

Réverbérateur hybride basé sur des réponses impulsionnelles directionnelles

Rapport de stage

Pierre Massé

Encadré par Thibaut Carpentier et Markus Noisternig

Équipe EAC, UMR 9912 STMS, IRCAM/CNRS/Sorbonne Université

M2 ATIAM, Sorbonne Université/IRCAM/Télécom ParisTech



2 février-27 juillet 2018

Table des matières

1	Introduction	5
2	Contexte	6
2.1	Bases théoriques	8
2.1.1	Spatialisation par décomposition en harmoniques sphériques	8
2.1.2	Modélisation stochastique de la réverbération	9
2.1.3	Mesures de diffusivité	12
2.1.4	Réverbération des espaces couplés	14
2.2	Autres travaux connexes	16
2.2.1	Caractérisation du temps de mélange	16
2.2.2	Analyse de la réverbération des espaces couplés	17
3	Travail réalisé	18
3.1	Caractérisation du temps de mélange	18
3.1.1	Implémentation des mesures de diffusivité	18
3.1.2	Algorithme de détermination du temps de mélange	22
3.1.3	Ajustement de l'algorithme d'hybridation	26
3.2	Synthèse de la réverbération des espaces couplés	27
3.2.1	Implémentation des méthodes d'analyse	28
3.2.2	Synthèse par double FDN	32
4	Travail ultérieur	36
4.1	Temps de mélange	36
4.1.1	Validation perceptive	36
4.1.2	Moyennage des mesures de diffusivité	37
4.1.3	Stabilité de la diffusivité par SDR	37
4.1.4	Théorie et complexité algorithmique	37
4.1.5	Diffusivité par covariance	37
4.2	FDN couplés	38
4.2.1	Détermination des paramètres de couplage	38
4.2.2	Extensions de la théorie de couplage	39
4.2.3	Hybridation multi-pente	39
4.2.4	Vérification des résultats temps-réel	39
4.2.5	Complexité algorithmique de l'analyse	40
4.2.6	Améliorations et considérations techniques	40
4.3	Perspectives générales	40
5	Conclusion	42
A	Temps de mélange	46
A.1	Cohérence spatiale	46
A.1.1	Signaux synthétiques stationnaires	46
A.1.2	Comparaison des ordres d'analyse	47
A.2	Détection de direction d'arrivée	48
A.2.1	Cartographie de l'erreur	48
A.2.2	Ajustement des valeurs autour de l'extrême azimutale	48
A.3	Bases de données de réponses impulsionnelles directionnelles mesurées	49
A.3.1	Espace de Projection (ESPRO), IRCAM	49

A.3.2	Kraftzentrale, Duisburg	50
A.4	Hybridation	51
A.4.1	Comparaison des temps de réverbération	51
A.4.2	Patch Max/MSP d'évaluation à l'écoute	52
B	FDN couplés	53
B.1	Analyse	53
B.1.1	Segmentation	53
B.2	Synthèse	53

Résumé

Un moteur de réverbération hybride [1] a été développé par l'équipe Espaces Acoustiques et Cognitifs (EAC) du STMS (Sciences et Technologies de la Musique et du Son) à l'IRCAM (Institut de Recherche et Coordination Acoustique/Musique) afin d'exploiter les avantages respectifs d'une reconstitution par convolution directe pour les réflexions précoces d'une réponse impulsionnelle directionnelle (DRIR) et d'une synthèse par réseaux de retards rebouclés (FDN) pour la queue de réverbération tardive. La transition entre les deux est faite au temps de mélange, un point critique après lequel la réverbération peut être décrite par un modèle stochastique [2] [3] [4] dont la synthèse par FDN est une implémentation valable [5]. La première partie de ce stage porte sur la caractérisation du temps de mélange à partir d'une mesure de la diffusivité du champ sonore décrit par la DRIR ; plusieurs définitions de la diffusivité [6] [7] [8] sont implémentées et évaluées, menant au développement d'un algorithme de détermination automatique du temps de mélange. Le processus d'hybridation est alors ajusté en conséquence et les DRIR hybrides générées sont évaluées à la fois qualitativement et quantitativement.

La deuxième partie du stage cherche à étendre la synthèse par FDN à la réverbération d'espaces plus complexes, et en particulier les volumes couplés. En se basant sur le modèle proposé par Cremer et Müller [9], un cadre d'analyse-synthèse est défini avec une recherche de paramètres sur le profil de décroissance énergétique (EDR), par segmentation, modélisation et inférence bayésienne, et maximisation de l'espérance. Une première implémentation de la synthèse par double FDN résultante est faite en MATLAB, parallèlement à l'analyse, puis une version temps-réel équivalente est recréée en Max/MSP. Ce cadre est globalement validé avec une synthèse paramétrée par des valeurs connues selon le modèle de Cremer et Müller. Finalement, une discussion des détails incomplets, des étapes ultérieures, et des perspectives possibles est présentée, relative à la continuation de ce travail au sein d'une thèse de doctorat à suivre.

Mots-clés : réverbération, son spatialisé, harmoniques sphériques, réponses impulsionnelles directionnelles, diffusivité, temps de mélange, réseaux de retards rebouclés, volumes couplés, décroissances à pentes multiples

Abstract

A hybrid reverberation engine [1] has been developed by the Espaces Acoustiques et Cognitifs (EAC) team of the STMS (Sciences et Technologies de la Musique et du Son) at IRCAM (Institut de Recherche et Coordination Acoustique/Musique) in order to exploit the respective advantages of reconstruction by direct convolution for the early reflexions of a Directional Room Impulse Response (DRIR) and synthesis by Feedback Delay Networks (FDN) for the late reverberation tail. The transition between the two occurs at the mixing time, a critical point after which the reverberation can be described by a stochastic model [2] [3] [4] for which FDN synthesis is a valid implementation [5]. The first part of this work relates to the characterization of the mixing time using a measure of the diffuseness of the sound field described by the DRIR; several definitions of the diffuseness [6] [7] [8] are implemented and evaluated, leading to the development of an algorithm capable of automatically determining the mixing time. The hybridization process is then consequently adjusted and the hybrid DRIRs generated are both qualitatively and quantitatively evaluated.

The second part of the work looks to extend the FDN synthesis methods to the reverberation of more complex spaces, and particularly coupled volumes. Based on the model proposed by Cremer and Müller [9], an analysis-synthesis framework is defined with a parameter search on the Energy Decay Relief (EDR), by segmentation, Bayesian model inference, and expectation maximization. The resulting double-FDN synthesis is first implemented in MATLAB, simultaneously to the analysis, before an equivalent real-time version is recreated in Max/MSP. The framework is globally verified with synthesis parametered by known values given by the Cremer and Müller model. Finally, a discussion of remaining incomplete details, possible next steps, and potential perspectives is presented relative to the continuation of this work within the context of an upcoming doctoral thesis.

Keywords : reverberation, spatialized sound, spherical harmonics, directional room impulse responses, diffuseness, mixing time, feedback delay networks, coupled volumes, multi-slope decays

1 Introduction

Ce stage s'inscrit dans le cadre du travail de l'équipe Espaces Acoustiques et Cognitifs (EAC) de l'UMR (Unité Mixte de Recherche) 9912 STMS (Sciences et Technologies de la Musique et du Son) à l'Institut de Recherche et Coordination Acoustique/Musique (IR-CAM) sur la reproduction et la manipulation de l'effet de réverbération de salle dans un contexte de son multidimensionnel ou "spatialisé". Plus particulièrement, il s'agit d'étendre les développements déjà entrepris par l'équipe autour d'un moteur de réverbération dit "hybride" [1], exploitant des techniques de synthèse par convolution et par réseaux de retards rebouclés (*Feedback Delay Networks* en anglais, ou FDN).

La réverbération d'une salle est assimilée à un processus de filtrage décrit par une réponse impulsionnelle, appelée "directionnelle" dans le cadre du son spatialisé (*Directional Room Impulse Response*, DRIR). L'idée de base du réverbérateur hybride découle d'une modélisation de cette réponse impulsionnelle en deux parties principales : une première comprenant le son direct transmis de la source au récepteur ainsi que les réflexions les plus précoces et discrètes, et une deuxième correspondant à la "queue" tardive de la réverbération, où la forte densité des réflexions permet de décrire le champ acoustique par un modèle stochastique [2] [3] [4]. La partie précoce étant difficile à modéliser et synthétiser de manière fidèle, surtout dans un contexte directionnel, on y exploite une convolution directe avec la réponse impulsionnelle. Cependant, le processus de convolution est coûteux en fonction de la longueur de la réponse impulsionnelle, et n'est pas toujours facile à paramétrer ou à manipuler ; la partie tardive est donc synthétisée par un FDN (réseau de retards rebouclés), en accord avec la modélisation stochastique [5].

