est configuré en français.

La configuration utilisée pour la réalisation de tests expérimentaux sur l'algorithme implémenté est montrée par la figure 5.14.

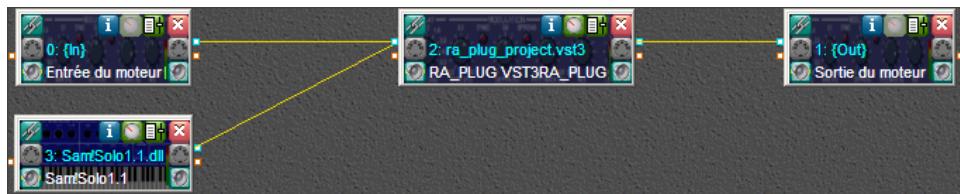


Figure 5.14 Branchement des plugins dans le « VST-Host » pour la réalisation de tests expérimentaux

Du côté gauche se trouve un module VST nommé « SamSolo » qui permet la lecture de

fichiers audio en format « wav ». Le rôle de la source audionumérique de la figure 1.5 est assuré par le module SamSolo. Le standard VST permet à chaque module d'avoir sa propre interface graphique. La figure 5.15 montre l'interface graphique du module SamSolo, on s'en sert entre autres pour charger un fichier audio.



Figure 5.15 Interface graphique du plugin « SamSolo »

Au centre se trouve le module VST contenant l'implémentation de l'algorithme développé dans le cadre de ce travail. Le module est nommé « ra-plug-project ». Le module-VST « ra-plug-project » assure le rôle de 3 blocs de la figure 1.5 : la modélisation, la conception du filtre égalisateur et le filtre EQ adaptatif. Pour ce faire, il doit avoir deux ports d'entrée, le premier lui permettant de recevoir le signal en provenance de la source audio, le second le signal en provenance du microphone placé dans l'environnement d'écoute. Les figures 5.16 et 5.17 montrent la connexion du signal en provenance du microphone et de la source audio en provenance du module SamSolo respectivement.

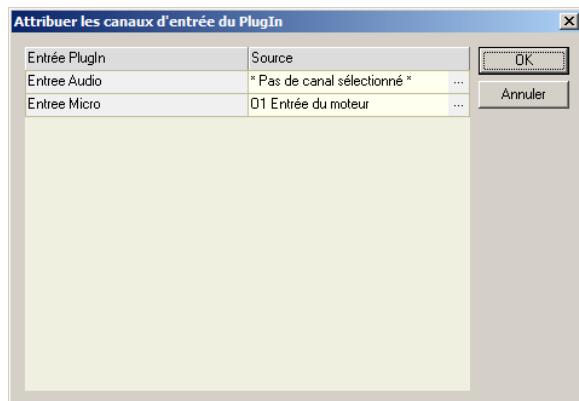


Figure 5.16 Entrée numérique microphone

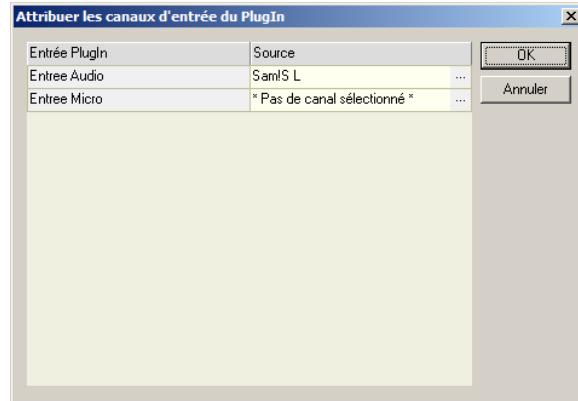


Figure 5.17 Entrée numérique audio

Le port « émis » du bloc « modélisation » de la figure 1.5 est un port interne du module « ra-plug-project », ce port lui permet de recevoir sa propre sortie. Il s'en sert pour comparer le signal transmis à la chaîne audio à celui capté par le microphone. Ces signaux sont utilisés par l'algorithme de modélisation décrit dans la section 4.5. Le module a un port de sortie connecté à la sortie du moteur permettant d'envoyer le signal vers le périphérique Focusrite et vers le haut-parleur B2030A. La figure 5.18 montre la connexion de cette sortie.



Figure 5.18 Entrée-numérique

La figure 5.19 montre l'interface graphique du module « ra-plug-project ». Le tableau 5.3 résume le fonctionnement des différents paramètres.

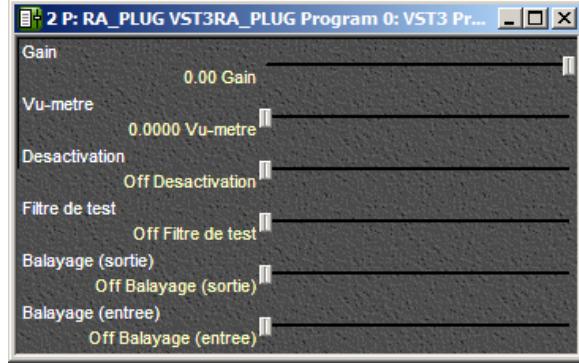


Figure 5.19 Interface graphique du plugin d'égalisation adaptative développé

Tableau 5.3 Paramètres du plugin de la figure 5.19

Paramètre	Entrée/Sortie	Type	Description
<i>Gain</i>	sortie	virgule flottante	Gain appliqué au signal audio en sortie du plugin
<i>Vu-mètre</i>	entrée	virgule flottante	Indique le niveau de la sortie du plugin
<i>Désactivation</i>	entrée	booléen	Permet de désactiver le plugin
<i>Filtre de test</i>	entrée	virgule flottante	Écrase la réponse du filtre égalisateur avec celle d'un filtre passe-bas avec fréquence de coupure contrôlable
<i>Balayage (sortie)</i>	entrée	booléen	Lorsqu'activé, applique un balayage sinusoïdal à la sortie du plugin. Permet de mesurer la réponse de la chaîne audio sans le filtre égalisateur.
<i>Balayage (entrée)</i>	entrée	booléen	Lorsqu'activé, applique un balayage sinusoïdal à l'entrée du plugin. Permet de mesurer la réponse de la chaîne audio avec le filtre égalisateur.

5.7 Ratio signal sur bruit de la chaîne

Pour obtenir le RSB instantané comme fonction de la fréquence, le bruit de fond de la salle fut enregistré. On a également enregistré l'émission sur haut-parleur d'un bruit blanc dont l'amplitude était établie à 0 dBFS, donc de façon à couvrir la pleine plage dynamique du système. Le spectrogramme de la figure 4.13 de la section 4.5 montre un bruit blanc tel que celui utilisé, sa pleine durée est de 60 secondes. Le bruit fut joué pendant 60 secondes dans la salle et la portion centrale de l'enregistrement d'une durée de 262 144 échantillons, soit environ 6 secondes, fut utilisée pour le calcul. Un enregistrement du bruit de fond de même durée fut utilisé. Les figures 5.20 et 5.21 montrent respectivement le spectrogramme de l'enregistrement du bruit blanc et du bruit de fond.

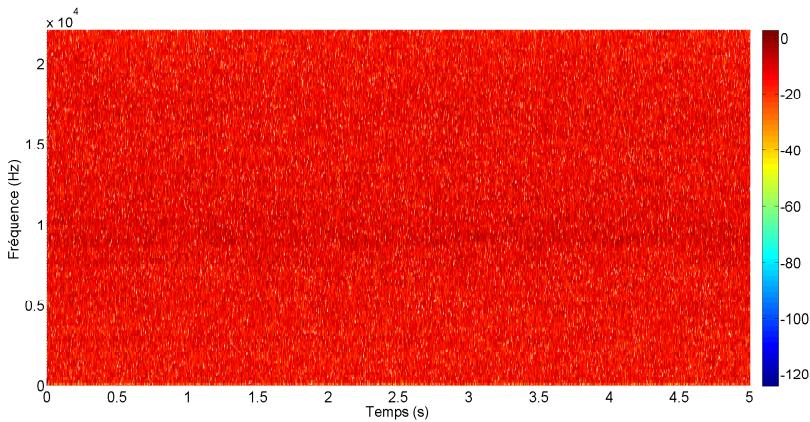


Figure 5.20 Enregistrement d'un bruit blanc à pleine amplitude utilisé pour le calcul du RSB

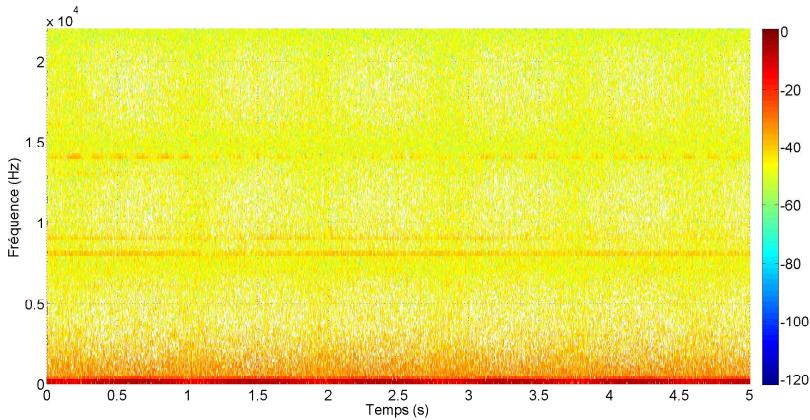


Figure 5.21 Enregistrement du bruit de fond utilisé pour le calcul du RSB

Le RSB instantané est calculé en divisant la magnitude de la transformée de Fourier de

l'enregistrement du bruit blanc par celle de l'enregistrement du silence. La figure 5.22 montre le RSB de la chaîne audio mesuré préalablement au reste des tests.

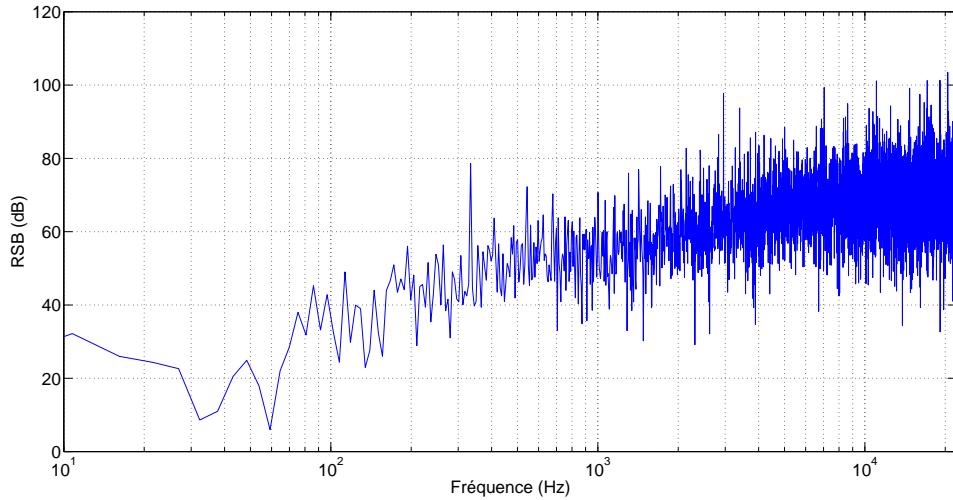


Figure 5.22 Rapport signal sur bruit de la chaîne audio utilisée pour la mesure

5.8 Obtention du modèle de référence

La présente section couvre l'obtention du modèle de référence. Celui-ci sera utilisé plus tard dans la section 5.9 à des fins de comparaison avec la méthode de modélisation suggérée dans le but d'en évaluer la performance. Le modèle de référence est tout simplement constitué d'une réponse impulsionnelle et de la réponse en fréquence y étant associée. Celui-ci est obtenu expérimentalement à l'aide d'une méthode standard bien connue et décrite dans la section 4.4. La distance entre le microphone et la source (haut-parleur) pour ce test est de 1 mètre. Le signal d'excitation $u_{ref}[n]$ utilisé pour la méthode de balayage sinusoïdal telle que décrite précédemment est montré par la figure 5.23.

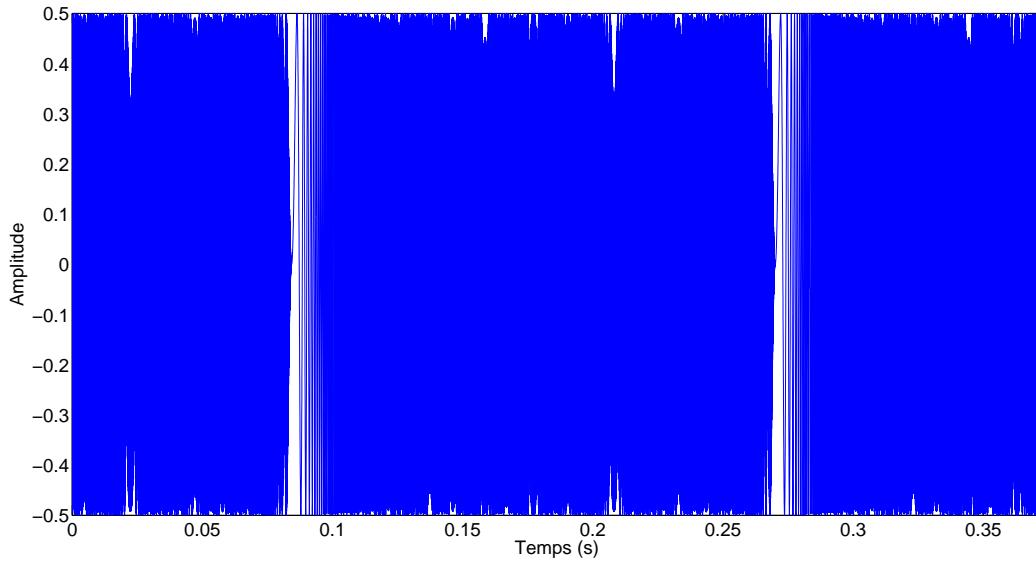


Figure 5.23 Signal de balayage sinusoïdal

Afin de mieux visualiser la variation temporelle du contenu fréquentiel de ce signal, le passage vers une représentation temps-fréquence du signal est de mise. La figure 5.24 montre le spectrogramme du signal d'excitation $u_{ref}[n]$.

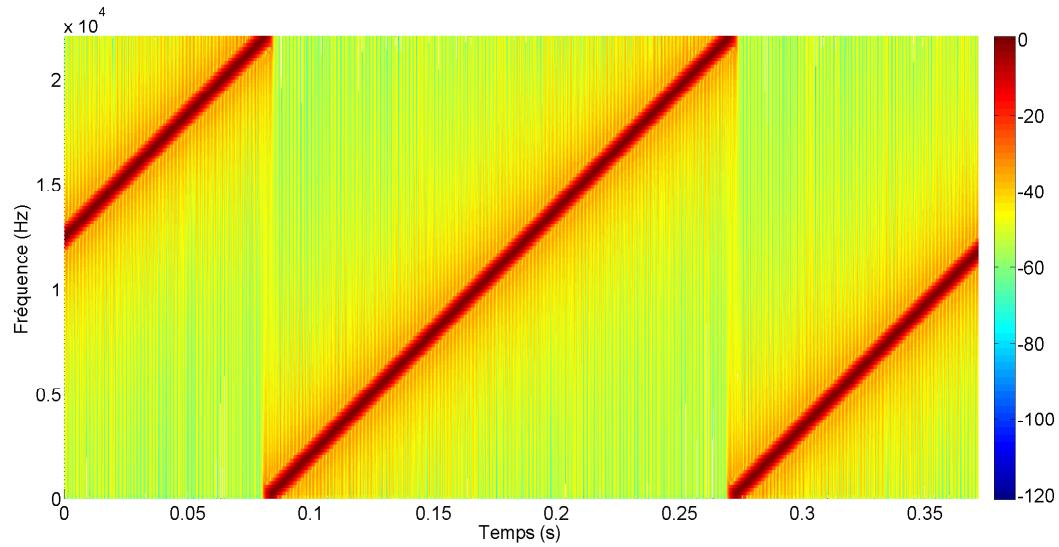


Figure 5.24 Spectrogramme du signal de balayage sinusoïdal (échelle de couleur en dBFS)

L'équation 4.5 nous indique que l'obtention de la réponse impulsionale de la chaîne audio $h_{ref}[n]$ passe par l'obtention de la réponse $y_{ref}[n]$ du système au signal d'excitation

$u_{ref}[n]$. La représentation temporelle ainsi que le spectrogramme de la réponse de la chaîne audio au signal d'excitation sont montrés respectivement par les figures 5.25 et 5.26.