La première partie du stage porte sur la caractérisation du "temps de mélange" (*mixing time*, noté t_{mix}), paramètre critique du processus d'hybridation qui définit le point de transition entre la convolution des réflexions précoces et la synthèse par FDN de la queue tardive. On cherche à prendre avantage du contexte spatialisé et définir celui-ci par rapport à la "diffusivité" du champ acoustique représenté par la DRIR (réponse impulsionnelle directionnelle), suivant la propriété fondamentale du modèle stochastique décrivant un champ parfaitement diffus (une infinité d'ondes planes décorréelées provenant d'une infinité de directions d'arrivée indépendantes). Plusieurs mesures de la diffusivité [6] [7] [8] sont donc implémentées et évaluées, et un algorithme de recherche est ensuite développé afin d'en extraire le point temporel en question.

La deuxième partie du travail est tournée vers la synthèse par FDN dans le cas de deux espaces réverbérants couplés l'un à l'autre, ce qui donne lieu à des profils de réverbération plus complexes pouvant présenter des décroissances dites en "double-pente" [9]. Un cadre d'analyse capable de prendre en compte cette complexité additionnelle [10] [11] est mis en place afin d'extraire les paramètres nécessaires pour implémenter une resynthèse avec des FDN, d'abord directement selon la théorie utilisée (Cremer-Müller [9]) et ensuite en l'adaptant à une simulation temps-réel (en Max/MSP) du transfert de la réverbération d'un espace à l'autre.

2 Contexte

Depuis plus de vingt ans maintenant [12], l'équipe EAC développe le Spatialisateur, ou Spat~, une suite d'outils de spatialisation du son dédiée à l'enregistrement, la manipulation, et la projection dans un contexte tridimensionnel. Le Spat~ se décline en plusieurs applications, notamment en tant que librairie de fonctions MATLAB (MathWorks), ainsi qu'en ensemble d'objets pour le langage de programmation visuel et temps-réel Max/MSP (Cycling '74). Parmi les fonctionnalités offertes, l'effet de réverbération de salle est implémenté selon deux techniques différentes : une convolution spatialisée avec une réponse impulsionnelle directionnelle (DRIR), et une synthèse par réseaux de retards rebouclés (FDN).

Récemment, l'idée est venue d'explorer la possibilité d'une "hybridation" entre ces deux méthodes, motivée particulièrement par une ambition d'implémentation temps-réel spatialisée [1]. La convolution permet une reconstitution idéale de l'effet de réverbération de par l'utilisation d'une réponse impulsionnelle directement mesurée dans la salle en question, mais celle-ci peut non seulement devenir rapidement coûteuse en termes de calcul numérique (surtout pour des longues réverbérations, même s'il existe aujourd'hui des algorithmes efficaces permettant son implémentation en temps-réel [13] [14]), elle est de plus difficile à paramétrer et à manipuler puisque cela nécessite d'agir directement sur la DRIR.

Historiquement, la réverbération a souvent été décrite par des modèles stochastiques [2] [3] [4] reposant sur l'idée qu'à partir d'un certain temps et au-delà d'une certaine fréquence la densité des réflexions présentes devient tellement élevée qu'une description déterministe paraît inutile. Cependant, le contenu de la réponse impulsionnelle avant ce "temps de mélange", la partie dite "précoce", est tout aussi critique à la perception de la qualité d'une salle (la limite fréquentielle, appelée "fréquence de Schroeder" d'après sa caractérisation par ce dernier, est un peu moins importante car assez basse dans la grande majorité des salles pour avoir peu d'effet perceptivement). Si des méthodes existent pour modéliser ces premières réflexions (dont certaines sont implémentées dans le Spat~), cela reste compliqué d'approcher précisément la réponse impulsionnelle d'une salle donnée, surtout dans un cadre multidimensionnel nécessitant une connaissance de la distribution spatiale. *Contrairement*, la modélisation stochastique requiert seulement le temps de décroissance en fonction de la fréquence et le profil fréquentiel de départ comme paramètres.

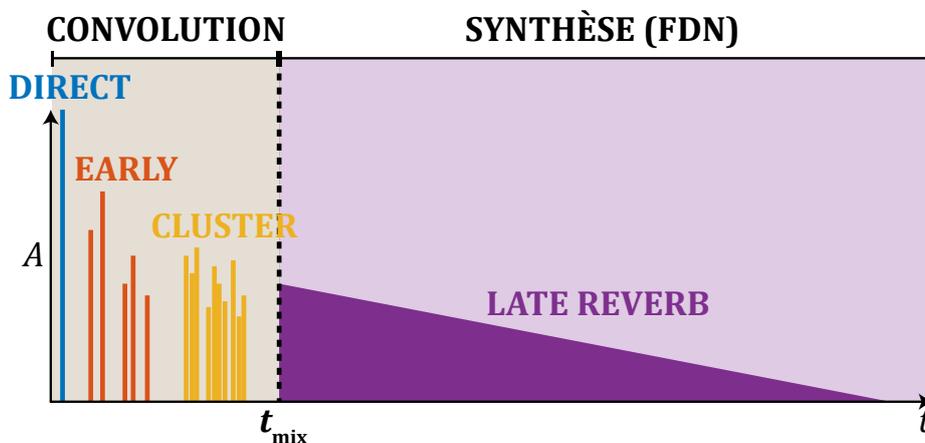


FIGURE 1 – Schématisation du modèle de réponse impulsionnelle utilisé dans le Spat~ (et en particulier pour le réverbérateur hybride)

De manière générale, le $\text{Spat}\sim$ se base sur une modélisation de la réponse impulsionnelle d'une salle en quatre parties (voir la figure 1) : le son *direct*, les toutes premières réflexions discrètes (*early*), une période de transition et d'augmentation de la densité d'échos (le *cluster*), et finalement la "queue" de réverbération tardive (*late reverb*) à décroissance exponentielle. L'idée de base derrière le moteur de réverbération hybride [1] est donc d'exploiter à la fois la reproduction idéale de la convolution sur les parties précoces (*direct/early/cluster*) et l'efficacité ainsi que la polyvalence des FDN sur la partie tardive. Le "temps de mélange" devient alors le moment critique où le réverbérateur hybride passe de la convolution au FDN, ou, en termes d'analyse, le point de transition entre les réflexions précoces et la réverbération tardive (voir à nouveau la figure 1).

Le moteur hybride est schématisé à la figure 2 et a déjà été implémenté pour des réponses impulsionnelles unidimensionnelles (omnidirectionnelles ou monophoniques) [1] [12] avec un temps de mélange paramétré "à la main" de façon empirique. Au sein de l'équipe EAC, un certain travail a aussi été effectué pour commencer à étendre ce processus à des réponses impulsionnelles directionnelles (DRIR) ou spatialisées ; le stage s'inscrit donc dans la continuité de cet élan de recherche. Plus particulièrement, il a pour premier objectif d'automatiser la paramétrisation du temps de mélange en analysant directement la DRIR. Compte tenu du contexte de son spatialisé, et surtout de la décomposition en harmoniques sphériques (à la base du format Ambisonics), ainsi que des spécificités théoriques du modèle stochastique associé à la queue de réverbération tardive, il semble judicieux de chercher à se baser sur une mesure de la "diffusivité" du champ acoustique décrit par la DRIR.

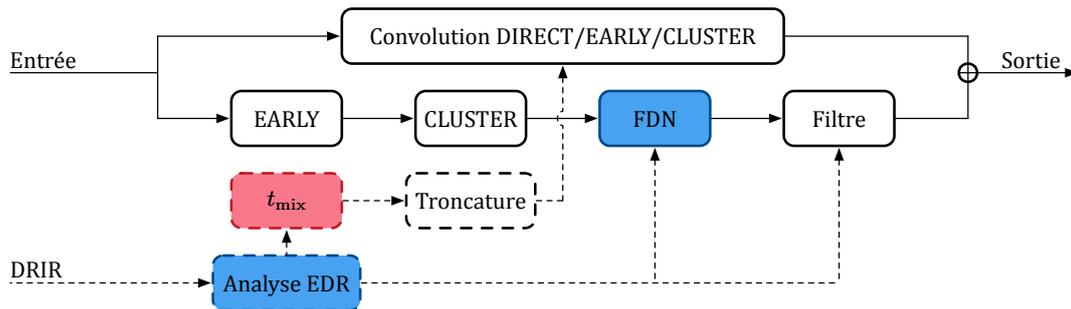


FIGURE 2 – Schématisation du réverbérateur hybride ; en rouge le module de calcul du temps de mélange sur lequel se focalise la première partie du stage, en bleu les points centraux de la deuxième partie ; les lignes hachées représentent la chaîne d'analyse conduite hors temps-réel

Le deuxième objectif est ensuite d'étendre le cadre de la synthèse par FDN et de sa paramétrisation afin de pouvoir y intégrer l'analyse et la reconstitution d'espaces couplés, telle qu'une salle de concert rattachée à une chambre réverbérante ou une petite pièce dont on ouvrirait la porte donnant sur un grand couloir ou hall. Le travail sera limité ici au cas de deux volumes reliés par une surface (éventuellement équivalente) de couplage, mais cherchera néanmoins à poser les bases nécessaires pour prendre en compte des configurations plus complexes ultérieurement.

2.1 Bases théoriques

2.1.1 Spatialisation par décomposition en harmoniques sphériques

Si le travail de Gerzon dans les années 1970 [15] représente la première formalisation d'une description tridimensionnelle du son dans le domaine des harmoniques sphériques, c'est Daniel [16] qui a ensuite étendu et généralisé cette représentation aux ordres supérieurs. Cette vue "temps-fréquence-espace" se base sur une transformée de Fourier spatiale qui peut être écrite, sur la surface d'une sphère de rayon $r = a$ [17],

$$X_{l,m}(f, t) = \int_{\Omega \in S^2} x(f, \Omega, t) Y_{l,m}(\Omega) d\Omega, \quad (1)$$

où $\Omega = (\theta, \phi)$ dans un repère sphérique (avec $\int_{\Omega \in S^2} d\Omega = \int_{-\pi}^{\pi} \int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} \sin(\theta) d\theta d\phi$) et $x(f, \Omega, t)$ est la valeur du champ acoustique mesuré à la fréquence f , au point (a, Ω) , et au temps t . $Y_{l,m}(\Omega)$ représente les harmoniques sphériques, fonctions solutions angulaires de l'équation de Helmholtz en coordonnées sphériques, d'ordre $l \in \mathbb{Z}^+$ et de degré $m = -l, -l+1, \dots, l-1, l$, formant la base orthonormée de l'espace de Hilbert.

Chaque élément $X_{l,m}(f, t)$ de la décomposition est donc la projection du champ sonore sur l'harmonique sphérique correspondante ; la transformée inverse est écrite

$$x(f, \Omega, t) = \sum_{l=0}^{\infty} \sum_{m=-l}^l X_{l,m}(f, t) Y_{l,m}^*(\Omega), \quad (2)$$

en prenant avantage de la relation d'orthonormalité :

$$\int_{\Omega \in S^2} Y_{l,m}(\Omega) Y_{l',m'}^*(\Omega) d\Omega = \delta_{l,l'} \delta_{m,m'}, \quad (3)$$

avec le symbole delta de Kronecker $\delta_{n,n'} = 1$ si $n = n'$, $\delta_{n,n'} = 0$ sinon (et $*$ dénote le conjugué complexe).

En réalité, une mesure continue $x(f, \Omega)$ est impossible à obtenir, tout comme le traitement d'une infinité de composantes $X_{l,m}(f, t)$; on ne peut qu'échantillonner ou reconstituer l'espace de façon discrète, avec un nombre limité de transducteurs. Suite notamment au travail de Rafaely [18], on privilégie aujourd'hui les antennes de microphones sphériques (*Spherical Microphone Arrays*, SMA) tel que l'Eigenmike de mh acoustics (voir la figure 3) afin de mesurer le champ acoustique. Celles-ci conduisent à remplacer l'intégrale dans l'équation 1 par une somme discrète sur l'ensemble des points $q = 1, \dots, Q$ échantillonnés :

$$\hat{X}_{l,m}(f, t) = \sum_{q=1}^Q \alpha_q(f) x(f, \Omega_q, t) Y_{l,m}(\Omega_q). \quad (4)$$

Les coefficients $\alpha_q(f)$ doivent alors être définis par rapport au SMA utilisé afin que $\hat{X}_{l,m}(f, t)$ approche $X_{l,m}(f, t)$ avec le moins d'erreur possible. De même, la somme infinie à l'équation 2 doit être tronquée à un ordre L , donnant $(L+1)^2$ termes ou composantes (et par conséquent le nombre de canaux dans une application de traitement). Pour éviter des effets de repliement (*aliasing*) spatial dans un tel système, il faut alors que $Q \geq (L+1)^2$ [18]. Par

exemple, l'Eigenmike (figure 3) contient 32 capsules, ce qui limite son encodage à l'ordre 4 (25 canaux).



FIGURE 3 – L'Eigenmike de mh acoustics

Ce sont ces types de capteurs qui permettent notamment la mesure de réponses impulsionnelles directionnelles (DRIR) en utilisant des méthodes de mesures néanmoins "classiques" (pistolet à air comprimé, balayage fréquentiel sinusoïdal, etc.) afin de caractériser la réverbération des salles de manière non seulement temporelle et fréquentielle mais aussi spatiale (d'où la terminologie "temps-fréquence-espace").

2.1.2 Modélisation stochastique de la réverbération

L'effet de réverbération des salles est étudié de manière statistique depuis les années 1950-1960 et les travaux de Schroeder [2] [3], qui furent raffinés ensuite par Polack [4] et Jot [5], entre autres. La considération de base est, qu'après un certain temps, les réflexions constituant la réverbération d'une salle deviennent suffisamment denses, à la fois temporellement et fréquentiellement, pour pouvoir être décrites de façon stochastique. Temporellement, on parle de densité d'échos, une mesure du nombre moyen de réflexions par seconde, et en fréquence d'une densité modale, le nombre moyen de modes de salle (les "fréquences naturelles" de la pièce) partageant une fréquence donnée.

La réponse impulsionnelle peut être modélisée temporellement par

$$h(t) = b(t)e^{-\delta t}, \quad (5)$$

où $b(t)$ est une réalisation d'un bruit blanc Gaussien stationnaire centré sur zéro et δ est le "temps" de décroissance, relié au temps de réverbération à -60dB (T_{60}) de Sabine par $T_{60} = 3 \ln(10)/\delta$. Afin de prendre en compte un temps de réverbération dépendant de la fréquence ainsi qu'un profil fréquentiel à l'origine arbitraire, Polack utilise la distribution temps-fréquence de Wigner-Ville [4] (ou plus particulièrement sa moyenne d'ensemble). Le modèle général résultant est donc valable au-delà d'une certaine fréquence, aujourd'hui communément appelée la "fréquence de Schroeder" ($f_{\text{Sch}} \simeq 2000\sqrt{T_{60}/V}$, avec V le volume de la salle), et après un certain temps, le "temps de mélange" (pour lequel Polack [4] donne

une borne supérieure de $t_{\text{mix}} \propto \sqrt{V}$), où les deux types de densités sont assez élevées (voir la figure 4).

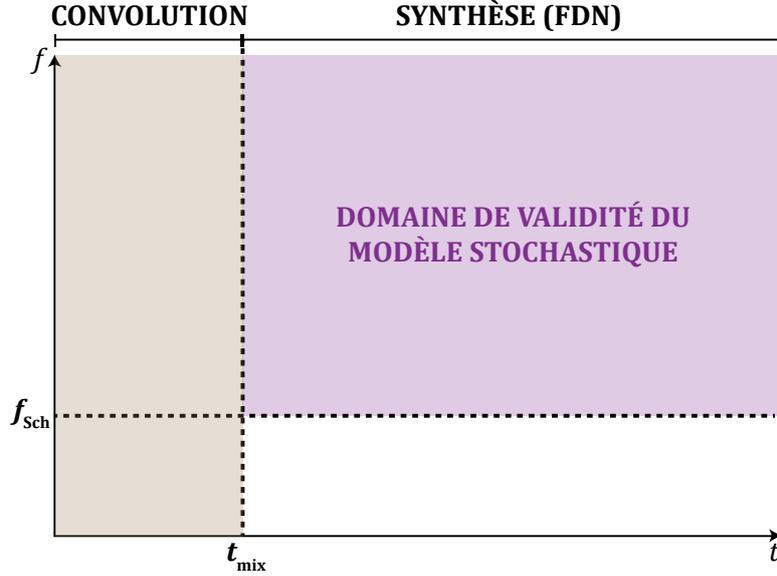


FIGURE 4 – Domaine de validité du modèle stochastique utilisé dans le réverbérateur hybride

Analyse par profil de décroissance énergétique L’analyse énergétique de la réverbération provient aussi de Schroeder [19] et de sa définition de la courbe de décroissance énergétique (communément appelée *Energy Decay Curve*, EDC). Essentiellement, la méthode consiste en une intégration rétrograde (*backward integration*) du carré de la réponse impulsionnelle; il démontre alors que celle-ci est équivalente à la moyenne d’ensemble du signal mesuré, simplement calculée à partir d’une seule réalisation¹ :

$$\text{EDC}_h(t) = \langle s^2(t) \rangle = \int_t^\infty h^2(\tau) d\tau. \quad (6)$$

L’analyse par EDC permet la détermination précise du temps de réverbération global de la salle (voir la figure 5a), mais puisque l’intégration ne comporte aucune information fréquentielle, on y est limité à des aspects purement temporels. Jot [5] a étendu cette formalisation en définissant une représentation temps-fréquence, le profil de décroissance énergétique (*Energy Decay Relief*, EDR; voir la figure 5b), qui peut être interprétée comme le calcul d’une EDC à chaque fréquence :

$$\text{EDR}_h(f, t) = \int_t^\infty H(f, \tau) d\tau. \quad (7)$$

Idéalement, le choix théorique de la distribution temps-fréquence sous-jacente à $H(f, t)$ devrait satisfaire certaines propriétés statistiques (menant Polack [4] à utiliser la distribution de Wigner-Ville, comme mentionné précédemment, et Jot [5] la distribution de Page ou Levin); en pratique, l’analyse numérique se fait avec la transformée de Fourier à court

1. À noter que cette formulation simplifiée néglige les effets du matériel de mesure sur la réponse impulsionnelle $h(t)$, ainsi que l’étape de normalisation, qui peut varier selon le choix de référence.

terme (*Short Term Fourier Transform*, STFT) discrète, qui peut être paramétrée de façon à approcher les propriétés désirables [5].