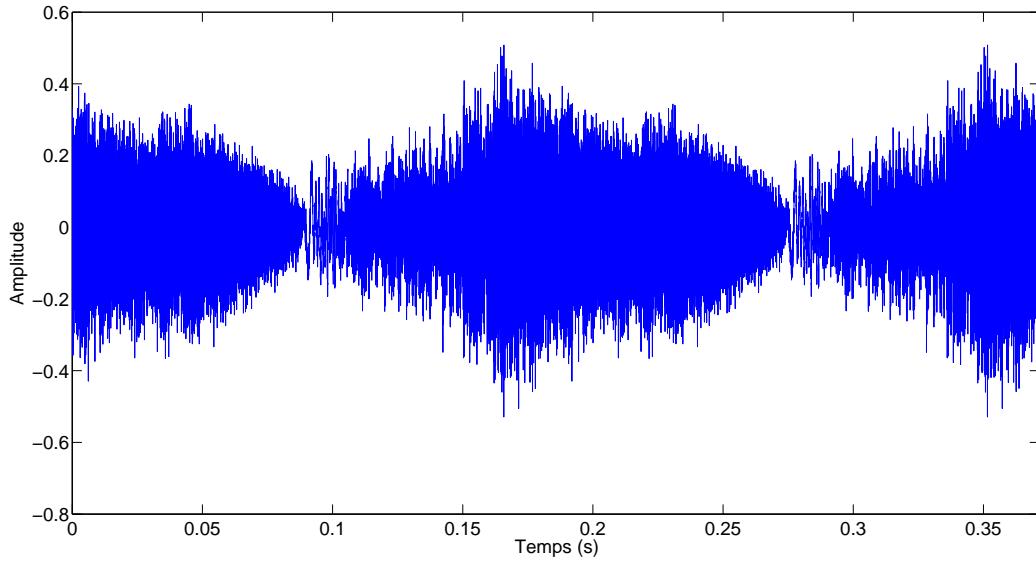


Figure 5.25 Réponse au balayage sinusoïdal

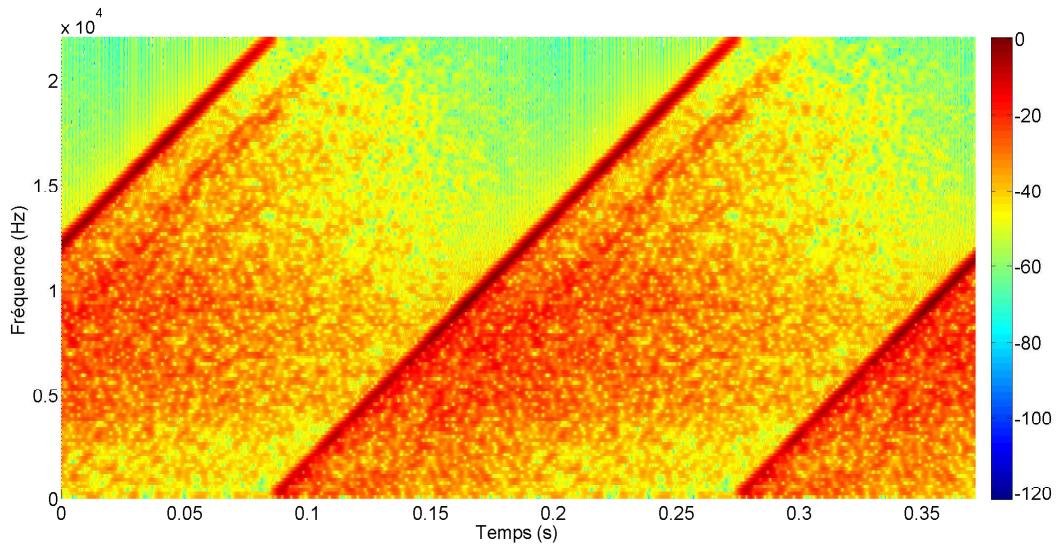


Figure 5.26 Spectrogramme de la réponse au balayage sinusoïdal (échelle de couleur en dBFS)

À l'aide de l'équation 4.5, la réponse impulsionnelle $h_{ref}[n]$ est obtenue et montrée dans la figure 5.27. Dans la majorité des travaux recensés, une échelle d'amplitude linéaire plutôt

que logarithmique est utilisée pour tracer les réponses impulsionales de salles. Cette même échelle est donc utilisée dans ce travail.

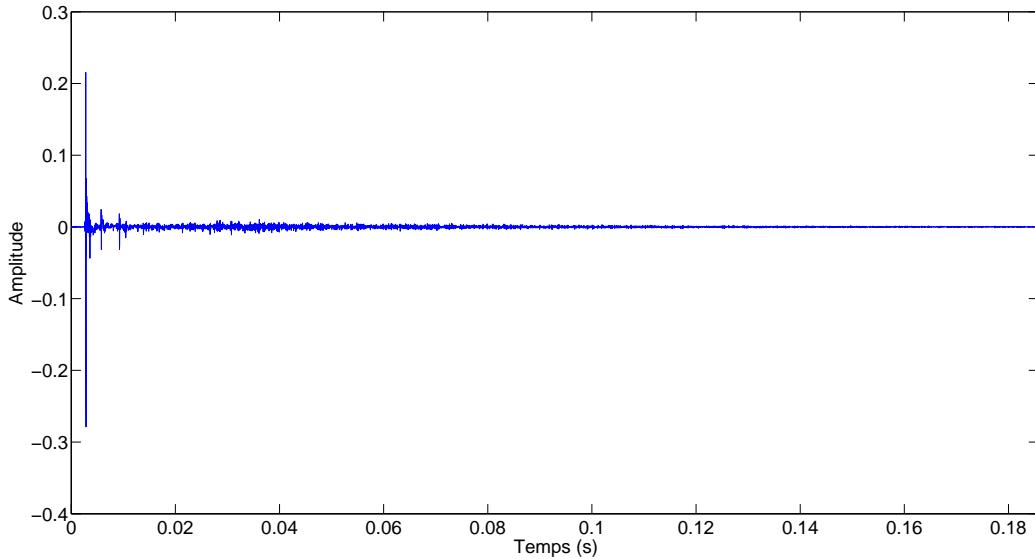


Figure 5.27 Réponse impulsionnelle servante de référence obtenue par balayage sinusoïdal

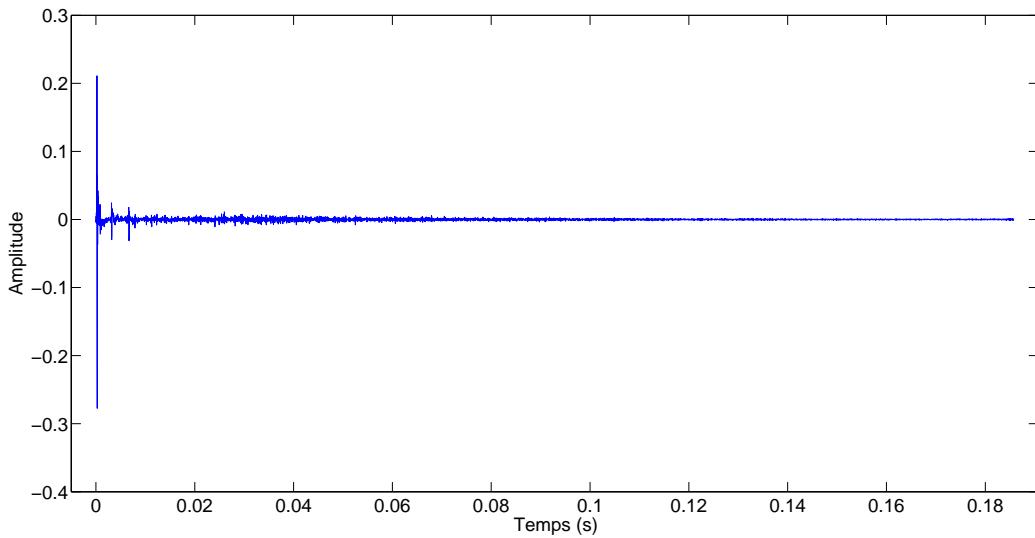


Figure 5.28 Suppression du délai dans la réponse impulsionnelle obtenue par balayage sinusoïdal

On note en observant la figure 5.27 qu'un certain délai est observé avant l'arrivée de la première impulsion (signal direct) dans la réponse impulsionnelle de la chaîne audio. Ce délai

de quelques millisecondes est dû au temps mis par l'onde de son à parcourir la distance de un mètre entre le haut-parleur et la position d'écoute. En utilisant $c = 330 \text{ m/s}$ comme vitesse du son, il est possible de calculer que ce délai est d'exactement 0,003 seconde. Ce délai est éliminé, comme ce sera le cas dans la section 5.9 pour le délai dans la réponse obtenue à l'aide de l'algorithme de modélisation suggéré.

La réponse en fréquence $H_{ref}[j\omega]$ peut être obtenue en appliquant une TFD à la réponse impulsionale $h_{ref}[n]$ à l'aide de la fonction FFT de Matlab. Dans la section 5.9, la réponse impulsionale $h_{ref}[n]$ ainsi que la réponse en fréquence $H_{ref}[j\omega]$ obtenues serviront de référence pour évaluer la performance de l'algorithme de modélisation suggéré.

5.9 Performance de l'algorithme de modélisation suggéré

Cette section contient les résultats expérimentaux permettant de quantifier la performance de l'algorithme de modélisation suggéré. Pour ce faire, les résultats obtenus sont comparés à l'aide de la méthode standard du balayage sinusoïdal de la section précédente (5.8). Dans la présente section, on ne s'intéresse pas encore au filtre égalisateur, mais plutôt simplement à la capacité de l'algorithme de modélisation suggéré d'estimer la réponse de la chaîne audio. Les concepts théoriques ainsi que les équations régissant l'algorithme de modélisation suggéré furent présentés à la section 4.5 du chapitre 4. À moins qu'on ne le précise, les résultats présentés dans cette section résultent de tests effectués avec un mètre séparant la source et le microphone de mesure.

À noter que le filtre égalisateur est désactivé pour cette partie et que l'algorithme de modélisation est actif.

L'algorithme de modélisation suggéré permet d'estimer la réponse de la chaîne audio. La réponse impulsionale donnée par l'algorithme est prise directement à la sortie du bloc de « modélisation » de la figure 1.5. On note d'ailleurs dans cette figure que l'estimation de la réponse impulsionale de la salle est donnée directement par la sortie de ce bloc. La figure 4.12 contient l'algorithme à l'intérieur. Puisque cet algorithme est itératif et qu'il repose sur une moyenne pondérée, la réponse impulsionale obtenue en sortie varie d'itération en itération. Si l'algorithme converge, cette réponse devrait, au fil du temps, s'approcher de celle du modèle de référence. La figure 5.29 montre la réponse impulsionale $h[n]$ obtenue à l'aide de la technique de modélisation suggérée après le traitement de 10 blocs de 8912 échantillons. Les 1.86 premières secondes de la chanson « Stricken » sont jouées et émises sur haut-parleur durant cet exercice de mesure où le microphone placé à une distance d'un mètre du haut-parleur capture la réponse de la salle à la position d'écoute. La figure 5.29 montre aussi la réponse $h_{ref}[n]$ obtenue dans la section précédente avec la technique du balayage

sinusoïdal. Les deux réponses impulsionales sont tracées une fois les délais supprimés dans une même figure à titre de comparaison.

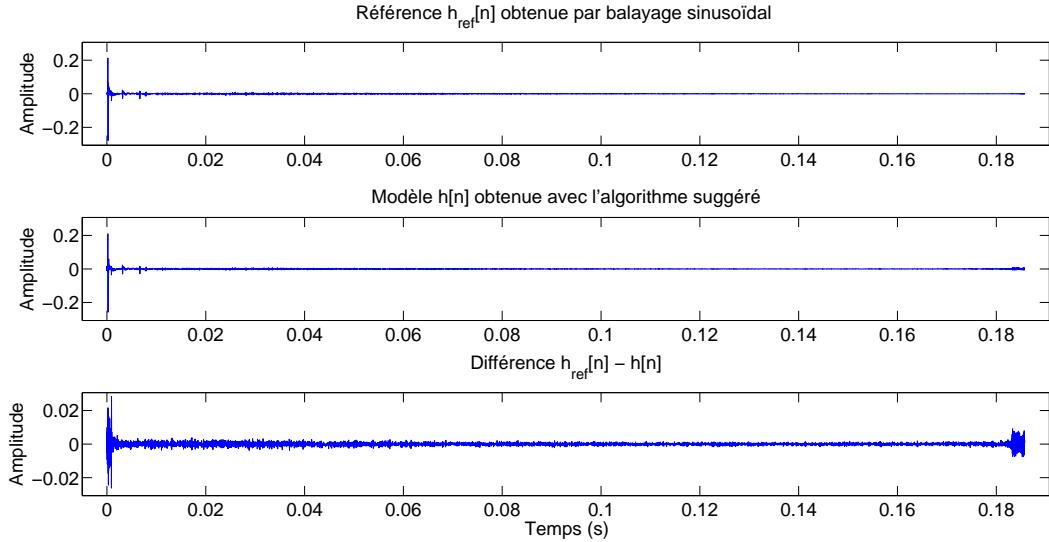


Figure 5.29 Réponse impulsionale obtenue après 10 blocs de données traités

On note que les deux réponses impulsionales de la figure 5.29 sont semblables. Une métrique de similitude entre ces deux réponses (EQM) sera introduite dans les paragraphes à venir dans le domaine fréquentiel plutôt que temporel et servira de critère de performance tel que le suggère la section 5.1. La soustraction des deux courbes indique que l'algorithme semble procurer une juste approximation de la réponse impulsionale de la chaîne audio. Il agrandir l'échelle d'amplitude 10 fois pour bien voir l'erreur.

Les deux réponses peuvent être comparées dans le domaine fréquentiel en appliquant une transformée de Fourier discrète aux réponses impulsionales $h_{ref}[n]$ et $h[n]$ pour obtenir les réponses en fréquence $H_{ref}[j\omega]$ et $H[j\omega]$ respectivement. Ces deux réponses en fréquence sont montrées dans la figure 5.30.

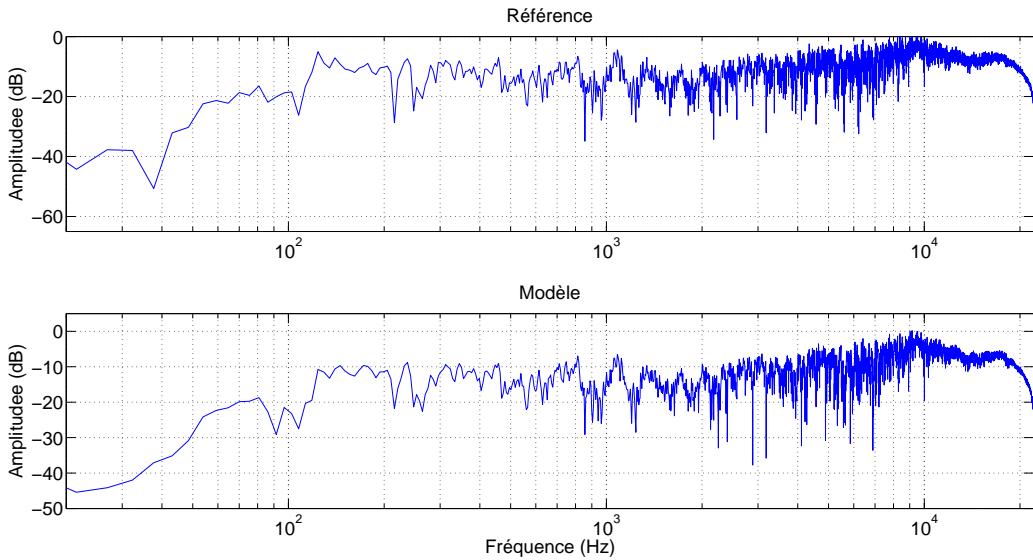


Figure 5.30 Réponse en fréquence obtenue après 10 blocs de données traités

Les réponses en fréquences montrent une similitude satisfaisante sur la pleine plage de fréquence. Afin de quantifier cette similitude, l'erreur quadratique moyenne telle que décrite dans la section 5.1 est utilisée comme métrique. L'avantage d'une telle métrique réside dans le fait qu'une seule quantité résume à elle seule la similitude entre les deux réponses en fréquences pour une itération donnée. Ceci nous permet de tracer l'erreur sur l'estimation de la réponse en fréquence en fonction du temps. Si l'algorithme de modélisation suggéré converge, cette erreur devrait diminuer à chaque itération, ce qui indique que l'estimation de la réponse en fréquence s'approche de la référence. La figure 5.31 montre l'évolution de l'erreur quadratique moyenne (notée EQM dans la figure) en fonction du temps pour les 30 premières secondes de divers extraits musicaux. L'exercice est répété avec six extraits des diverses pièces musicales pour lesquels les noms des pièces, les noms des artistes ainsi que les noms des albums sur lesquels elles figurent, sont donnés en annexe A. Dans le but d'en montrer le contenu fréquentiel, cette annexe contient également le spectrogram de chacun des six extraits utilisés. On répète également l'exercice avec un bruit blanc. Le riche contenu spectral d'un bruit blanc nous permet d'observer la convergence maximale de l'algorithme de modélisation dans le cas où les spectres du signal musical sont extrêmement riches. La figure 5.31 permet de comparer la capacité de l'algorithme de modélisation à bien approximer la réponse de la salle au défilement des trames de musique pour ces six extraits et pour un bruit blanc.