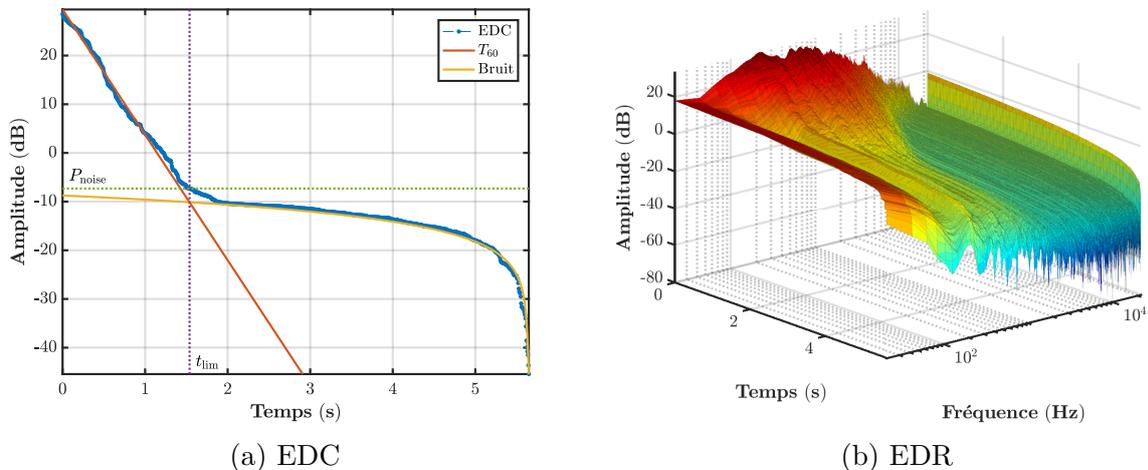


FIGURE 5

La sensibilité à la présence de bruit de mesure de l'EDC fût remarquée dès sa conception (et notée par Schroeder lui-même). En effet, puisqu'elle est définie par une intégration sur la réponse impulsionnelle à partir du point tempore considéré, sa valeur à chaque instant contient la somme de l'énergie due au bruit de fond sur cette même période. Jot [5] propose ainsi un algorithme itératif d'analyse et de débruitage combiné, remplaçant la partie de l'EDC au-delà du "temps limite" (le moment où la courbe de décroissance exponentielle idéale, ou linéaire en déciBel, rencontre le profil idéal du bruit de mesure; voir fig.) par une extension de la décroissance exponentielle idéale paramétrée par l'estimation du temps de réverbération (T_{60}). Une analyse complète par EDR donne donc les $T_{60}(f)$ et $t_{\text{lim}}(f)$ en fonction de la fréquence, ainsi que le profil énergétique au temps zéro ($P_0(f)$) et le niveau où l'EDC tombe dans le bruit de mesure ($P_{\text{noise}}(f)$).

Synthèse par réseaux de retards rebouclés L'utilisation de retards rebouclés pour simuler la réverbération vient encore une fois du travail de Schroeder dans les années 1960 [20], avant d'être étendu aux applications numériques dans les années 1990 encore une fois par Jot [21] (entre autres). Tel qu'implémenté dans le *Spat~*, il s'agit d'un réseau multi-canal mélangeant et produisant un nombre arbitraire² de canaux de réverbération décorrélés (voir la figure 6). Ceux-ci peuvent ensuite être sommés afin de créer un seul canal monophonique, ou sinon leur décorrélation peut être exploitée afin de créer un champ de réverbération multi-canal parfaitement diffus (en accord avec le modèle statistique présenté ci-dessus, car un champ diffus est défini comme étant composé d'une infinité d'ondes planes provenant d'une infinité de directions aléatoires et différentes et de phases aléatoires et décorrélées).

En règle générale, plus le nombre de canaux est élevé, plus le réseau convergera rapidement vers une réverbération assez dense au sens du modèle stochastique, mais plus la synthèse demandera alors de ressources de calcul. L'équipe EAC a développé des modules de synthèse des premières réflexions (*early/cluster*, voir la figure 1) afin de remédier non

2. En réalité les nombres de canaux puissances de 2 sont privilégiés pour des raisons pratiques au niveau des matrices de mélange (utilisation de matrices de Hadamard).

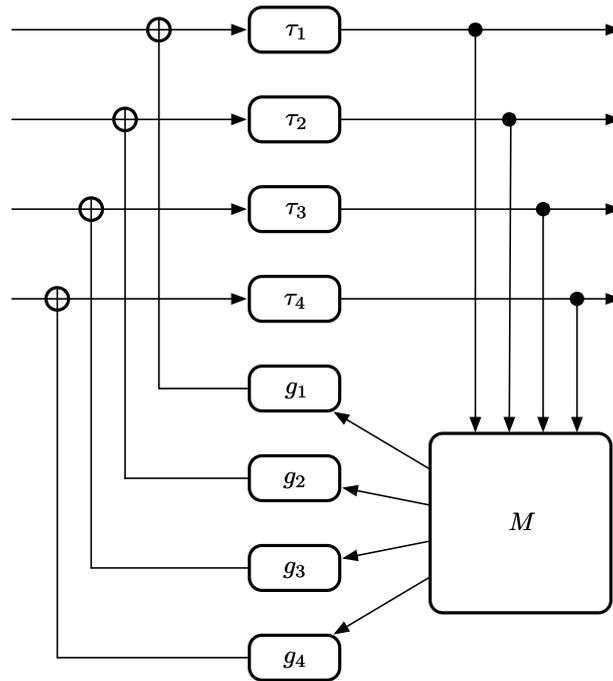


FIGURE 6 – Schéma simplifié d'un réseau de retards rebouclés multi-canal ; M est une matrice mélangeante, et les retards τ_n et gains $g_n < 1$ sont calculés afin de maximiser la décorrélation en sortie tout en respectant le T_{60} visé

seulement au fait qu'avant le temps de mélange la réverbération ne se décrit pas avec des méthodes statistiques, mais aussi au temps d'attente nécessaire avant que le FDN produise une densité d'échos suffisamment élevée (la densité augmente plus rapidement avec un groupe d'impulsions en entrée du réseau qu'avec une seule).

2.1.3 Mesures de diffusivité

Le premier objectif de ce stage est de trouver une manière de caractériser le temps de mélange de la réverbération, le point de transition entre les réflexions précoces (partie convolutive du réverbérateur hybride) et la queue tardive (synthétisée par FDN ; voir la figure 1). Comme noté ci-dessus, la description acoustique d'un champ diffus dans un repère sphérique, tel que celui considéré dans le contexte d'une spatialisation en harmoniques sphériques (voir la section 2.1.1), est celle d'une infinité d'ondes planes provenant d'une infinité de directions d'arrivée différentes, aléatoires, et décorrélées, et ayant des phases pareillement différentes, aléatoires, et décorrélées. Compte tenu du lien entre cette définition, la synthèse par FDN multi-canal, et le modèle stochastique sous-jacent, il semble judicieux d'essayer de se baser sur une mesure du degré de "diffusivité" de la réponse impulsionnelle directionnelle (DRIR), un descripteur contenu entre 0 (diffusivité minimale ou directivité maximale, c'est-à-dire la présence d'une seule onde plane) et 1 (diffusivité maximale ou directivité minimale, c'est-à-dire un champ entièrement diffus). Un aperçu des autres options explorées dans la littérature est présenté à la section 2.2.

Vecteur d'intensité Au cours de leurs travaux sur la synthèse de réponses impulsionnelles spatialisées, Merimaa et Pulkki [7] ont défini la diffusivité à partir d'un rapport entre l'énergie "directive" du champ sonore à son énergie totale. Ce rapport est décrit dans le

cadre du "B-format" suite au travail de Gerzon [15]; après Daniel [16] et l'établissement du format Ambisonics, le B-format peut être vu comme un système d'ordre 1 avec une nomenclature et une normalisation particulière. Les quatre composantes, notées $X_{0,0}$, $X_{1,-1}$, $X_{1,0}$, et $X_{1,1}$, en harmoniques sphériques (voir la section 2.1.1), y sont appelées respectivement W ou "omni", Y , Z , et X , en raison de leur orientation dans un repère cartésien³.

Le vecteur d'intensité d'un champ acoustique est la projection de la pression sur le vecteur de débit; ici, Merimaa et Pulkki assimilent la composante omnidirectionnelle $X_{0,0}$ (ou W) à la pression acoustique, et construisent un vecteur de débit à partir des trois composantes directionnelles: $\mathbf{U} = X_{1,1}\mathbf{e}_x + X_{1,-1}\mathbf{e}_y + X_{1,0}\mathbf{e}_z$, où \mathbf{e}_x , \mathbf{e}_y , et \mathbf{e}_z sont les vecteurs unitaires cartésiens. Ils peuvent alors écrire l'intensité "active" dans une fenêtre de STFT (transformée de Fourier à court terme) $\mathbf{I}_a(f) = \sqrt{2}\text{Re}\{X_{0,0}^*(f)\mathbf{U}(f)\}/Z_0$, où $\text{Re}\{\cdot\}$ fait référence à la partie réelle d'une valeur complexe et $Z_0 = \rho_0 c$ est l'impédance acoustique d'un champ libre (avec c la célérité du son dans l'air et ρ_0 la masse volumique de l'air), et l'énergie totale réelle de la fenêtre $E(f) = \rho_0(|X_{0,0}(f)|^2/Z_0^2 + |\mathbf{U}(f)|^2)/2$, où $|\cdot|^2$ représente le module carré (carré de la valeur absolue d'un scalaire, produit scalaire de la valeur absolue d'un vecteur avec lui-même). La diffusivité peut alors être finalement calculée à partir d'une représentation temps-fréquence (e.g. STFT) des composantes :

$$\Psi_{\text{VI}}(f, t) = 1 - \frac{\sqrt{2}||\text{Re}\{X_{0,0}^*(f, t)\mathbf{U}(f, t)\}||}{2(|X_{0,0}(f, t)|^2 + |\mathbf{U}(f, t)|^2/2)}, \quad (8)$$

où $||\cdot||$ dénote la norme l_2 . À noter que le vecteur d'intensité peut aussi être utilisé afin d'estimer la direction d'arrivée prédominante dans la fenêtre d'analyse, $\Omega_{\text{dir}} = (\theta_{\text{dir}}, \phi_{\text{dir}})$. Si l'on décompose le vecteur d'intensité de manière cartésienne, alors les angles sphériques sont donnés par :

$$\theta_{\text{dir}}(f, t) = \arctan \left[\frac{\text{Re}\{X_{0,0}^*(f, t)X_{1,1}(f, t)\}}{\text{Re}\{X_{0,0}^*(f, t)X_{1,-1}(f, t)\}} \right] \quad (9)$$

et

$$\phi_{\text{dir}}(f, t) = \arctan \left[\frac{\text{Re}\{X_{0,0}^*(f, t)X_{1,0}(f, t)\}}{\sqrt{\text{Re}\{X_{0,0}^*(f, t)X_{1,1}(f, t)\}^2 + \text{Re}\{X_{0,0}^*(f, t)X_{1,-1}(f, t)\}^2}} \right]. \quad (10)$$

Rapport signal-sur-diffus Jarrett *et al.* [6] ont eux cherché à exploiter la définition d'un champ diffus telle que donnée ci-dessus, et plus particulièrement par rapport à la "cohérence spatiale" des ondes planes. Puisque dans ce contexte un champ diffus peut être décrit par une infinité d'ondes planes parfaitement décorréelées, la cohérence entre les différentes composantes en harmoniques sphériques tend vers zéro. Ils démontrent alors que cette cohérence $\gamma_{l_1, m_1; l_2, m_2}(f, t)$ entre deux composantes (l_1, m_1) et (l_2, m_2) peut être utilisée afin d'estimer un rapport "signal-sur-diffus" (*Signal-to-Diffuse Ratio*, SDR) $\hat{\Gamma}_{l_1, m_1; l_2, m_2}(f, t)$, d'abord entre les deux composantes et ensuite globalement ($\hat{\Gamma}(f, t)$) en moyennant sur l'ensemble des combinaisons possible (en excluant les égalités et les symétries) avec une pondération bien choisie. Le critère de diffusivité s'écrit alors "simplement" :

3. Pour l'uniformité et la cohérence du travail présenté à la suite, la première notation sera néanmoins utilisée.

$$\Psi_{\text{SDR}}(f, t) = \frac{1}{1 + \hat{\Gamma}(f, t)}. \quad (11)$$

Il est à noter ici que la fonction de pondération donnée par Jarrett *et al.* nécessite une estimation de la direction d'arrivée (voir ci-dessus). De plus, et contrairement à celle proposée par Merimaa et Pulkki, cette définition de la diffusivité ne comporte aucune restriction par rapport à l'ordre d'analyse.

Covariance Plus récemment, dans le cadre d'un travail sur la déréverbération, Epain et Jin [22] [8] se sont penchés sur la caractérisation de la diffusivité en partant de l'évolution de la matrice de covariance des composantes en harmoniques sphériques du signal enregistré. Si l'on rassemble ces composantes dans un vecteur $\mathbf{X}(n)$, à une fréquence donnée sur les trames de la STFT (n dénotant l'indice discret de la trame temporelle), alors la matrice de covariance peut être définie par :

$$\mathbb{C}(n) = (1 - \beta)\mathbb{C}(n - 1) + \beta\mathbf{X}(n)\mathbf{X}^H(n), \quad (12)$$

où β est le facteur "d'oubli" et \cdot^H dénote la matrice adjointe ou conjuguée hermitienne (conjuguée et transposée). La diffusivité est ensuite définie par rapport à la distance de chaque valeur propre $\lambda_m(n)$ de la matrice de covariance à leur moyenne $\bar{\lambda}(n)$:

$$\Lambda(n) = \frac{1}{\bar{\lambda}(n)} \sum_{m=1}^M |\lambda_m(n) - \bar{\lambda}(n)| \Rightarrow \Psi_{\text{CV}}(n) = 1 - \frac{\Lambda(n)}{\Lambda_0} \quad (13)$$

où $M = (L + 1)^2$ est le nombre de composantes à l'ordre L et $\Lambda_0 = 2(M - 1)$ est la valeur maximale de Λ dans le cas le plus directif (une seule onde plane présente).

2.1.4 Réverbération des espaces couplés

Le deuxième objectif de ce stage est d'étendre les capacités de la synthèse par FDN afin de pouvoir simuler des espaces plus complexes tels que des volumes couplés. Si l'existence de décroissances non-exponentielles avait déjà été notée par Schroeder [20], c'est bien le travail de Cremer et Müller [9] qui fit le lien entre les propriétés physiques du couplage et l'apparence d'une réverbération dite à "double-pente". Ils décrivent l'équilibre dynamique de l'énergie totale des deux espaces couplés (respectivement $E_1(t)$ et $E_2(t)$) par le système d'équations

$$\begin{cases} \frac{c}{4}(A_{1,1}E_1(t) - \zeta S_{1,2}E_2(t)) = -V_1 \frac{dE_1(t)}{dt} \\ \frac{c}{4}(-\zeta S_{1,2}E_1(t) + A_{2,2}E_2(t)) = -V_2 \frac{dE_2(t)}{dt} \end{cases}, \quad (14)$$

où V_1 et V_2 sont les volumes respectifs, $S_{1,2}$ est la surface de couplage entre les deux (voir la figure 7), ζ est le coefficient de transmission sur la surface de couplage, et $A_{1,1} = A_{1,0} + a_{1,S}S_{1,2}$ et $A_{2,2} = A_{2,0} + a_{2,S}S_{1,2}$ sont les surfaces équivalentes d'absorption totales, incluant la surface de couplage et les surfaces équivalentes d'absorption séparées $A_{1,0}$ et

$A_{2,0}$ (sans l'effet de la surface de couplage, i.e. $A_{i,0} = \sum_j a_{i,j}A_{i,j}$ où $a_{i,j}$ est le coefficient d'absorption de la surface $A_{i,j}$ de la pièce i). Les processus de réverbération solutions du système sont alors supposés de la forme $E_i(t) = E_{0,i}e^{-2\delta t}$ (le facteur additionnel de 2 dans l'exponentielle, par rapport à l'équation 5, traduit la relation carrée entre la pression et l'énergie).