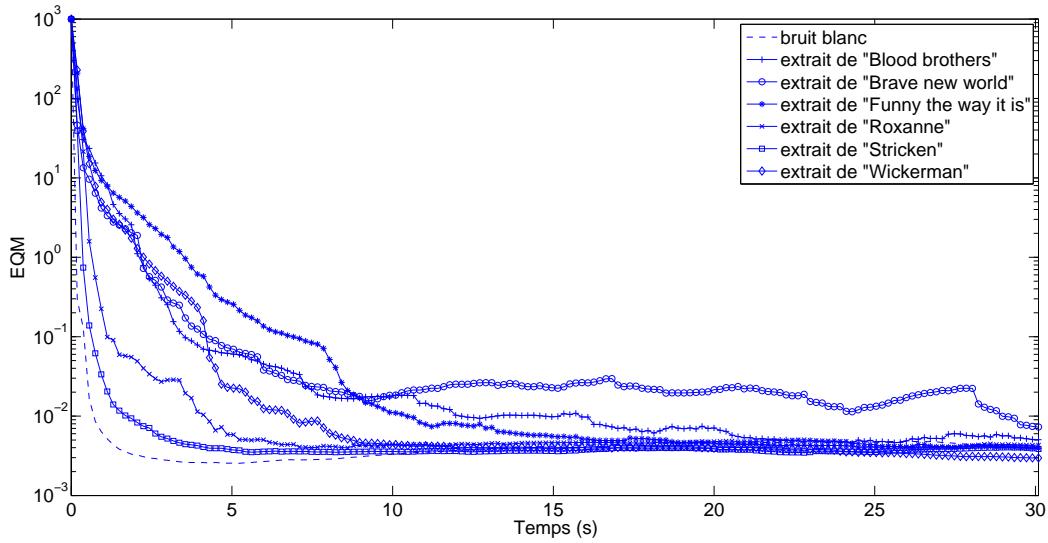


Figure 5.31 Erreur quadratique moyenne sur la réponse en fréquence lorsque le filtre est désactivé

On note tout d'abord que pour chaque extrait musical, l'erreur décroît rapidement durant les premières secondes. Durant cette période, chaque bloc de 8192 échantillons analysés permet une grande amélioration de l'estimation de la réponse en fréquence qui tend à s'approcher de plus en plus de la référence. On note que la convergence est plus rapide pour certains extraits musicaux. Dans les paragraphes qui suivent (figures 5.35 et 5.34), nous verrons que le contenu spectral d'extrait musical peut influencer la rapidité de cette convergence. On note en observant la figure 5.31, qu'après plusieurs secondes, l'erreur de l'ensemble des extraits musicaux converge vers la même valeur. Ce phénomène est dû à la capacité de l'algorithme de modélisation à reconstruire la réponse en fréquence au fil du temps, exploitant la richesse spectrale qu'ont les signaux musicaux sur une longue période, richesse qui ne se manifeste pas nécessairement ponctuellement à tous les endroits des signaux. Le fait que l'erreur pour l'ensemble des extraits musicaux converge vers la même valeur que celle du bruit blanc démontre bien cette capacité qu'a l'algorithme à reconstruire la réponse de la chaîne audio au fil du temps. Cependant, en observant l'évolution de l'erreur pour le bruit blanc, on note que la convergence est beaucoup plus rapide pour ce dernier qui contient, pour chaque trame de 8192 échantillons analysés, toutes les fréquences.

La méthode suggérée est donc satisfaisante et la réponse obtenue à l'aide de cette méthode est suffisamment similaire à celle obtenue à l'aide d'une méthode standard telle que le balayage sinusoïdal, pour servir à la conception d'un filtre égalisateur.

Dans les tests expérimentaux pour lesquels des résultats ont été présentés jusqu'à présent,

le filtre égalisateur était désactivé. Avant de s'intéresser à la performance du filtre égalisateur adaptatif et à son impact sur la réponse du système global, nous voulons nous assurer que l'activation du filtre égalisateur n'empêche pas l'algorithme de modélisation de faire son travail. Pour ce faire, nous répétons l'expérience ayant servi à générer les résultats de la figure 5.31 mais cette fois avec le filtre égalisateur adaptatif actif. Le but de cet exercice est de s'assurer qu'une convergence vers la réponse en fréquence de référence au fil du temps a lieu et donc que l'erreur quadratique moyenne sur l'estimation de la réponse en fréquence diminue au fil du temps.

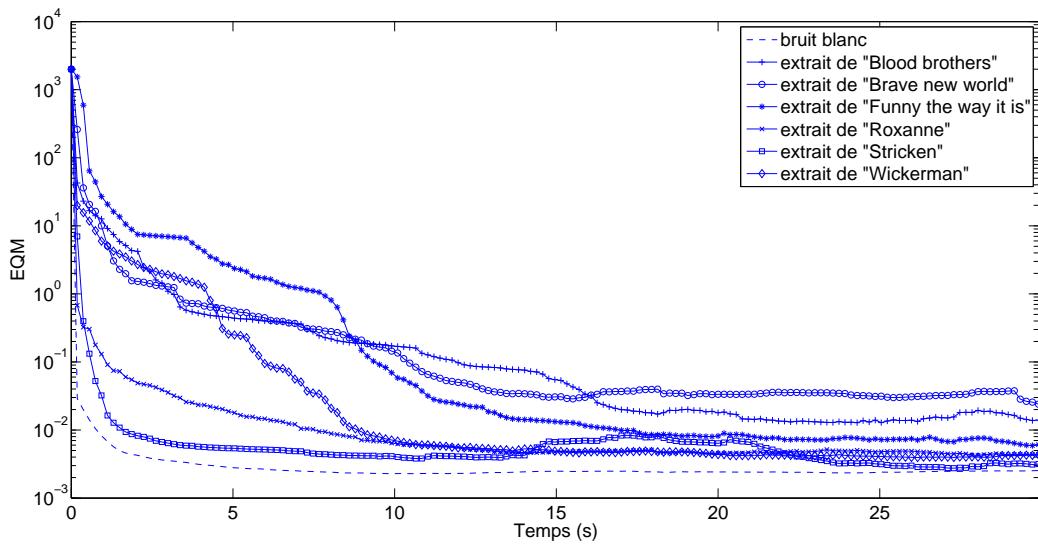


Figure 5.32 Erreur quadratique moyenne sur la réponse en fréquence lorsque le filtre est activé

Les résultats de la figure 5.32 démontrent qu'une diminution de l'erreur a toujours lieu et que la convergence est maintenue. Cependant, lorsque comparée aux résultats de la figure 5.31, on sent que cette convergence est légèrement plus lente. Afin de mieux évaluer l'impact de l'activation du filtre égalisateur sur la capacité de l'algorithme de modélisation à bien estimer la réponse de la salle, on utilise une moyenne arithmétique pour combiner l'ensemble des résultats pour les six extraits musicaux sans filtre, et la même moyenne pour combiner l'ensemble des résultats pour les six extraits musicaux avec filtre. La figure 5.33 montre les deux moyennes.

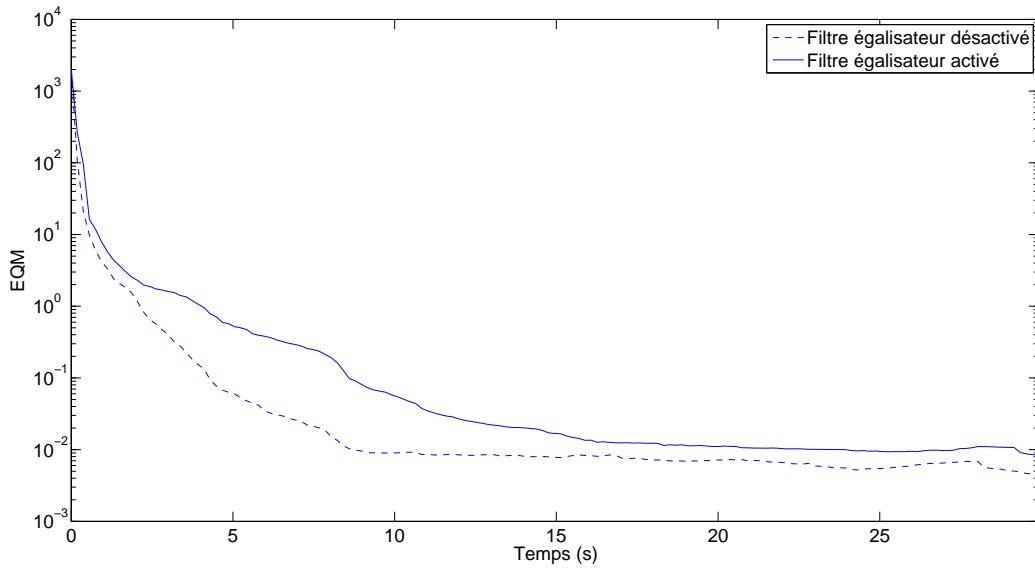


Figure 5.33 Impact de l'activation du filtre sur l'erreur quadratique moyenne sur la réponse en fréquence

On conclut que l'activation du filtre n'empêche pas le fonctionnement de l'algorithme de modélisation mais qu'il en ralentit légèrement la convergence. Le filtre fonctionne en principalement en atténuant les fréquences dont la transmission est favorisée par la chaîne audio. Cette atténuation de certaines fréquences résulte en une excitation moins importante de la chaîne audio pour ces fréquences qui ont été atténuées. Moins l'excitation procurée est grande, moins la MMEP décrite dans la section 4.5 est en mesure de procurer une mesure précise de l'atténuation pour cette fréquence, ce qui résulte en une erreur quadratique plus grandes pour ces fréquences. Finalement, nous savons, de par la nature de l'algorithme de modélisation, que la capacité de ce dernier à rapidement converger et bien estimer la réponse de la salle repose sur la richesse spectrale du signal musical qui le traverse. Il est donc intéressant de tracer, dans une même figure, l'erreur quadratique moyenne sur la réponse en fréquence (sans et avec filtre égalisateur actif) ainsi que le spectrogramme du signal musical ayant servi de signal d'excitation à ce moment. En plus, la figure 5.34 montre l'évolution de l'erreur pour un bruit blanc, ce qui correspond à la performance maximale de l'algorithme dans le cas où le signal musical contiendrait ponctuellement toutes les fréquences.

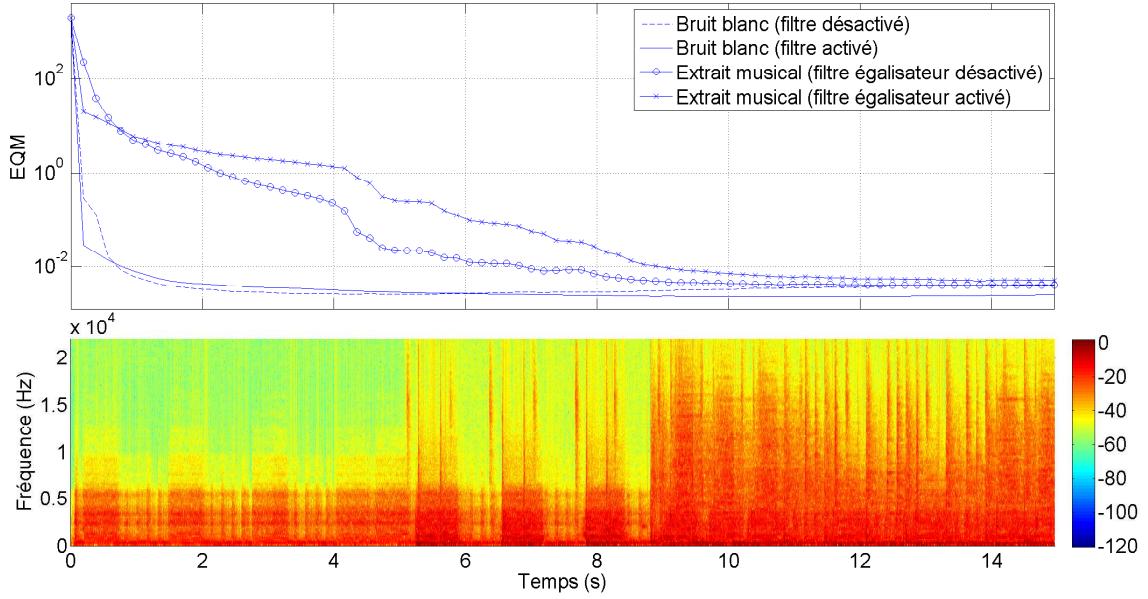


Figure 5.34 EQM (sans et avec filtre) accompagnée du spectrogramme de l'extrait de la pièce « Wikerman »

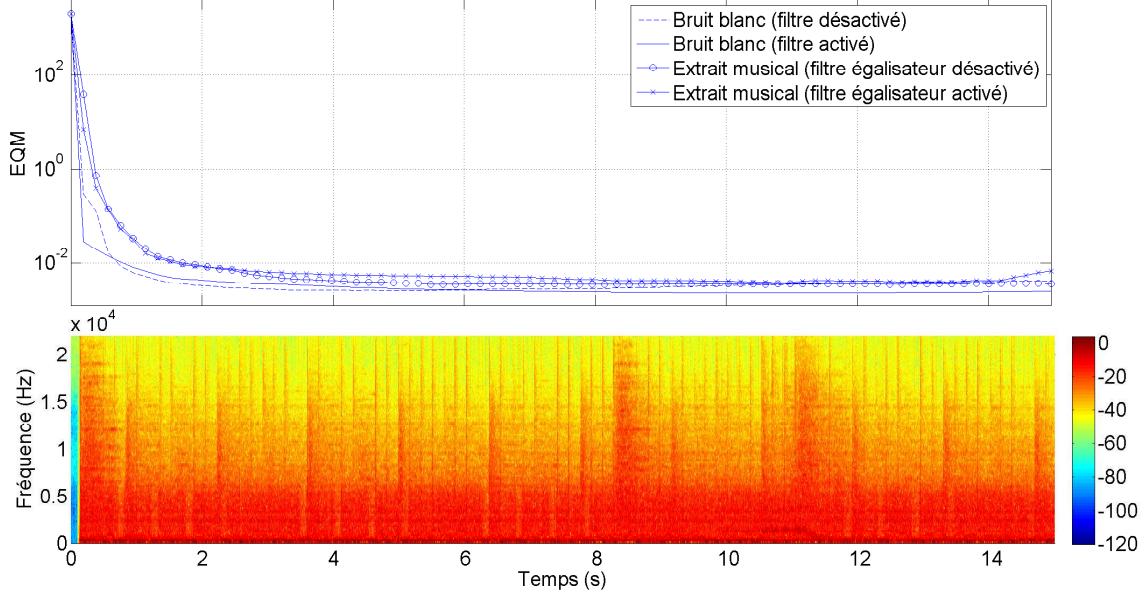


Figure 5.35 EQM (sans et avec filtre) accompagnée du spectrogramme de l'extrait de la pièce « Stricken »

La figure 5.34 nous permet également d'observer que les régions riches en contenu fréquentiel accélèrent la convergence de l'algorithme de modélisation. Ceci est autant valide lorsque le filtre égalisateur est désactivé que lorsqu'il est activé. La figure 5.35 démontre que

la convergence est accélérée lorsque l'extrait musical utilisé comme signal d'excitation est très riche en fréquences. Pour cet extrait musical, la performance de l'algorithme de modélisation s'approche de celle observée lorsqu'un bruit blanc gaussien est utilisé. En comparant les deux spectrogrammes (5.34 et 5.35), on note que les premières 15 secondes de la pièce « Stricken » sont plus riches en fréquences que celles de « Wikerman ». L'annexe A contient les résultats pour l'ensemble des 6 autres extraits musicaux.

Finalement, la composition spectrale des signaux musicaux joués ne constitue pas l'unique facteur potentiellement capable d'influencer la performance du système proposé. La distance entre la source et la position d'écoute doit elle aussi être considérée. L'ensemble des tests ayant servi à générer les résultats présentés précédemment ont été effectués avec un espacement d'un mètre entre le microphone de mesure et la source. La figure 5.36 montre une comparaison de l'évolution de l'erreur sur l'estimation de la réponse en fréquence pour le test à 1 mètre et un second test effectué à deux 2 mètres sur le même échantillon de musique.

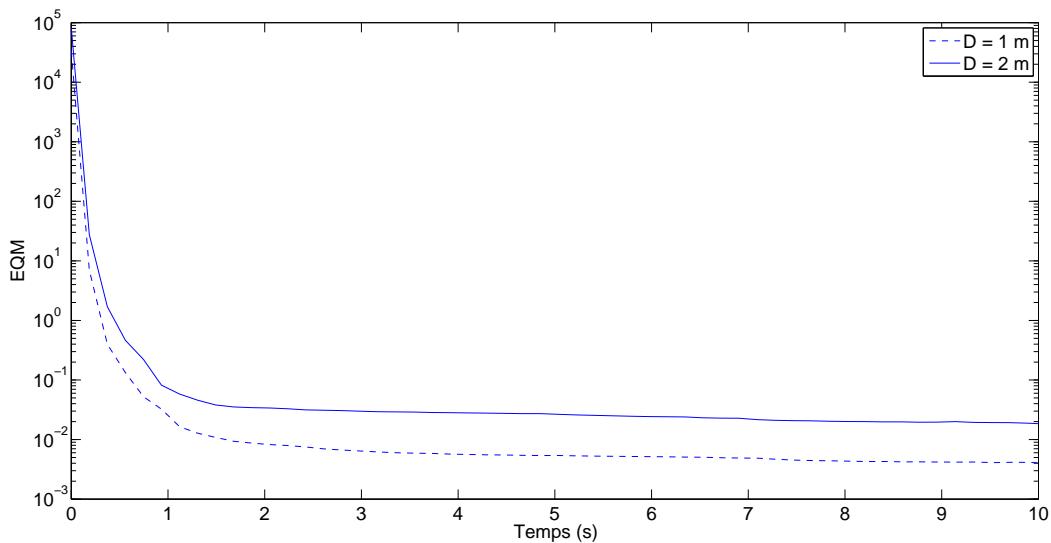


Figure 5.36 Effet de la distance entre le haut-parleur et la position d'écoute sur la capacité à estimer la salle

Les résultats obtenus sont très satisfaisants et montrent que l'algorithme continue de converger même lorsqu'on laisse le champ réverbéré gagner en importance en éloignant le microphone de la source. La plage $D \in [1 \text{ m}, 2 \text{ m}]$ couvre les positions d'écoute typiques dans le cas de la salle de régie du Studio CME à l'étude dans ce travail.