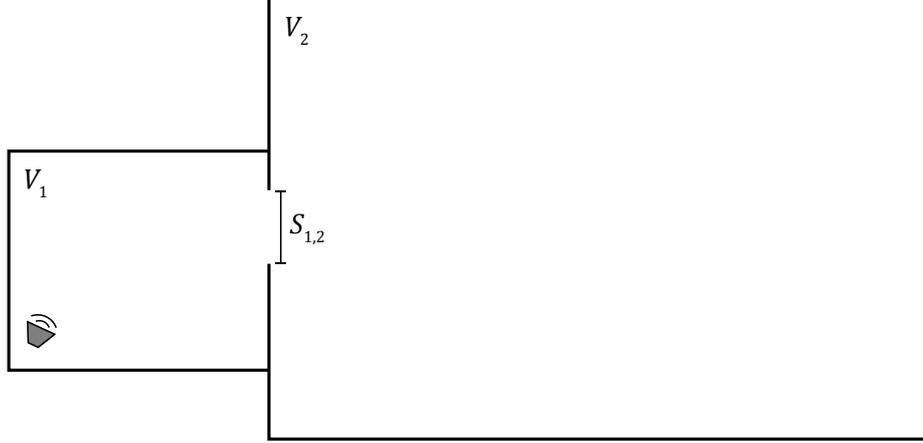


FIGURE 7 – Système de volumes V_1 et V_2 couplés par une surface $S_{1,2}$

La description des deux salles peut être simplifiée en définissant $\delta_1 = cA_{1,1}/8V_1$ et $\delta_2 = cA_{2,2}/8V_2$, les facteurs de décroissance des deux salles séparées (non couplées mais incluant la surface de couplage), ainsi que les "facteurs de couplage" $k_1 = \zeta S_{1,2}/A_{1,1}$ et $k_2 = \zeta S_{1,2}/A_{2,2}$ et leur moyenne géométrique $\kappa = \sqrt{k_1 k_2}$. Les valeurs propres δ_I et δ_{II} qui résolvent le système 14 peuvent alors être exprimées :

$$\delta_{I,II} = \frac{1}{2}(\delta_1 + \delta_2) \pm \sqrt{\frac{1}{4}(\delta_1 - \delta_2)^2 + \kappa^2 \delta_1 \delta_2}, \quad (15)$$

et les processus de réverbération par

$$\begin{cases} E_1(t) = E_{I,1}e^{-2\delta_I t} + E_{II,1}e^{-2\delta_{II} t} \\ E_2(t) = E_{I,2}e^{-2\delta_I t} + E_{II,2}e^{-2\delta_{II} t} \end{cases} \quad (16)$$

Deux des quatre coefficients peuvent ensuite être éliminés en revenant au système 14, notamment $E_{I,2}$ et $E_{II,1}$ qui sont les termes dûs au couplage :

$$\begin{cases} E_{I,2} = E_{I,1} \frac{k_2}{1 - \delta_I/\delta_2} \\ E_{II,1} = E_{II,2} \frac{k_1}{1 - \delta_{II}/\delta_1} \end{cases}, \quad (17)$$

et finalement les deux coefficients qui restent sont exprimés en termes des valeurs initiales $E_{0,1} = E_1(t=0)$ et $E_{0,2} = E_2(t=0)$:

$$\begin{cases} E_{I,1} = \frac{E_{0,1} - E_{0,2}k_1/(1 - \delta_{II}/\delta_1)}{1 - \kappa^2/(1 - \delta_I/\delta_2)(1 - \delta_{II}/\delta_1)} \\ E_{II,2} = \frac{E_{0,2} - E_{0,1}k_2/(1 - \delta_I/\delta_2)}{1 - \kappa^2/(1 - \delta_I/\delta_2)(1 - \delta_{II}/\delta_1)} \end{cases} \quad (18)$$

La caractérisation de la réverbération des espaces couplés par une somme de processus exponentiels est aussi à la base du travail d'analyse par recherche bayésienne proposé par Xiang *et al.* [23] [10]. Afin de pouvoir exploiter les propriétés de l'inférence bayésienne, ils modélisent l'EDC (courbe intégrée de Schroeder, voir la section 2.1.2) d'une réverbération couplée par une somme généralisée :

$$\widehat{\text{EDC}}(\mathbf{E}, \mathbf{D}, t_n) = \sum_{k=1}^{K+1} E_k G_k(\delta_k, t_n), \quad (19)$$

où

$$G_k(\delta_k, t_n) = \begin{cases} e^{-\delta_k t_n}, & k \leq K \\ t_N - t_n, & k = K + 1 \end{cases}, \quad (20)$$

\mathbf{E} est le vecteur des amplitudes E_k , \mathbf{D} le vecteur des facteurs de décroissance δ_k , K est le nombre de pentes à modéliser (e.g. 2 pour un "simple" couplage entre deux volumes tel que décrit par Cremer et Müller), et N est le nombre de trames temporelles de l'EDC. La probabilité bayésienne *a posteriori* étant donné les valeurs des paramètres \mathbf{E} et \mathbf{D} peut être réduite à l'inverse d'une fonction de coût euclidienne (Xiang *et al.* travaillent avec les logarithmes népériens par souci de stabilité numérique) :

$$\mathcal{L}(\mathbf{E}, \mathbf{D}) = \left\{ \frac{1}{N} \sqrt{\sum_{n=1}^N [\text{EDC}(t_n) - \widehat{\text{EDC}}(\mathbf{E}, \mathbf{D}, t_n)]^2} \right\}^{-1} \quad (21)$$

Le maximum global de cette fonction d'inférence (ou postérieur) correspond alors aux valeurs des paramètres qui modélisent le mieux l'EDC.

2.2 Autres travaux connexes

Malgré leur contexte et leur perspective de recherche particulière, les objectifs de ce stage recouvrent certains aspects déjà abordés dans la littérature du domaine. Un petit tour d'horizon est effectué ici afin de mieux justifier les approches retenues et développées par la suite.

2.2.1 Caractérisation du temps de mélange

Outre la borne supérieure donnée par Polack (voir la section 2.1.2), d'autres mesures ont été utilisées pour tenter de décrire et calculer le temps de mélange. Abel et Huang [24], notamment, utilisent un "profil de densité d'échos" (*Echo Density Profile*, EDP), tandis

que plusieurs autres définitions se basent sur un calcul du kurtosis de la réponse impulsionnelle (une version normalisée du moment statistique d'ordre 4 de la distribution des échantillons) ; ces deux méthodes sont d'ailleurs implémentées au sein du `Spat~`. Cependant, ces mesures restent purement statistiques et n'exploitent pas les propriétés spatiales du champ acoustique ; de plus, elles sont définies seulement pour des réponses impulsionnelles monophoniques, ce qui engendre des questions d'implémentation peu triviales dans le cadre spatialisé.

2.2.2 Analyse de la réverbération des espaces couplés

Luizard et Katz [11] présentent une méthode d'analyse de l'EDC plus alignée avec les algorithmes déjà mis en place par l'équipe EAC, c'est à dire par segmentation et régression linéaire locale. Si celle-ci est plus simple à implémenter et plus rapide au calcul, le fait qu'elle ne soit pas basée sur une modélisation de la réverbération la rend plus délicate à relier à la théorie de Cremer et Müller (voir ci-dessus).

3 Travail réalisé

Cette section présente le travail réalisé envers les principaux objectifs du stage : dans un premier temps la détermination du temps de mélange d'une réponse impulsionnelle directionnelle (DRIR) à partir d'un critère de diffusivité, et ensuite le développement d'un cadre d'analyse-synthèse par réseaux de retards rebouclés (FDN) capable de prendre un compte des réverbérations plus complexes, telles que celles des volumes couplés.

3.1 Caractérisation du temps de mélange

Afin d'arriver à un algorithme d'hybridation où le moment de transition entre la convolution et la synthèse par FDN (voir les figures 1 et 2) est déterminé automatiquement à partir d'une mesure de la diffusivité de la DRIR, il est nécessaire de développer une analyse robuste de ce critère. La première étape consiste donc à implémenter les différentes mesures de diffusivité présentées à la section 2.1.3 et s'assurer de leur validité une fois appliquées à des DRIR réelles. Vient ensuite le développement de l'algorithme chargé d'en extraire le point temporel représentant au mieux le temps de mélange, et finalement le processus d'hybridation déjà mis en place par l'équipe EAC peut être ajusté pour prendre en compte cette détermination automatique.

3.1.1 Implémentation des mesures de diffusivité

Un algorithme de traitement est implémenté en MATLAB pour chacune des trois mesures décrites à la section 2.1.3. Celles-ci sont ensuite évaluées en trois étapes principales :

1. Calcul sur des signaux numériques synthétisés en MATLAB, encodés aux ordres 1 et 4 dans le domaine des harmoniques sphériques (Ambisonics),
2. Application à des enregistrements par Eigenmike de ces signaux d'ordre 4, décodés à l'ordre 3 et envoyés sur les 24 haut-parleurs du Studio 4 à l'IRCAM, puis ré-encodés à l'ordre 4,
3. Application à des mesures de DRIR réelles, enregistrées par Eigenmike (base de données de l'équipe EAC) et encodées à l'ordre 4.

Signaux de synthèse Les signaux sont synthétisés avec des bruits blancs (distributions aléatoires d'échantillons) pour éviter, dans un premier temps, toute influence d'une éventuelle dépendance fréquentielle sur l'évaluation des résultats ; de plus, le bruit blanc approche la description stochastique de la réverbération (voir la section 2.1.2). Trois types de signaux sont alors synthétisés, simulant grossièrement les différentes parties d'une réponse impulsionnelle idéale (voir la figure 1) :

1. Un son purement directif avec une seule direction d'arrivée bien précise et statique (le son direct),
2. Différentes répartitions de 4 à 25 sources (selon l'ordre du signal synthétisé) réparties sur autant de directions d'arrivée uniques (les réflexions précoces, plus ou moins denses),

3. La distribution d'un grand nombre de sources (par rapport à l'ordre d'encodage) sur toute la sphère des directions d'arrivée, selon une répartition cherchant à maximiser la décorrélation (le champ diffus).