5.10 Performance du filtre égalisateur

La présente section contient les résultats expérimentaux relatifs au filtre égalisateur adaptatif. Alors que la section précédente (5.9) se limite à l'algorithme de modélisation et la validation de sa capacité à approximer la réponse d'une chaîne audio, la présente section s'intéresse à l'algorithme utilisant cette réponse pour calculer quelle devrait être la réponse du filtre égalisateur permettant de la corriger. On présente également les résultats obtenus après l'égalisation, l'avant/après.

L'algorithme de conception du filtre égalisateur adaptatif pour lequel les résultats expérimentaux sont présentés dans cette section est représenté par le bloc « Conception du filtre égalisateur » dans la figure 1.5. La figure 4.26 montre l'implémentation de ce bloc. La première étape consiste à passer la réponse impulsionale de la chaîne audio dans un banc de filtres assurant sa décomposition par bandes de fréquence. La motivation derrière l'utilisation d'une telle approche est expliquée dans la section 4.8 du chapitre 4. Le tableau 4.1 résume l'ensemble des paramètres pour chacune des bandes. Les figures 5.37 à 5.46 montrent le résultat de la convolution de la réponse en fréquence de la chaîne audio $h[n]$ par chacun de ces filtres et la fenêtre ensuite appliquée dans le but d'éliminer la réverbération tardive du signal.

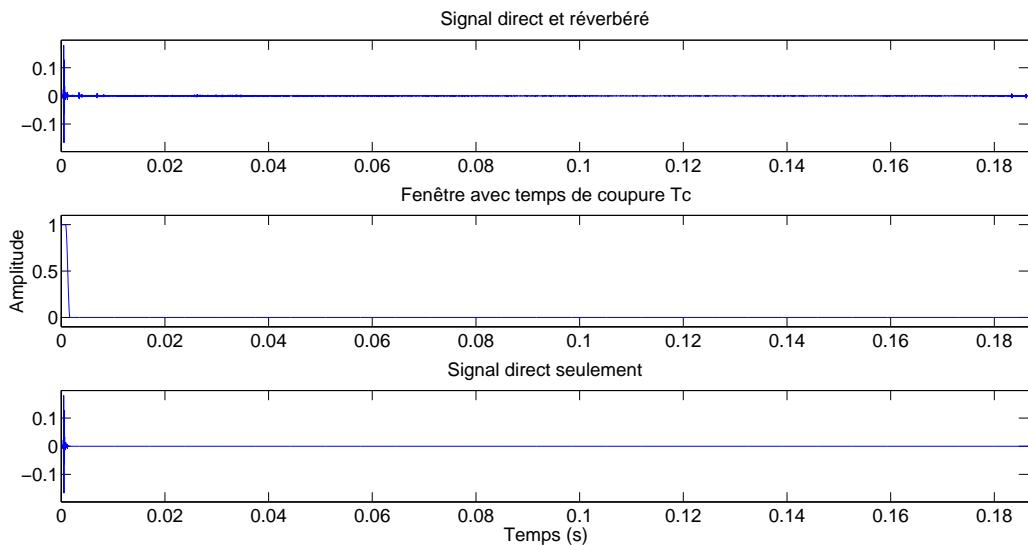


Figure 5.37 Fenêtrage de la réponse impulsionale pour la bande A (9454 à 22050 Hz)

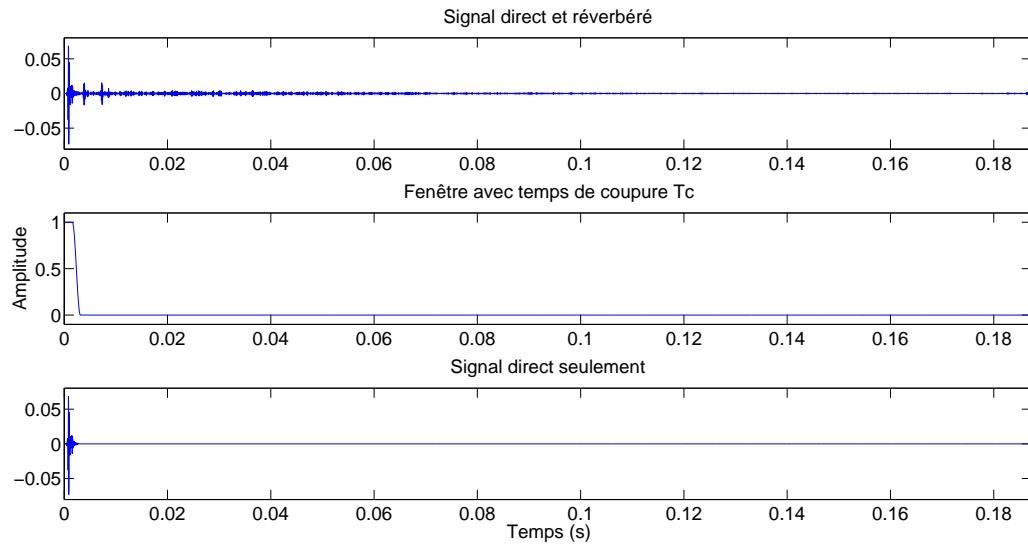


Figure 5.38 Fenêtrage de la réponse impulsionale pour la bande B (4663 à 9454 Hz)

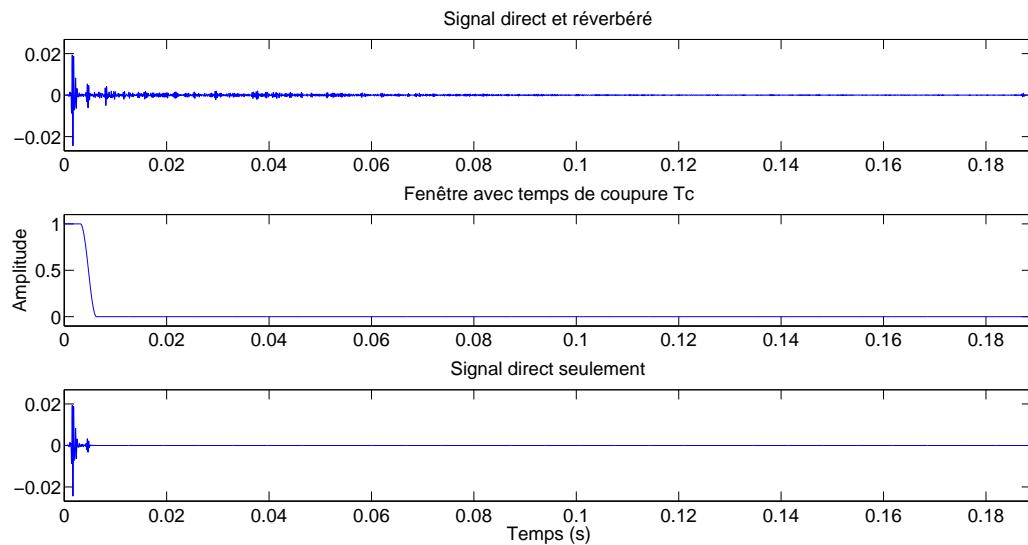


Figure 5.39 Fenêtrage de la réponse impulsionale pour la bande C (2318 à 4663 Hz)

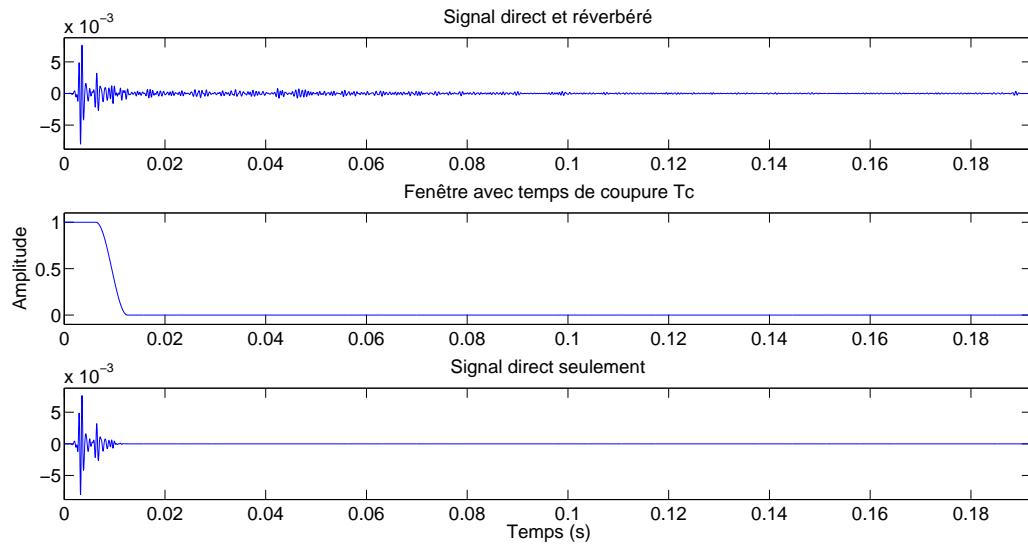


Figure 5.40 Fenêtrage de la réponse impulsionale pour la bande D (1158 à 2318 Hz)

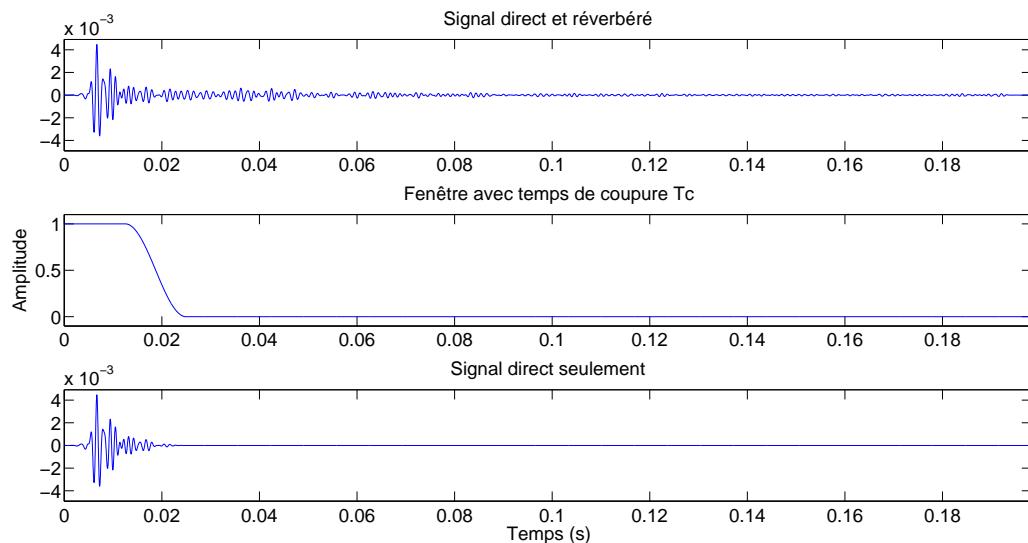


Figure 5.41 Fenêtrage de la réponse impulsionale pour la bande E (580.6 à 1158 Hz)

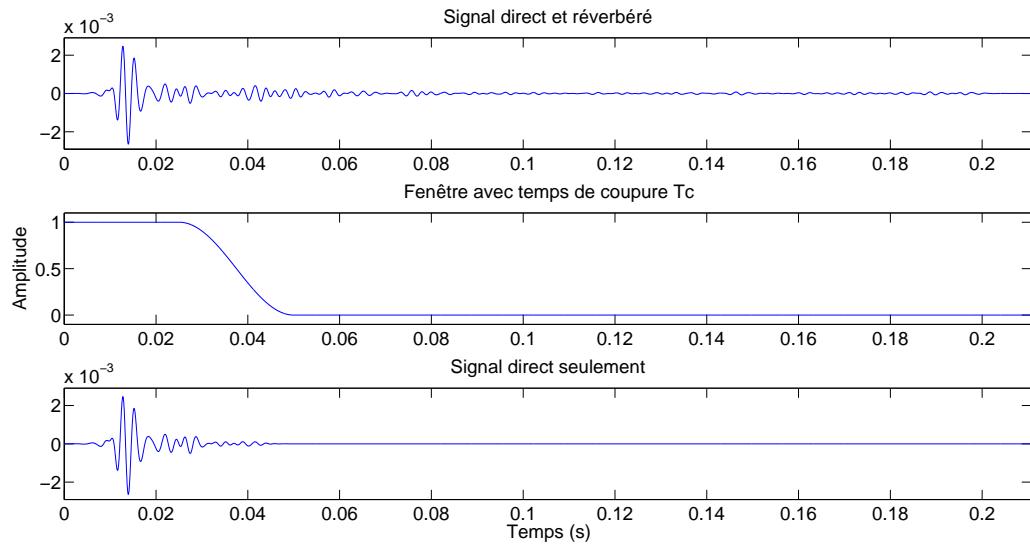


Figure 5.42 Fenêtrage de la réponse impulsionale pour la bande F (292.7 à 580.6 Hz)

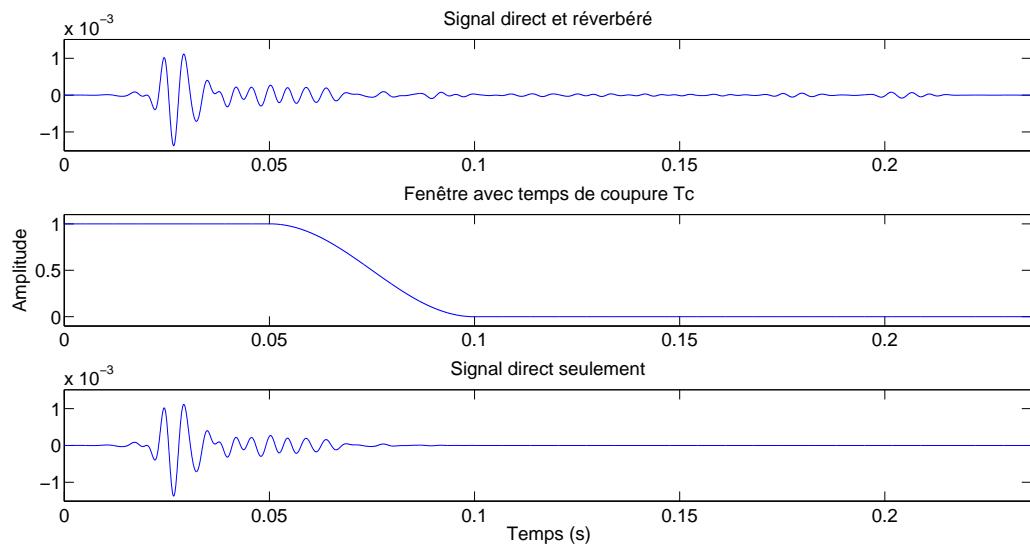


Figure 5.43 Fenêtrage de la réponse impulsionale pour la bande G (149.0 à 292.7 Hz)

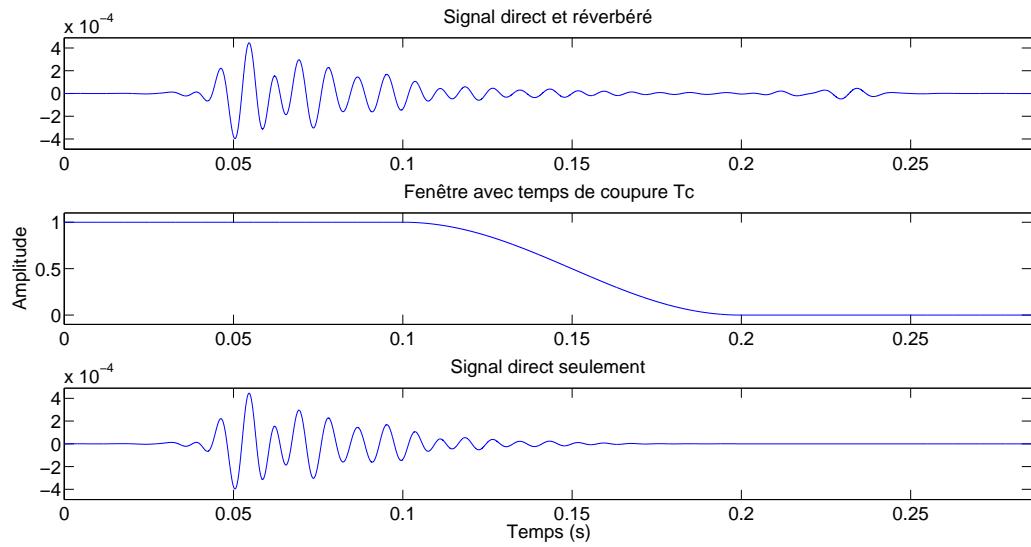


Figure 5.44 Fenêtrage de la réponse impulsionale pour la bande H (77.12 à 149.0 Hz)