Les trois types de signaux sont enregistrés numériquement d'abord séparément puis en les enchainant avec des transitions en fondu croisé (*crossfade*) et une enveloppe d'amplitude globale décroissante (mais non exponentielle); ce sont ces enchainements évolutifs qui sont aussi enregistrés en Studio 4. La synthèse est implémentée en MATLAB avec la librairie de fonctions de l'équipe EAC, et ensuite recréée en Max/MSP avec la suite d'objets du Spat~ (équivalents aux fonctions MATLAB) afin de faciliter l'enregistrement par Eigenmike des signaux envoyés sur le système de haut-parleurs du Studio 4 de manière simultanée.

Vecteur d'intensité La figure 8 montre la diffusivité calculée sur les signaux purement numériques (gauche, figure 8a) et ceux enregistrés avec l'Eigenmike en Studio 4 (droite, figure 8b), avec ici une direction d'arrivée du son direct $\Omega_{\text{dir}} = (\theta_{\text{dir}}, \phi_{\text{dir}}) = (45^\circ, 45^\circ)$. La définition temps-fréquence de Merimaa et Pulkki [7] (voir la section 2.1.3) est réduite à la seule dimension temporelle en moyennant sur les fréquences (voir la note sur la plage fréquentielle utile de l'Eigenmike ci-dessous).

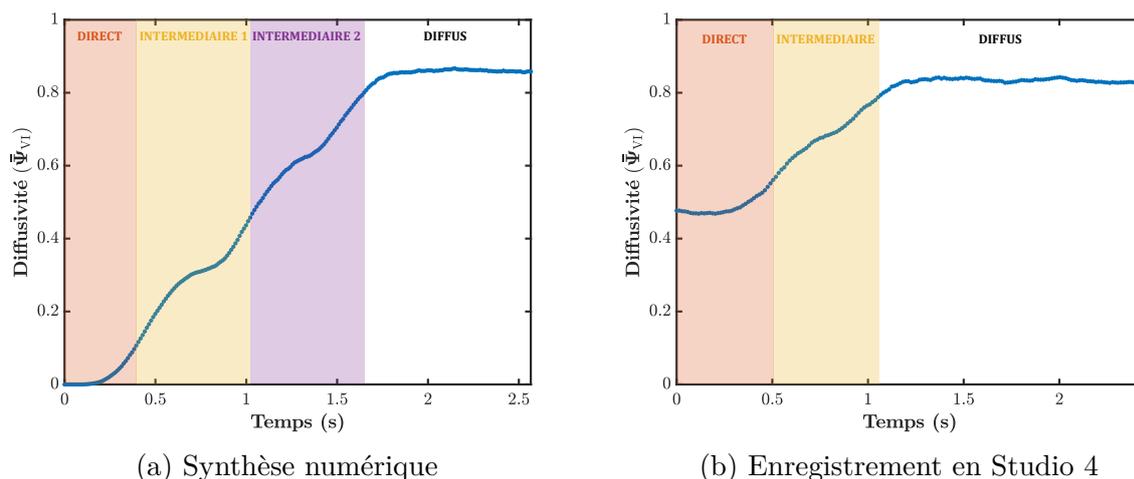


FIGURE 8 – Diffusivité par vecteur d'intensité ($\bar{\Psi}_{VI}$) calculée sur des signaux de synthèse encodés à l'ordre 4 ($\Omega_{\text{dir}} = (45^\circ, 45^\circ)$), moyennée sur les fréquences

L'analyse donne une diffusivité maximale aux alentours de 0,85 dans les deux cas, la différence avec la valeur idéale de 1 pour un champ parfaitement diffus venant sûrement de son approximation par un nombre fini d'ondes planes. La valeur minimale vers 0,55 pour l'enregistrement en Studio 4 est peut être attribuée soit au fait que le champ diffus était synthétisé et encodé sur une sphère entière alors que la cage de haut-parleur du Studio 4 ne forme qu'une demie-sphère, soit à la forte présence de réflexions du sol non-absorbant du studio (ou bien à une combinaison des deux). Par manque de temps d'accès au studio, ces phénomènes n'ont malheureusement pas (encore) pu être corrigés par de nouvelles prises de mesures.

Cependant, on retrouve bien le même profil de courbe dans les deux cas, correspondant aux signaux synthétisés décrits ci-dessus; il est à noter que les versions envoyées dans le Studio 4 par Max/MSP ne comportait qu'une seule étape de réflexions intermédiaires, ce qui se voit aussi dans la mesure de diffusivité (figure 8b).

Rapport signal-sur-diffus Avant de se lancer dans l'implémentation complète du calcul de la diffusivité selon le rapport signal-sur-diffus (SDR) de Jarrett *et al.* [6] (voir la section 2.1.3), une étude préliminaire de la cohérence spatiale entre les composantes en harmoniques sphériques sur laquelle repose la définition du SDR est menée afin de vérifier sa capacité à décrire les caractéristiques directionnelles du champ sonore. L'idée est que la cohérence tendra vers 1 pour une paire de composantes décrivant la même information (par exemple entre la composante omnidirectionnelle $X_{0,0}$ et une autre contenant la direction d'arrivée de la seule onde plane d'un champ parfaitement directionnel), et vice-versa vers 0 pour une paire de composantes ayant aucune information commune (par exemple dans le cas d'un champ diffus, où chaque composante décrit une collection unique d'ondes planes décorréliées).

La figure 9 montre la cohérence pour les six combinaisons inter-composantes uniques possibles à l'ordre 1, moyennée sur les fréquences comme pour la diffusivité, pour les mêmes signaux de synthèse qu'à la figure 8. Pour une direction d'arrivée du son direct $\Omega_{\text{dir}} = (\theta_{\text{dir}}, \phi_{\text{dir}}) = (45^\circ, 45^\circ)$, toutes les composantes de l'ordre 1 décrivent en partie l'onde plane incidente, et la cohérence tend vers 1 pour toutes les combinaisons (parfaitement, dans le cas du signal numérique). On retrouve par ailleurs les mêmes imperfections pour l'enregistrement en Studio 4 que celles notées dans la mesure de la diffusivité par vecteur d'intensité : la cohérence entre la composante omnidirectionnelle et celle du premier ordre décrivant la "hauteur", $X_{1,0}$, appelée Z en B-format, ne tend ni parfaitement vers 1 ni idéalement vers 0 en raison de la forte réflexion au sol et de l'encodage sur une seule demie-sphère d'un champ diffus défini sur la sphère entière.

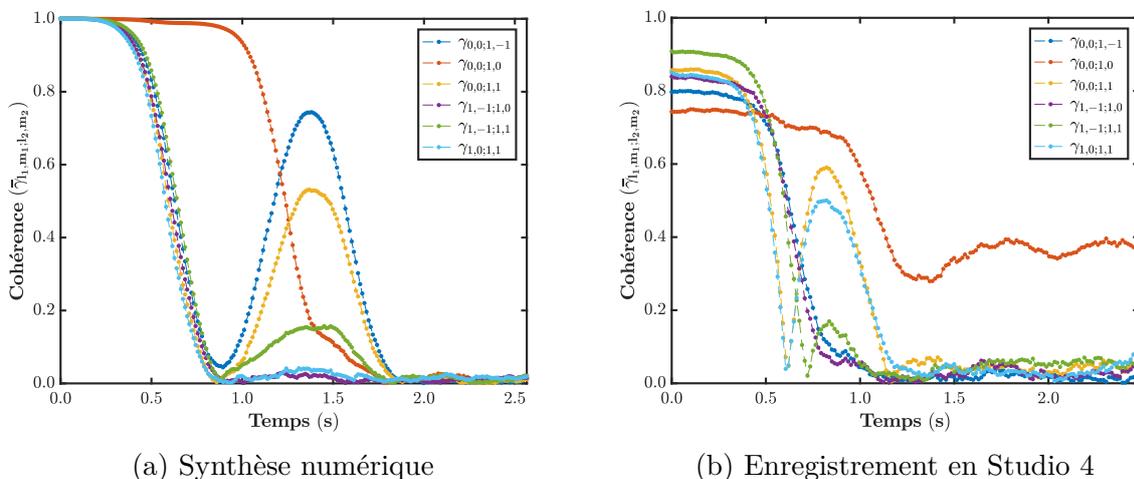


FIGURE 9 – Les six cohérences inter-composantes du premier ordre, calculées sur des signaux de synthèse encodés à l'ordre 4 ($\Omega_{\text{dir}} = (45^\circ, 45^\circ)$), moyennées sur les fréquences

La figure 10 est ensuite l'équivalente de la 8 pour la diffusivité calculée par SDR. On y voit que cette définition basé sur les cohérences inter-composantes semble moins sensible au nombre d'ondes planes utilisé pour simuler le champ diffus et sa valeur maximale se rapproche donc plus facilement de 1 pour un champ diffus qu'avec la méthode par vecteur d'intensité. Cependant, la fonction de pondération utilisée pour moyenner le SDR entre les différentes paires de composantes est calculée avec la direction d'arrivée du son direct (voir la section 2.1.3), ce qui non seulement nécessite son estimation au préalable, et donc le développement d'un algorithme dédié (voir ci-dessous), mais introduit aussi une source d'instabilité numérique due au fait que certaines harmoniques sphériques tendent vers zéro

pour certaines valeurs angulaires. Cette instabilité est surtout un problème pour l'analyse des signaux de synthèse numériques, où la détection de la direction d'arrivée peut donner des valeurs "trop" parfaites menant aux sus-dites extrêmes numériques.