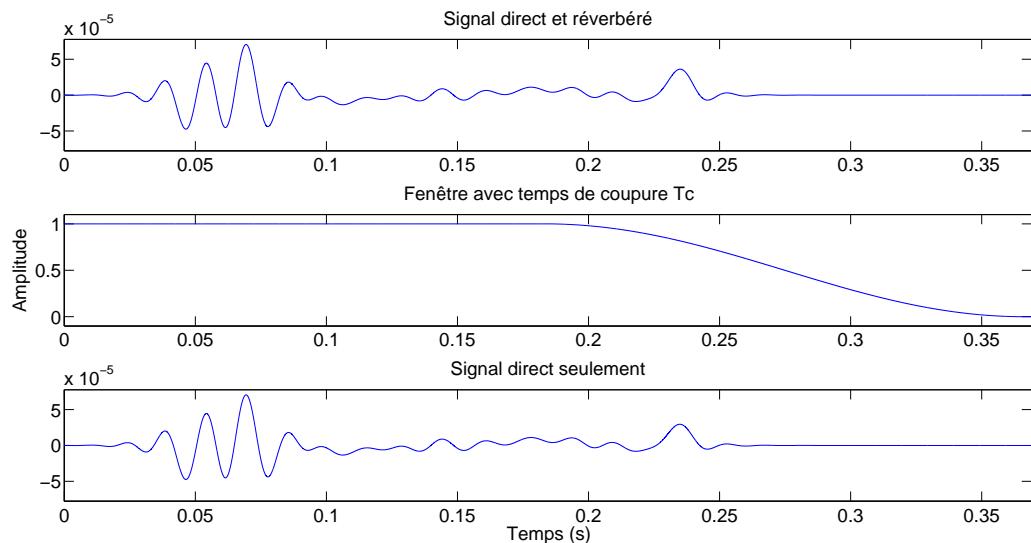


Figure 5.45 Fenêtrage de la réponse impulsionale pour la bande H (41.25 à 77.12 Hz)

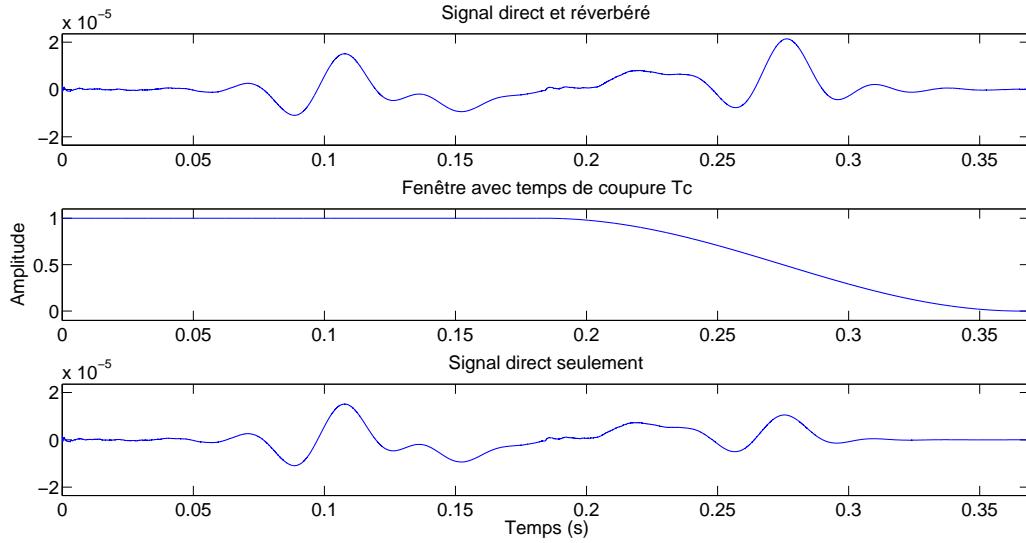


Figure 5.46 Fenêtrage de la réponse impulsionale pour la bande H (0 à 41.25 Hz)

On remarque que, pour l'ensemble de ces bandes, le comportement est tel que prescrit dans la section 4.8. La première impulsion, celle du signal direct, ainsi que les premières réflexions, sont conservées pour chaque bande. La réverbération tardive est cependant éliminée pour chaque bande. Ensuite, les réponses fenêtrées pour l'ensemble des dix bandes sont réalignées et sommées.

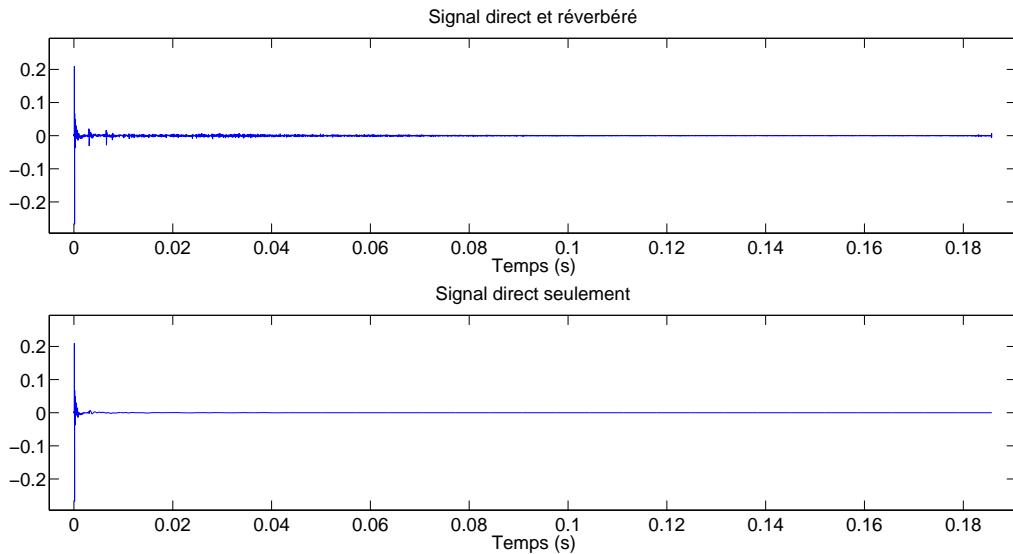


Figure 5.47 Réponse impulsionale complète (direct et réverbéré) et réponse impulsionale après fenêtrage par bandes d'octave (direct seulement)

La figure 5.47 montre la réponse impulsionnelle $h[n]$ avant qu'elle ne soit passée à travers le banc de filtres (signal direct et réverbéré) et après (signal direct seulement). Tel que vu dans la section 2.1, on utilise ici le terme « direct » lorsqu'on fait référence à la portion de la réponse impulsionnelle de la salle que l'on veut égaliser. Par « direct » on entend la combinaison du signal direct et des premières réflexions, excluant la réverbération tardive.

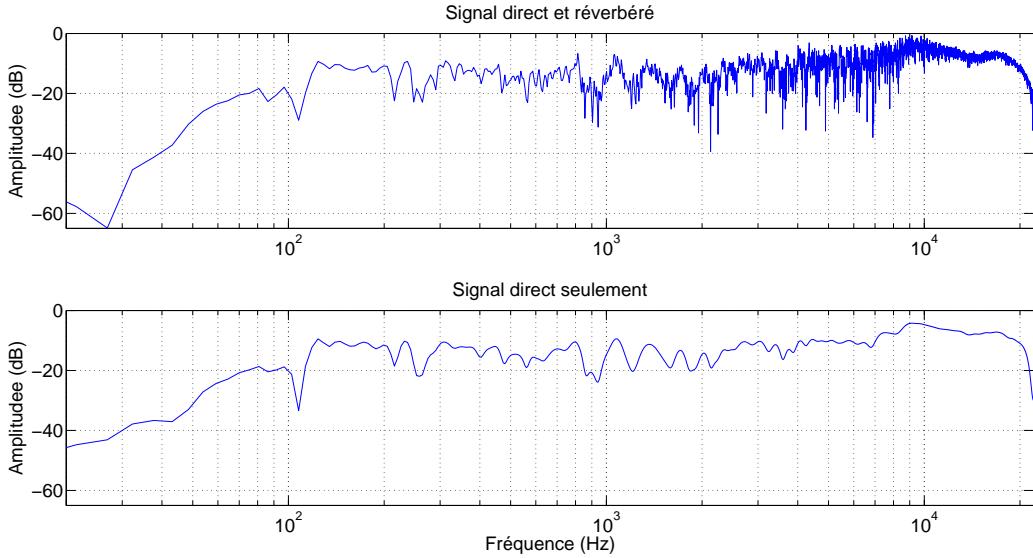


Figure 5.48 Réponse en fréquence complète (direct et réverbéré) et réponse en fréquence après fenêtrage par bandes (direct seulement)

La figure 5.48 montre la réponse en fréquence avant et après le passage de cette dernière à travers le banc de filtres, obtenue en appliquant une transformée de Fourier discrète aux deux réponses impulsionales de la figure 5.47.

Dans la section 4.8, on explique pourquoi seul le signal direct ainsi que les premières réflexions devraient être corrigés par un filtre égalisateur lorsque le but est d'améliorer la qualité de l'écoute musicale sur haut-parleur dans une salle. La réponse impulsionnelle de laquelle la réverbération tardive fut éliminée par la fenêtre par bandes de fréquence est inversée afin de créer la réponse en impulsion du filtre égalisateur. Cette inversion ($1/x$) a lieu dans le domaine fréquentiel à l'aide d'une division complexe, le code protège contre la division par zéro. Afin d'éviter la situation où le filtre tenterait d'amplifier exagérément les extrémités du spectre à cause de l'incapacité du haut-parleur utilisé à transmettre ces très hautes ou très basses fréquences, l'algorithme fut implémenté avec deux paramètres permettant de définir la plage de fréquence sur laquelle l'égalisation prend lieu. À l'extérieur des ces limites, la réponse en fréquence du filtre reste constante. La limite inférieure était de 60 Hz et celle supérieure de 18000 Hz lors du test ayant mené à l'obtention du filtre

dont la réponse en fréquence est montrée par la figure 5.49. Cette plage est, par mesure de précaution, légèrement plus restrictive que celle supportée par les haut-parleurs utilisés soit de 50 à 21000 Hz.

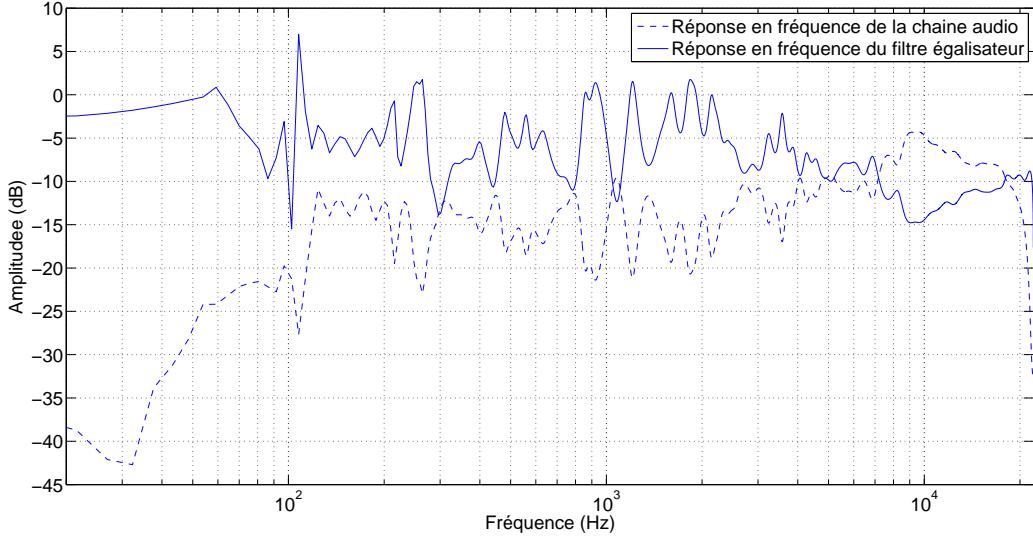


Figure 5.49 Réponse en fréquence de la salle (direct seulement) et réponse en fréquence du filtre égalisateur

La section 4.9 couvre en détail comment un filtre convolutif de réponse relativement longue, soit 8192 échantillons dans notre cas, peut être réalisé efficacement et avec un minimum de délai entre l'entrée et la sortie.

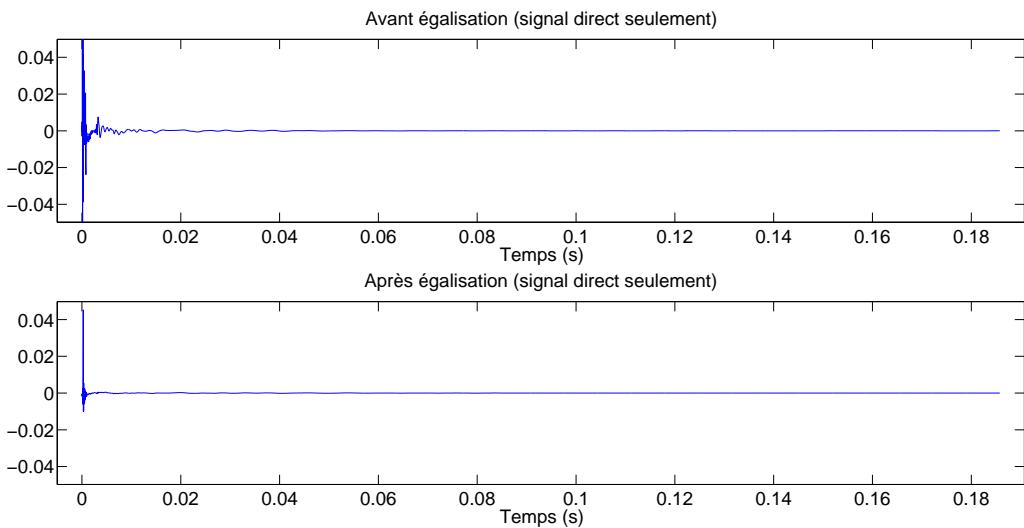


Figure 5.50 Réponse impulsionnelle de la salle (direct seulement) avant et après égalisation

Le filtre égalisateur convolutif réalisé exploite l'algorithme « overlap-add » pour calculer la convolution linéaire du fil de donnée audio avec la réponse impulsionnelle du filtre avec un délai de 1024 échantillons entre l'entrée du filtre et sa sortie, soit 0,0232 seconde. Ce filtre procure donc l'effet d'égalisation pour la portion composée du champ direct et des premières réflexions de la réponse impulsionnelle de la chaîne audio. La figure 5.50 montre la réponse impulsionnelle avant et après l'égalisation.

La réponse en fréquence avant et après l'égalisation permet de valider que l'effet d'égalisation désiré est bel et bien obtenu ; la réponse en fréquence devient plus plate qu'elle ne l'était avant l'égalisation.

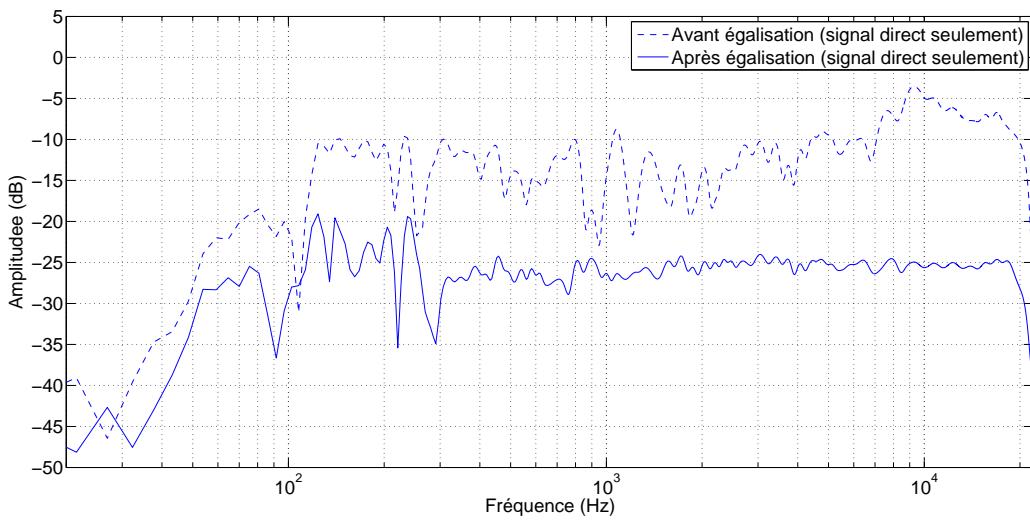


Figure 5.51 Réponse en fréquence de la salle (direct seulement) avant et après égalisation

Bien que l'égalisation ne tente d'adresser que le champ direct et les premières réflexions de la réponse en fréquence et non la totalité de celle-ci, il peut être intéressant d'observer l'effet de l'égalisation sur la pleine réponse en fréquence pour voir si celle-ci est aussi aplatie ou pas. La technique suggérée pourrait être remise en question si elle amplifiait exagérément une partie du spectre et que l'effet se fasse entendre dans la partie réverbérée de la réponse impulsionnelle. Ce ne semble pas être le cas. La figure 5.52 montre que même la pleine réponse en fréquence (incluant la réverbération tartive) se voit aplatie par le filtre égalisateur.

L'équation 5.2 de la section 5.1 suggère l'utilisation de l'erreur quadratique moyenne afin de quantifier l'aplatissement de la réponse en fréquence suite à l'activation du filtre égalisateur. Cette équation permet de mesurer la distance entre la réponse en fréquence et sa propre moyenne sur l'ensemble de sa longueur. Plus la réponse est plate, plus l'erreur quadratique moyenne entre celle-ci et sa propre moyenne diminue. Nous utilisons ce critère pour évaluer

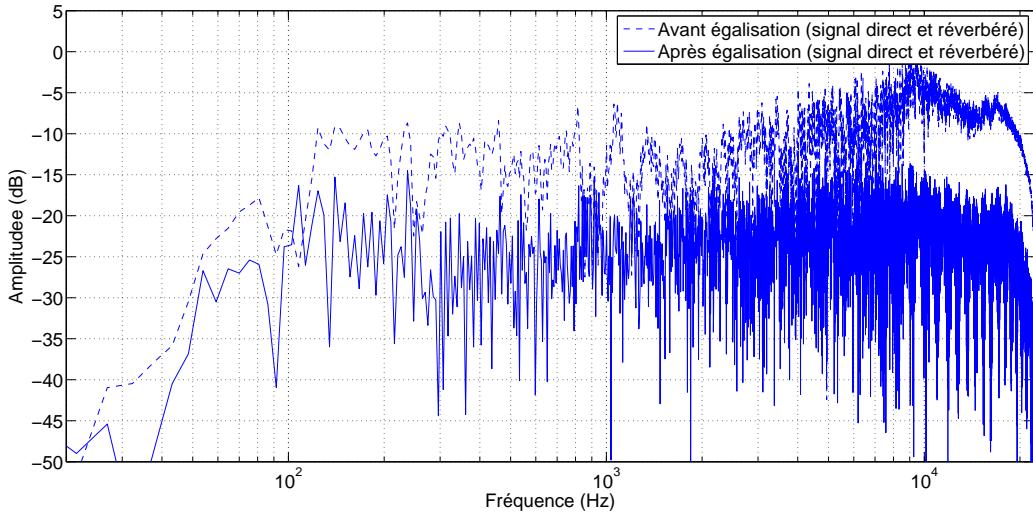


Figure 5.52 Réponse en fréquence de la salle (direct et réverbéré) avant et après égalisation

la performance globale de la solution suggérée pour l'égalisation adaptative. Cette métrique évolue naturellement au fil du temps à mesure que l'estimation de la réponse de la salle fournie par l'algorithme de modélisation s'approche de la vraie réponse et que le filtre s'y adapte dans le but d'assurer l'égalisation. Puisque l'objectif est de ne pas tenter d'égaliser la réverbération tardive, celle-ci est exclue du calcul de l'erreur quadratique moyenne. La réponse en fréquence après égalisation du signal direct seulement de la figure 5.51 sert au calcul de l'erreur. La figure 5.53 montre l'évolution de l'erreur pour les 35 premières secondes suivant l'activation du filtre pour six extraits musicaux.

Une diminution importante de l'erreur dans les premières secondes et qui converge au fil du temps nous permet de conclure que le filtre adaptatif égalise avec succès la réponse en fréquence de la salle à mesure que l'estimation fournie par l'algorithme de modélisation s'affine.

Un test à 2 mètres est nécessaire pour valider que la méthode suggérée converge sur la plage de distances de $D \in [1 \text{ m}, 2 \text{ m}]$ couvrant les positions d'écoute typiques dans la salle à l'étude. L'extrait de la pièce « Stricken » est utilisé.

Les résultats obtenus montrent une diminution de l'erreur et donc une amélioration de la réponse en fréquence suite à l'introduction du filtre égalisateur. Cependant, la performance est moins satisfaisante à une distance de 2 mètres comparativement aux résultats obtenus à un mètre. Bien que la performance soit satisfaisante dans le contexte pour lequel il fut développé, l'algorithme pourrait bénéficier d'un mécanisme supplémentaire assurant une meilleure immunité à la distance entre la source et la position d'écoute.

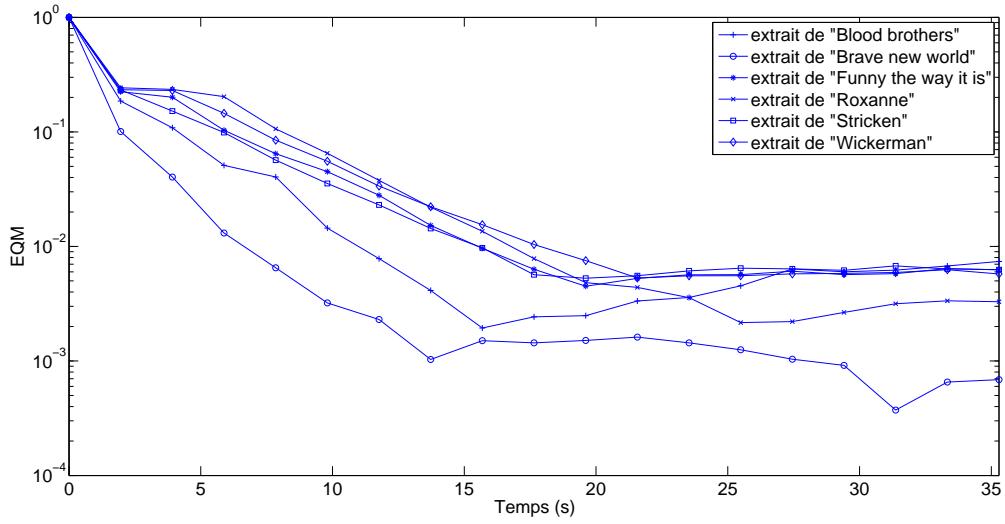


Figure 5.53 EQM entre la réponse en fréquence de la salle (direct seulement) et une réponse idéale (plate)

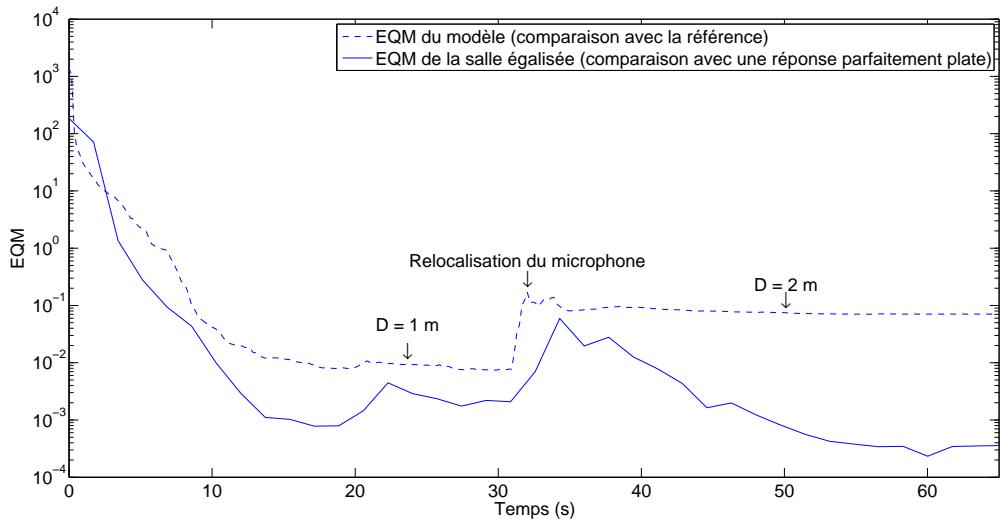


Figure 5.55 Capacité d'adaptation de la correction de la réponse de la salle lorsque celle-ci change soudainement

Le but ultime d'avoir un filtre égalisateur adaptatif était d'assurer l'égalisation de la réponse de la chaîne audio en tout temps, notamment même lorsque celle-ci est altérée durant l'écoute. La méthode suggérée fut conçue à cet effet. La reconstruction de la réponse en fréquence de la salle au fil du temps repose sur une moyenne mobile exponentielle pondérée dotée d'une constante de temps α permettant d'ajuster la dynamique avec laquelle le filtre

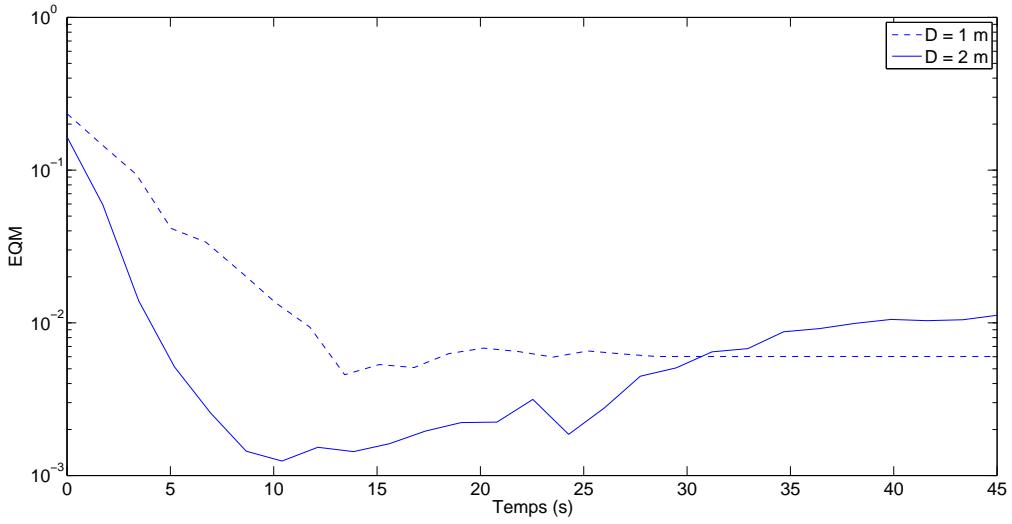


Figure 5.54 Effet de la distance entre le haut-parleur et la position d'écoute sur la capacité à égaliser la salle

s'adapte à un changement. Une façon simple dans notre cas d'altérer subitement la réponse de la salle durant l'écoute est de déplacer rapidement (manuellement) le microphone au milieu d'un test.

La figure 5.55 montre à la fois l'erreur sur l'estimation de la réponse en fréquence ainsi que la comparaison entre la réponse égalisée et une réponse parfaitement plate. Après 31 secondes d'écoute musicale, le microphone orginalement placé à 1 mètre de la source fut subitement déplacé à deux mètres de la source. Naturellement on ne s'attend pas à ce que l'erreur sur l'estimation de la réponse en fréquence diminue suite à la relocalisation du microphone puisque celle-ci se réfère à une réponse en fréquence obtenue préalablement par balayage sinusoïdal au moment où 1 mètre séparait la source du microphone. L'erreur sur l'estimation de la réponse en fréquence est simplement tracée pour aider à visualiser le déplacement du microphone qui se manifeste alors comme une augmentation irrécupérable de cette erreur se produisant après 31 secondes. L'erreur sur la correction de la réponse de la salle, elle, devrait être perturbée au moment de la relocalisation mais devrait recommencer à diminuer dès que le système est au repos. La figure 5.55 montre qu'il faut 11 secondes suite à la relocalisation du microphone à la réponse en fréquence égalisée pour atteindre la performance enregistrée au moment de la perturbation. On voit dans la section 4.5 que le choix d' α permet d'ajuster la dynamique du système. Un α petit permet au système de réagir plus rapidement, cependant en échange d'une perte de robustesse et d'une moins bonne capacité à bien estimer la réponse de la salle lorsque le contenu spectral de la musique jouée varie au fil du temps.

CHAPITRE 6

APPLICATION DE LA MÉTHODE AU PROBLÈME DE L'ACOUSTIQUE DANS UN BROYEUR

La méthode présentée à la section 4.5 reposant sur la déconvolution circulaire d'un signal quelconque en entrée d'un système de celui obtenu en sortie grâce à un microphone ou autre type de capteur peut être exploitée dans plusieurs contextes. Cette section présente deux publications issues de la techniques de modélisation présentée. La chaîne audio considérée est similaire à celle de la figure 1.5, à l'exception que le signal d'excitation fut procuré à l'aide d'un marteau à impact muni d'un capteur de pression plutôt qu'à l'aide d'un signal audio émis sur haut-parleur comme la figure le suggère. Dans les travaux aillant mené à ces deux publications, seul l'aspects modélisation était considéré. Le bloc de « conception du filtre égalisateur » et celui du « filtre EQ adaptatif » de la figure 1.5 étaient désactivés.

La solution aux problèmes fondamentaux de la modélisation expérimentale abordée dans le chapitre 1 est commune à plusieurs domaines d'application. Les techniques de base suggèrent l'utilisation de signaux dédiés à la mesure. Ces signaux d'excitation ont des propriétés mathématiques simplifiant le calcul de réponse impulsionale ou de réponse en fréquence. C'est le cas de la parfaite impulsion, d'une sinusoïdale balayant la pleine plage de fréquence, ou d'un bruit blanc de puissance constante sur cette même plage par exemple. Cependant, il n'est pas rare que le contexte ne permette pas l'utilisation de tels signaux. Par exemple, dans le cadre du présent travail, les contraintes imposées ne permettaient que l'exploitation de signaux musicaux pour exciter le système dans le but d'en extraire sa réponse à l'impulsion. Le contenu fréquentiel de bien des signaux d'excitation du monde réel n'est pas constant. Une méthode permettant l'obtention de la réponse impulsionale est donc suggérée et validée dans ce travail. Cette méthode vue dans la section 4.5 repose sur la déconvolution du signal observé à la sortie d'un système de celui ayant servi à l'exciter. Dans ce cas, une approximation de la déconvolution linéaire reposant sur la déconvolution circulaire calculée dans le domaine fréquentiel fut utilisée. Nous avons vu dans la section 3.3 qu'il est acceptable de considérer la déconvolution circulaire comme une bonne approximation de la déconvolution linéaire.

Cette même technique ne se limite pas forcément à la caractérisation d'une chaîne audio mais peut très bien servir dans d'autres cas où le signal d'excitation utilisé pour caractériser un système souffre d'imperfections. C'est le cas lorsqu'on utilise un marteau dans le but d'exciter avec une quasi impulsion un système mécanique. Un coup de marteau ne générant

jamais exactement précisément la même pression lorsqu'un impact survient, il est nécessaire de déconvoluer le signal de pression durant l'impact du signal observé par exemple par un microphone placé non loin si on désire établir la relation entre une force d'impact sur une surface et le signal acoustique observé en champs proche. Certains marteaux sont munis d'un capteur de pression permettant de les raccorder à une carte de conversionn analogique-numérique afin de stocker une version échantillonnée du signal de pression qu'il génère durant l'impact.

Dans le cadre d'un travail mené conjointement avec le département de génie mécanique de l'Université McGill à Montréal, on s'est intéressé à la méthode par laquelle la relation entre un seul impact, à l'intérieur de la coquille métallique d'un moulin servant au traitement des minéraux à l'état brut. Pour l'expérience, un broyeur à billes de laboratoire est utilisé. Un disque de grand diamètre en aluminium est fixé sur un arbre monté sur un palier. Le tambour du broyeur (ou coquille) est boulonné sur ce disque. Une face en plexiglas transparent fermé permet l'observation de ce qui se passe à l'intérieur de la coquille. Le tambour est constitué d'un cylindre en acier, d'un diamètre de 1,5 m et d'une longueur de 0,3 m. Un ensemble de douze plaques sont fixées à la surface intérieure de la coque, ces plaques sont appelées « releveurs ». Leur rôle est de promouvoir l'action de culbutage de la charge, comme induit par la rotation du moulin. Les billes, qui forment la charge, ont un diamètre moyen de 5,1 cm et un poids moyen de 43 g. La figure 6.1 montre le broyeur utilisé pour l'expérience.

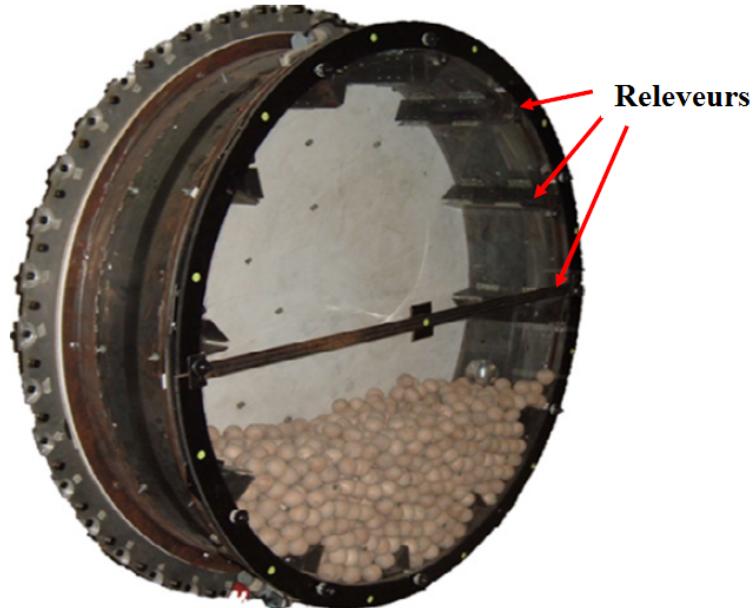


Figure 6.1 Broyeur utilisé pour l'expérience (figure tirée de Hosseini *et al.* (2011))

Une relation liant une force d'impact à une position spécifique à l'intérieur de la coquille métallique au signal acoustique en champ proche à l'extérieur de la coque devait être établie. La force de l'impact et celle du signal acoustique résultant sont mesurées à l'aide d'un marteau d'impact avec capteur de force intégrée et un microphone de champ de pression, respectivement. La figure 6.2 montre la position du microphone de mesure, les endroits où des impacts faits à l'aide du marteau sont numérotés.

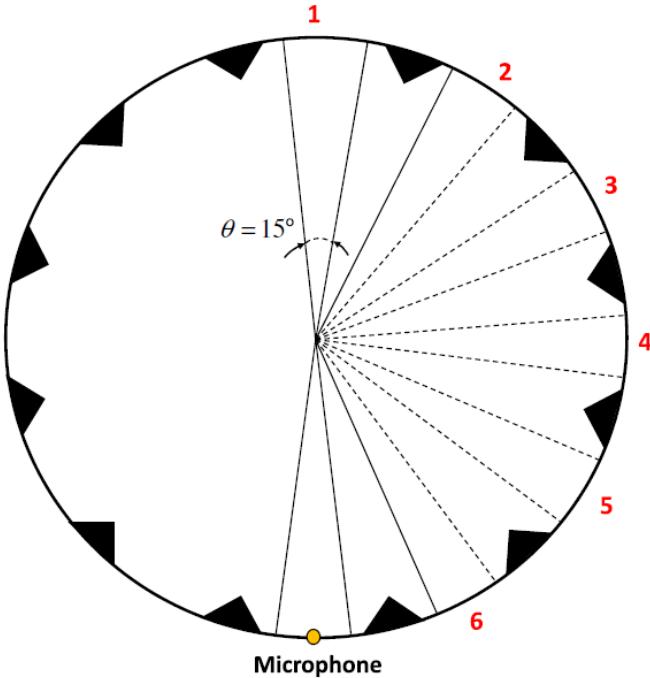


Figure 6.2 Position du microphone et endroits percutés avec le marteau (figure tirée de Hosseini *et al.* (2011))

Un système d'acquisition de données multicanal est utilisé pour amplifier simultanément et numériser les deux signaux et pour diffuser les signaux résultants à un ordinateur. La figure 6.3 représente un échantillon du signal de force d'impact avec le signal de pression acoustique résultante.

En supposant que les vibrations de la coque sont linéaires et invariantes dans le temps, ce qui signifie des effets différents sur la même position à différents moments de produire la même réponse, la réponse de la coquille à la force d'impact peut être caractérisée par la réponse impulsionale du système $h(t)$. Par conséquent, la relation entre la pression acoustique $p(t)$ et la force d'impact $i(t)$ peut être exprimée en utilisant le théorème de la convolution.

$$p[n] = h[n] *_{lin} i[n] \quad (6.1)$$

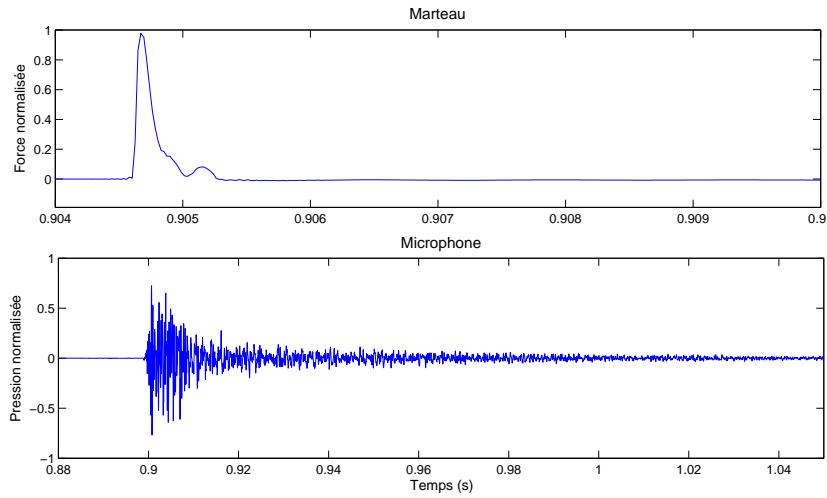


Figure 6.3 Force d'impact du marteau et pression acoustique résultante

où $h[n]$, $p[n]$ et $i[n]$ sont les versions échantillonnées de $h(t)$, $p(t)$ et $i(t)$ respectivement.

En observant la force d'impact du marteau de la figure 6.3, il est évident qu'il ne s'agit pas d'une impulsion pure et que le passage de la réponse impulsionnelle du système passe forcément par la déconvolution \div_{circ} du signal d'excitation.

$$\begin{aligned} h[n] &= p[n] \div_{lin} i[n] \\ &\approx p[n] \div_{circ} i[n] = \mathcal{T}\mathcal{FDI}\left(\frac{\mathcal{T}\mathcal{FD}(p)}{\mathcal{T}\mathcal{FD}(i)}\right) \end{aligned} \quad (6.2)$$

Comme pour la chaîne audio, il est possible d'utiliser la déconvolution circulaire calculée dans le domaine fréquentiel pour approximer la déconvolution linéaire et par le fait même obtenir la réponse impulsionnelle liant la force d'impact exercée par le marteau sur la coque à la pression acoustique captée par le microphone.

Cette technique fut utilisée et documentée dans un article scientifique (Hosseini *et al.*, 2011) publié dans la revue Minerals Engineering, en octobre 2011. Les travaux furent réalisés et l'article fut rédigé conjointement par Poorya Hosseini, Sudarshan Martins, et Dr Peter Radziszewski du département de génie mécanique de l'Université McGill, et Dr François-Raymond Boyer du département de génie informatique et génie logiciel de l'École polytechnique de Montréal. Un article de conférence (Martins *et al.*, 2011) fut également présenté à la conférence SAG à Vancouver en 2011.

CHAPITRE 7

CONCLUSION

7.1 Synthèse des travaux

Les techniques classiques dites « invasives » pour la modélisation expérimentale de l’acoustique d’une salle ont le désavantage d’utiliser des signaux d’excitation dédiés à la mesure. Par ce fait même, l’exercice de modélisation cause problème puisque ces mesures ne peuvent pas être prises en présence d’un auditoire et l’exercice doit être recommencé lorsque l’acoustique de la salle est changée suite à la reconfiguration de la salle, par exemple. Un algorithme capable de modéliser l’acoustique de la salle à l’insu de l’auditoire fut développé. Cette technique innovatrice repose sur l’excitation procurée par les signaux musicaux lors de l’écoute sur haut-parleur afin de modéliser la salle.

La décomposition du signal musical en blocs et la déconvolution de chaque bloc du signal musical émis sur haut-parleur au bloc du même signal capté par un microphone placé dans l’environnement d’écoute permet le calcul de la réponse à un instant donné. La déconvolution circulaire calculée dans le domaine fréquentiel sert à approximer efficacement la déconvolution linéaire qui est mal adaptée à l’application temps-réel. La méthode est utilisée dans des travaux réalisés dans le cadre d’une collaboration entre des étudiants et professeurs de l’École polytechnique de Montréal et de l’Université McGill ayant mené à une co-publication dans une revue à comité de lecture (Hosseini *et al.*, 2011) ainsi qu’un article de conférence (Martins *et al.*, 2011).

Une moyenne mobile exponentielle pondérée fut utilisée pour combiner des réponses individuelles afin de reconstruire statistiquement la réponse de la salle durant l’écoute. Cette méthode exploite le riche contenu fréquentiel des signaux musicaux. Résultat : la résolution spectrale du modèle obtenu (4097 points) est nettement supérieure à celle obtenue par d’autres techniques permettant de caractériser la salle à l’insu de l’auditoire telles que l’insertion de tonalités masquées (20 points). Cette résolution spectrale fait de la méthode suggérée une méthode mieux adaptée à l’égalisation adaptative de la réponse de salles. L’erreur quadratique moyenne fut utilisée pour quantifier la capacité de la méthode non invasive suggérée à approximer la réponse de la salle. La réponse en fréquence obtenue à l’aide de la méthode classique du balayage sinusoïdal fut utilisée comme référence.

Un partenariat fut formé avec un studio d’enregistrement de la région de Montréal. La salle de régie du studio fut utilisé pour la réalisation de tests expérimentaux dans le but

de valider la méthode suggérée et d'en quantifier la performance. La figure 5.30 montre la similitude entre la réponse obtenue dans cette salle à l'aide de la méthode non invasive suggérée et la méthode invasive standard. L'évolution de l'erreur quadratique moyenne au fil du temps montrée dans la figure 5.31 indique que l'algorithme de modélisation suggéré converge aussi peu que 5 à 10 secondes après que la lecture d'un extrait musical soit lancé.

Les filtres aux coefficients fixes souvent utilisés pour l'égalisation de la réponse d'une chaîne audio (amplificateur/haut-parleur/salle) souffrent tous du même problème : l'environnement dans lequel a lieu l'écoute change et le filtre est incapable de s'adapter automatiquement. Un filtre égalisateur s'adaptant périodiquement à la réponse changeante du milieu d'écoute s'appuie sur la technique de modélisation non invasive développée. La réponse du filtre est obtenue par l'inversion de la réponse de la salle en ne considérant que le champ direct et les premières réflexions, la réverbération tardive n'est pas utilisée. L'erreur quadratique moyenne est utilisée cette fois pour quantifier l'amélioration apportée par l'insertion d'un tel filtre dans une chaîne audio.

La figure 5.51 montre la réponse en fréquence avant et après correction. Le fonctionnement du filtre égalisateur adaptatif est validé lorsque 1 mètre sépare la source du microphone de mesure et ensuite à deux mètres. Le filtre ainsi que la réponse égalisée du système global restent stables même lorsque la réponse de la salle est changée de façon draconienne durant l'exercice en déplaçant le microphone d'un endroit à un autre dans la salle pendant que le filtre compense, comme le montre la figure 5.55. La réponse en fréquence de la salle après égalisation est comparée à une réponse parfaitement plate. L'évolution de l'erreur quadratique sur la réponse égalisée au fil du temps montre que l'erreur diminue à mesure que la musique défile. L'erreur quadratique moyenne sur la réponse égalisée est réduite par un facteur pouvant dépasser 100, aussi peu que 15 secondes après le lancement d'un extrait musical. La figure 5.53 montre l'évolution de l'erreur sur la réponse corrigée au fil du temps pour six extraits musicaux.

7.2 Limitations de la solution proposée

La première limitation importante de l'approche suggérée est que, sous sa forme actuelle, elle se limite à une chaîne audio à une seule entrée et une seule sortie. Elle ne permet pas ni la modélisation ni la correction dans le cas où deux haut-parleurs sont considérés. La déconvolution utilisée pour l'analyse ponctuelle de blocs de données devrait être remplacée par un système d'équation liant chacune des entrées à la sortie, pour qu'une fonction de transfert multi-entrée puisse être obtenue. Cette avenue n'a pas été explorée dans le présent travail.

La seconde limitation importante de l'approche suggérée est que la correction est ponctuelle dans l'espace. Un seul microphone est utilisé et donc la réponse corrigée n'est bonne que pour une seule position dans l'environnement d'écoute. Il suffit de déplacer le microphone de l'auditeur de quelques dizaines de centimètres pour que ce genre de correction rende la situation potentiellement pire qu'elle ne l'était initialement (sans égalisation). Les trois premiers articles répertoriés mentionnent ce problème (Radlovic *et al.*, 2000), (Elliott et Nelson, 1989) et (Mourjopoulos, 1992). Quatre autres articles répertoriés suggèrent différentes solutions à ce problème en tentant de faire une moyenne dans l'espace de réponses en fréquence mesurées à différentes positions d'écoute dans une salle dans le but d'en faire l'égalisation (Bharitkar et Kyriakakis, 2003), (Bharitkar *et al.*, 2002), (Bharitkar et Kyriakakis, 2007) et (Bharitkar et Kyriakakis, 2001). Aucune de ces méthodes ne fut explorée plus en profondeur dans le cadre de ce travail.

7.3 Améliorations futures

On a vu dans le chapitre 4 une analogie entre le problème étudié dans le cadre de ce travail et un problème classique en analyse et contrôle des systèmes dynamiques, soit le contrôle d'un moteur à courant continu. Cette analogie entre le contrôle d'un moteur et l'égalisation d'une salle nous a permis de noter que la méthode suggérée pour l'adaptation des coefficients du filtre égalisateur partage beaucoup de similitudes avec l'approche « feed-forward » décrite dans la littérature pour le contrôle d'un moteur à courant continu. Dans cette approche, l'erreur du système n'est pas réinjectée dans le contrôleur, comme dans notre cas, l'erreur sur la réponse en fréquence égalisée n'est pas réinjectée dans l'algorithme de filtrage en rétroaction. Pourtant, lorsqu'il s'agit du contrôle d'un moteur, rares sont les cas où on se satisfait d'un contrôleur « feed-forward » seul. Le contrôleur « feed-forward » est normalement utilisé en combinaison avec un contrôleur en « feedback » qui permet de pousser plus loin la correction de l'erreur statique. Dans bien des cas, avec des connaissances très minimales a priori sur la nature du système $h(t)$ dont il est question, un contrôleur « feedback » qui altérera la commande $u(t)$ en fonction de l'erreur $e(t)$ observée peut contrôler à lui seul la sortie d'un système.

Pour l'algorithme de filtrage adaptatif implémenté dans le cadre de ce travail, nous suggérons l'ajout d'une contrôleur proportionnel intégral dérivé (PID) prenant comme entrée en rétroaction l'erreur de la figure 4.30. Un tel contrôleur permettrait de réduire encore plus l'erreur quadratique moyenne sur la réponse en fréquence corrigée vers laquelle l'algorithme converge.

7.4 Applications éventuelles

L'avenir est prometteur pour cette méthode innovatrice développée permettant d'égaliser automatiquement la réponse de salles à géométrie complexe et changeante au fil du temps. Le module VST développé dans le cadre de ce travail pourrait lui-même faire partie intégrante de systèmes d'égalisation de salles sur PC pour les salles de spectacles, pour studios d'enregistrement ou pour salles de cinéma. Le choix C++ comme langage de programmation garantit également la portabilité du code vers une autre plateforme.

RÉFÉRENCES

(2000). Measurement microphone ecm8000 technical specifications v1.2, behringer.

(2011). Scarlett 2i2 user guide v1.0, focusrite.

BHARITKAR, S., HILMES, P. et KYRIAKAKIS, C. (2002). Robustness of spatial averaging equalization methods : a statistical approach. *Signals, Systems and Computers, 2002. Conference Record of the Thirty-Sixth Asilomar Conference on.* vol. 1, 184–187 vol.1.

BHARITKAR, S. et KYRIAKAKIS, C. (2001). A classification scheme for acoustical room responses. *Signal Processing and its Applications, Sixth International, Symposium on.* 2001. vol. 2, 671–674 vol.2.

BHARITKAR, S. et KYRIAKAKIS, C. (2003). A comparison between multi-channel audio equalization filters using warping. *Applications of Signal Processing to Audio and Acoustics, 2003 IEEE Workshop on.* 63–66.

BHARITKAR, S. et KYRIAKAKIS, C. (2007). Visualization of multiple listener room acoustic equalization with the sammon map. *Audio, Speech, and Language Processing, IEEE Transactions on,* 15, 542–551.

(BSI), B. S. I. (2008). Acoustics - Measurement of room acoustic parameters - Part 2 : Reverberation time in ordinary rooms.

CANCLINI, A., MARKOVIC, D., ANTONACCI, E., SARTI, A. et TUBARO, S. (2012). A room-compensated virtual surround system exploiting early reflections in a reverberant room. *Signal Processing Conference (EUSIPCO), 2012 Proceedings of the 20th European.* IEEE, 1029–1033.

COX, T. J., LI, F. et DARLINGTON, P. (2001). Extracting room reverberation time from speech using artificial neural networks. *Journal of the audio engineering society,* 49, 219–230.

ELLIOTT, S. J. et NELSON, P. A. (1989). Multiple-point equalization in a room using adaptive digital filters. *J. Audio Eng. Soc,* 37, 899–907.

FARINA, A. (2000). Simultaneous measurement of impulse response and distortion with a swept-sine technique. *Audio Engineering Society Convention 108.* Audio Engineering Society.

FERRARA, E. (1980). Fast implementations of lms adaptive filters. *Acoustics, Speech and Signal Processing, IEEE Transactions on,* 28, 474–475.

HATZIANTONIOU, P. D. et MOURJOPPOULOS, J. N. (2003). Results for room acoustics equalisation based on smoothed responses. *Audio Engineering Society Convention 114.* Audio Engineering Society.

HATZIANTONIOU, P. D. et MOURJOPOULOS, J. N. (2004). Errors in real-time room acoustics dereverberation. *Journal of the Audio Engineering Society*, 52, 883–899.

HOSSEINI, P., MARTINS, S., MARTIN, T., RADZISZEWSKI, P. et BOYER, F.-R. (2011). Acoustic emissions simulation of tumbling mills using charge dynamics. *Minerals Engineering*, 24, 1440–1447.

KENDRICK, P., COX, T. J., ZHANG, Y., CHAMBERS, J. A. et LI, F. F. (2006). Room acoustic parameter extraction from music signals. *Acoustics, Speech and Signal Processing, 2006. ICASSP 2006 Proceedings. 2006 IEEE International Conference on*. IEEE, vol. 5, V–V.

MARTINS, S., HOSSEINI, P., MARTIN, T., RADZISZEWSKI, P., BOYER, F.-R., FAUCHER, A., MAKNI, S. et SABIH, A. (2011). Simulating tumbling mill acoustic signals using DEM. *Fifth International Conference on Autogenous & Semiautogenous Grinding Technology (SAG)*. Canadian Institute of Mining, Metallurgy and Petroleum.

MIYOSHI, M. et KANEDA, Y. (1988). Inverse filtering of room acoustics. *Acoustics, Speech and Signal Processing, IEEE Transactions on*, 36, 145–152.

MOSAYYEBPOUR, S., ESMAEILI, M. et GULLIVER, T. (2013). Single-microphone early and late reverberation suppression in noisy speech. *Audio, Speech, and Language Processing, IEEE Transactions on*, 21, 322–335.

MOSAYYEBPOUR, S., SHEIKHZADEH, H., GULLIVER, T. A. et ESMAEILI, M. (2012). Single-microphone lp residual skewness-based inverse filtering of the room impulse response. *Audio, Speech, and Language Processing, IEEE Transactions on*, 20, 1617–1632.

MOURJOPOULOS, J. (1992). Digital equalization of room acoustics. *Audio Engineering Society Convention 92*.

NELSON, G., PFEIFER, L. et WOOD, R. (1972). High-speed octave band digital filtering. *Audio and Electroacoustics, IEEE Transactions on*, 20, 58–65.

NOWLIN, W. C., GUTHART, G. S. et TOTH, G. K. (2000). Noninvasive system identification for multichannel broadband active noise control. *The Journal of the Acoustical Society of America*, 107, 2049.

PLOMPO, R. (1964). The ear as a frequency analyzer. *The Journal of the Acoustical Society of America*, 36, 1628.

RADLOVIC, B., WILLIAMSON, R. et KENNEDY, R. (2000). Equalization in an acoustic reverberant environment : robustness results. *Speech and Audio Processing, IEEE Transactions on*, 8, 311–319.

RIETA, J. et SANCHIS, J. (2006). A new method to determine the frequency response of enclosures using masked tones. *Applied Acoustics*, 67, 297 – 305.

ROCHA, A. F., LEITE, A., PINTO, F. et FERREIRA, A. J. (2006). Adaptive audio equalization of rooms based on a technique of transparent insertion of acoustic probe signals. *Audio Engineering Society 120th Convention, Paris, France*. 20–23.

RUBAK, P. et JOHANSEN, L. (2001). Listening test results from a new digital loudspeaker/room correction systems.

SARRIS, J. C. et CAMBOURAKIS, G. E. (2003). Time frequency analysis and parametric approximation of room impulse responses. *Acoustics, Speech, and Signal Processing, 2003. Proceedings.(ICASSP'03). 2003 IEEE International Conference on*. IEEE, vol. 6, VI-297.

SCHROEDER, M. R. (1979). Integrated-impulse method measuring sound decay without using impulses. *The Journal of the Acoustical Society of America*, *66*, 497.

SEIB, H. (2013). Vsthost, a program to run vst-compatible plugins, v1.53.

SREENIVAS, T. V., SINGH, K. et NIEDERJOHN, R. J. (1990). Spectral resolution and noise robustness in auditory modeling. *Acoustics, Speech, and Signal Processing, 1990. ICASSP-90., 1990 International Conference on*. IEEE, 817–820.

STEWART, R. et SANDLER, M. (2007). Statistical measures of early reflections of room impulse responses. *Proc. of the 10th int. conference on digital audio effects (DAFx-07), Bordeaux, France*. 59–62.

STOCKHAM, JR., T. G. (1966). High-speed convolution and correlation. *Proceedings of the April 26-28, 1966, Spring joint computer conference*. ACM, New York, NY, USA, AFIPS '66 (Spring), 229–233.

ZHANG, W., KHONG, A. W. et NAYLOR, P. A. (2008). Adaptive inverse filtering of room acoustics. *Signals, Systems and Computers, 2008 42nd Asilomar Conference on*. IEEE, 788–792.

ANNEXE A

EQM (sans et avec filtre) accompagnée du spectrogramme des six extraits musicaux

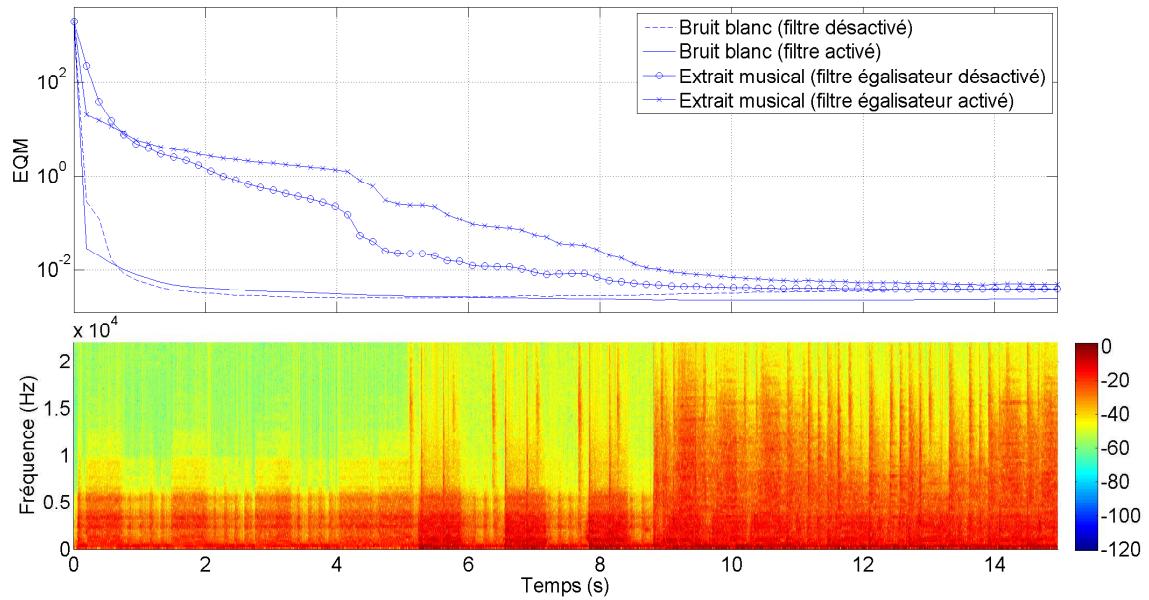


Figure A.1 EQM (sans et avec filtre) accompagnée du spectrogramme de l'extrait de la pièce « Wickeman », du groupe « Iron Maiden », sur l'album « Brave New World », paru en 2000

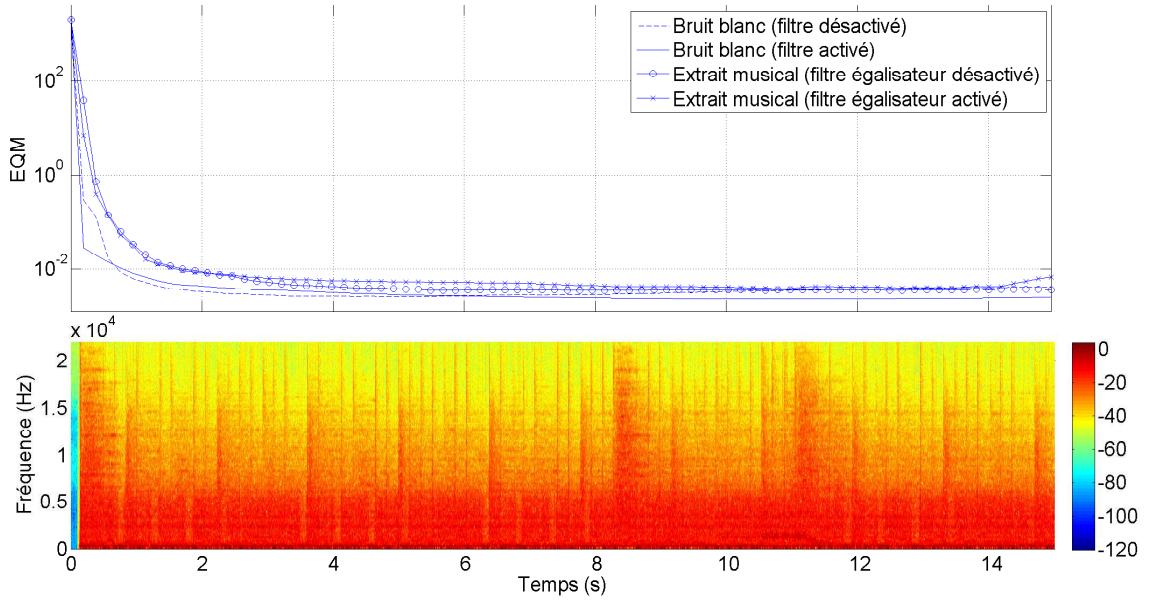


Figure A.2 EQM (sans et avec filtre) accompagnée du spectrogramme de l'extrait de la pièce « Stricken », du groupe « Disturbed », sur l'album « Ten Thousand Fists », paru en 2005

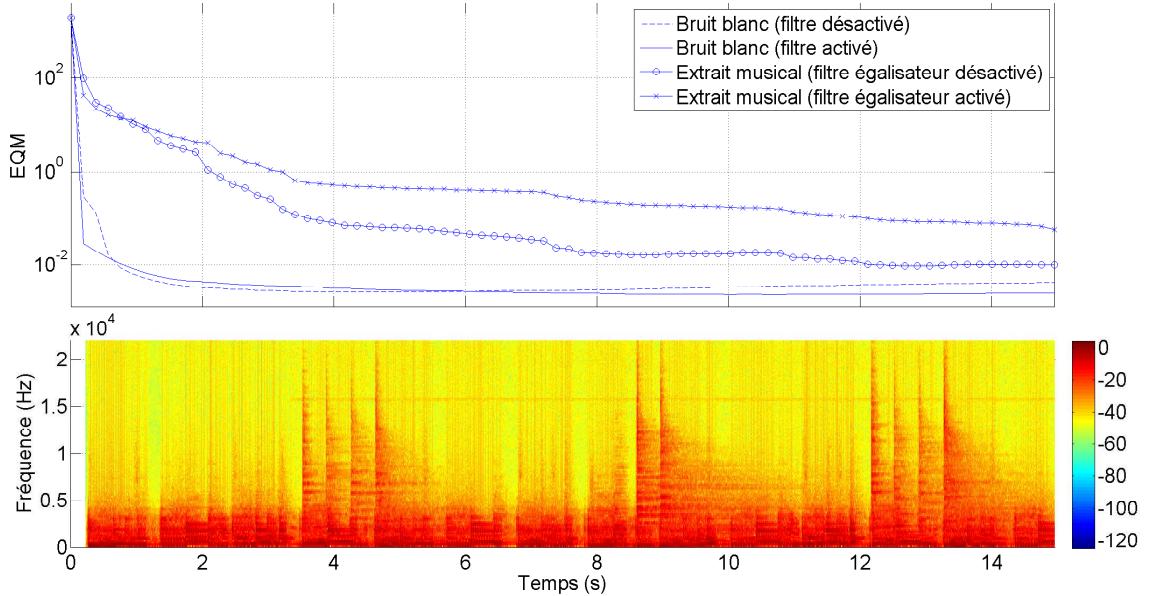


Figure A.3 EQM (sans et avec filtre) accompagnée du spectrogramme de l'extrait de la pièce « Blood Brothers », du groupe « Iron Maiden », sur l'album « Brave New World », paru en 2000

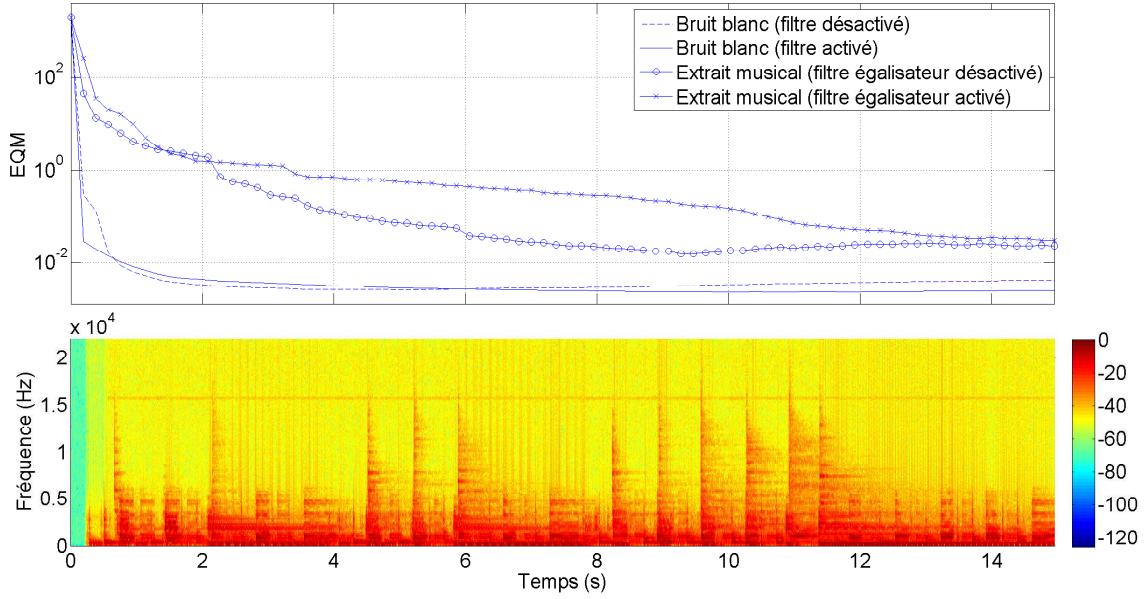


Figure A.4 EQM (sans et avec filtre) accompagnée du spectrogramme de l'extrait de la pièce « Brave new world », du groupe « Iron Maiden », sur l'album « Brave New World », paru en 2000

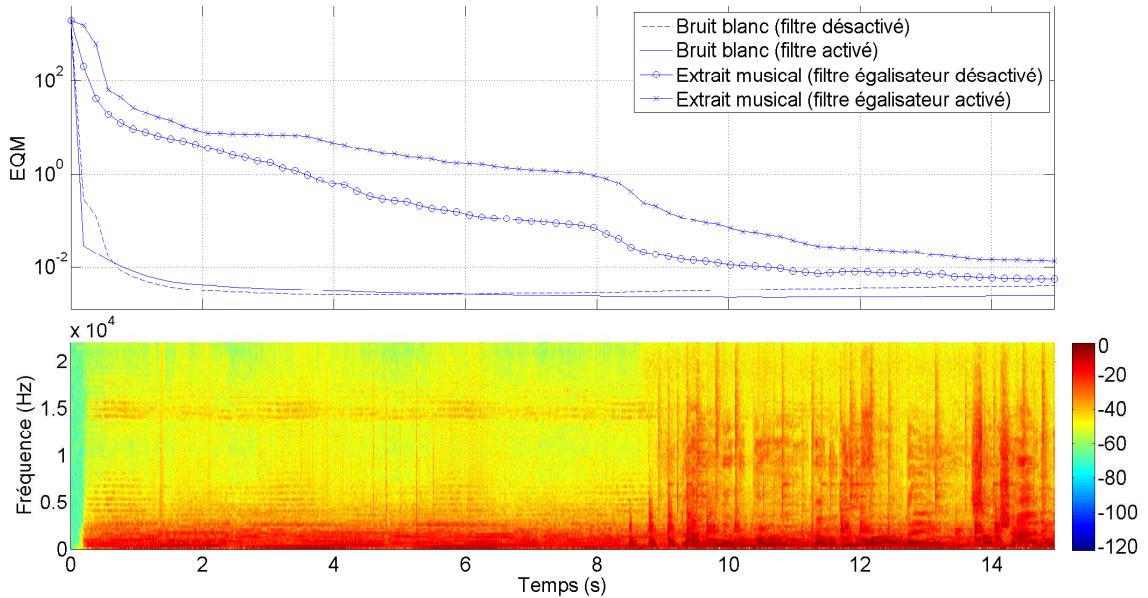


Figure A.5 EQM (sans et avec filtre) accompagnée du spectrogramme de l'extrait de la pièce « Funny the way it is », du groupe « Dave Matthews Band's », sur l'album « Big Whiskey & the GrooGrux King », paru en 2009

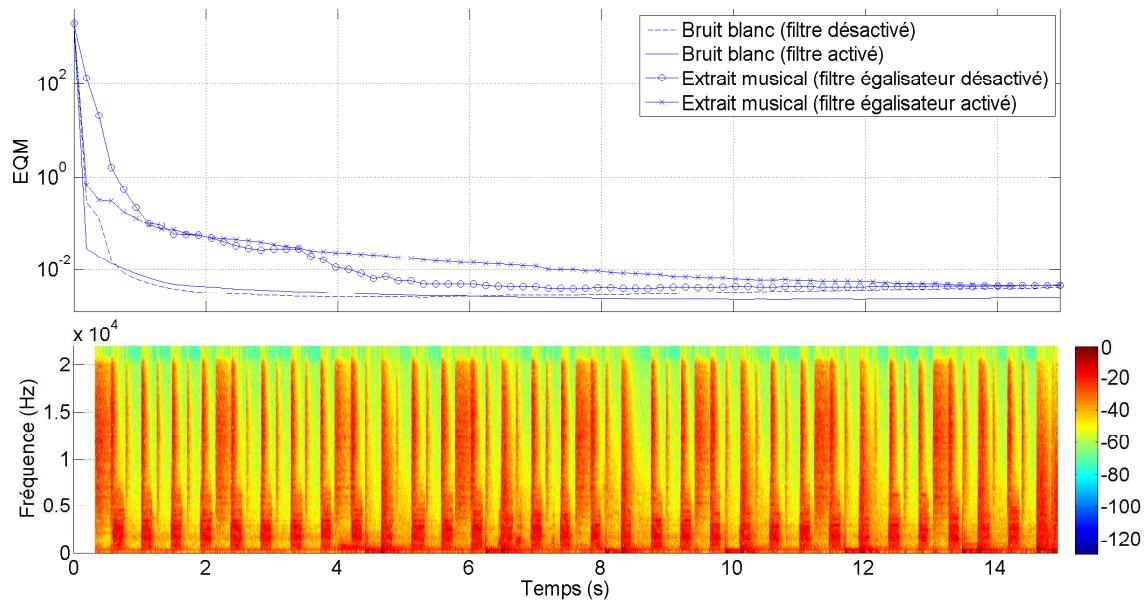


Figure A.6 EQM (sans et avec filtre) accompagnée du spectrogramme de l'extrait de la pièce « Roxanne », du groupe « The Police », sur l'album « Outlandos d'Amour », paru en 1978