Contrairement au vecteur d'intensité, la définition du SDR n'est pas limitée au premier ordre; les figures 28 et 29 (annexe A.1.2) comparent l'analyse complète à l'ordre 4 avec la version tronquée aux quatre premières composantes telle que présentée ici aux figures 9 et 10. Globalement, les différences minimales entre les deux résultats ne justifient pas le coût de calcul sévèrement plus élevé à l'ordre 4 (le nombre de composantes augmentant en $(L + 1)^2$).

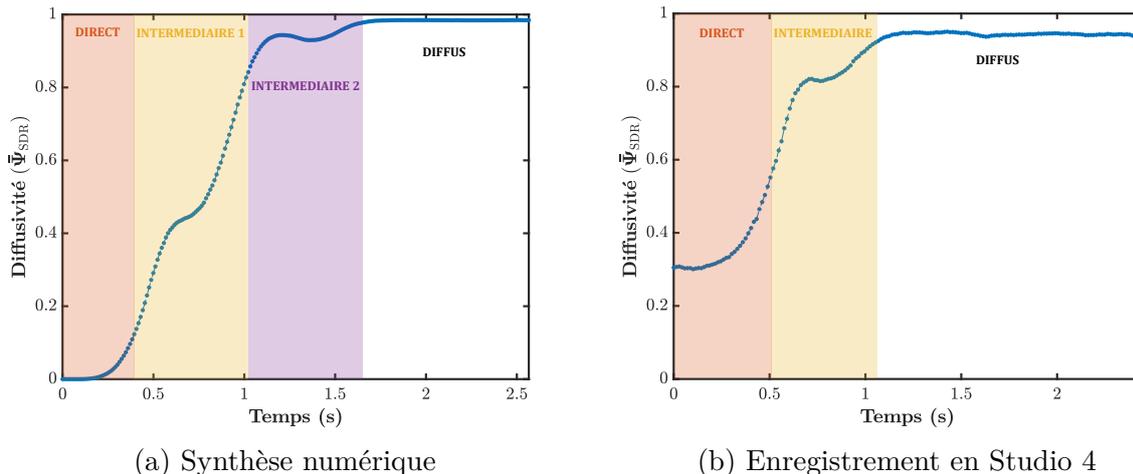


FIGURE 10 – Diffusivité par rapport signal-sur-diffus (SDR, $\bar{\Psi}_{\text{SDR}}$) calculée au premier ordre sur des signaux de synthèse encodés à l'ordre 4 ($\Omega_{\text{dir}} = (45^\circ, 45^\circ)$), moyennée sur les fréquences

À noter que la motivation derrière la construction des signaux évolutifs analysés dans toutes ces figures découle d'une première étude de la cohérence spatiale sur des signaux stationnaires plus ou moins directionnels ou diffus : ceux-ci sont présentés à l'annexe A.1.1.

Détection de direction d'arrivée Comme cela a été noté ci-dessus, le calcul de la fonction de pondération du SDR nécessite une estimation de la direction d'arrivée du son direct. Il s'agit donc de développer un algorithme basé sur les équations données par Merimaa et Pulkki [7] (présentées à la section 2.1.3), appliquées trame par trame (de STFT) afin d'obtenir une paire de valeurs azimut-élévation $\Omega_{\text{dir}} = (\theta_{\text{dir}}, \phi_{\text{dir}})$ à chaque point temps-fréquence. Dans un premier temps, ce calcul est opéré tel-quel sur une base de signaux numériques directifs quadrillant l'espace des valeurs angulaires possibles ; l'erreur résultante est cartographiée à la figure 30 (annexe A.2.1) et représente les limites numériques par rapport aux extrêmes mathématiques.

Suivant les étapes détaillées au début de cette section, l'algorithme est ensuite étendu à des signaux enregistrés par Eigenmike dans le Studio 4, ici des bruits blancs envoyés tour à tour de chaque haut-parleur. Les positions précises de ceux-ci étant connues, il est ensuite possible d'analyser l'erreur selon plusieurs critères. La figure 11a montre un exemple du processus d'estimation pour le haut-parleur n°11, où chaque point orange représente la valeur de Ω_{dir} à une fréquence discrète donnée, moyennée dans le temps (les signaux étant stationnaires sur 3 secondes). Ce calcul est fait pour les 24 haut-parleurs du studio, ce

qui permet d’obtenir l’erreur en fonction de la fréquence (figure 12b) moyennée sur cet ensemble. On trouve alors une région assez stable entre 300 Hz et 10 KHz sur laquelle on peut moyenner les valeurs fréquentielles de Ω_{dir} avec confiance, menant à l’estimation finale de la direction d’arrivée, représentée par le point jaune étoilé sur la figure 11a⁴. Le résultat de ce processus appliqué à l’ensemble des 24 haut-parleurs est présenté à la figure 11b.

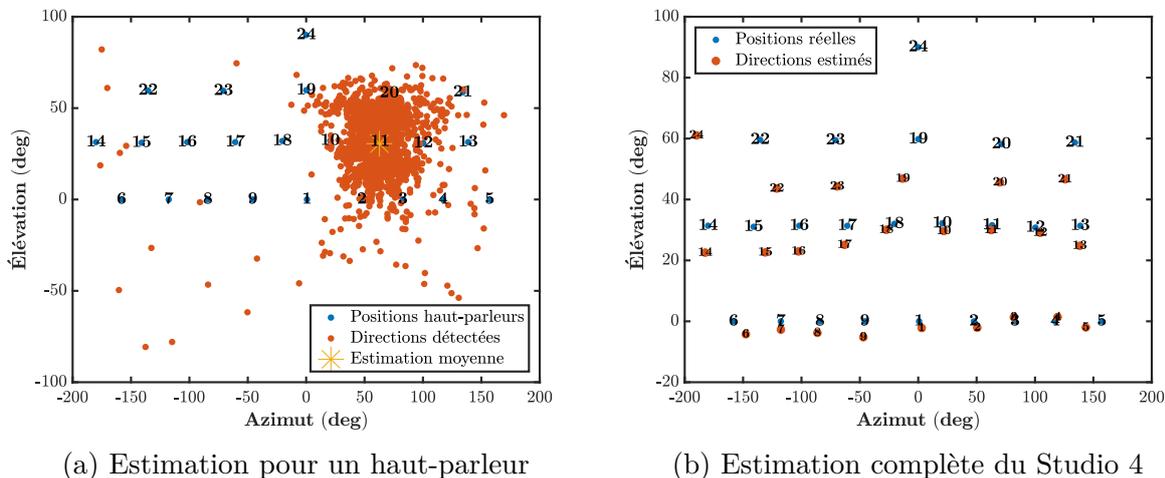


FIGURE 11 – Détection de direction d’arrivée sur le système de haut-parleurs du Studio 4 en enregistrant tour à tour par Eigenmike un bruit blanc provenant de chaque haut-parleur

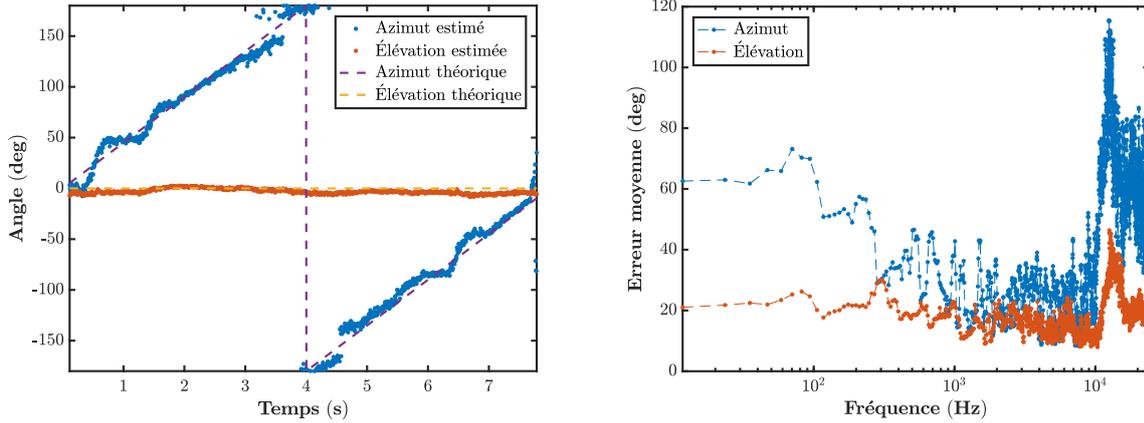
Des signaux dynamiques, et notamment des trajectoires azimutales à vitesse de rotation et élévation constantes, sont ensuite traitées afin d’approcher un cadre d’analyse plus approprié aux DRIR ; il n’y a donc plus de moyennage temporel outre le fenêtrage intrinsèque de la STFT. Les résultats, présentés à la figure 12a, démontrent un calcul plutôt précis de la direction d’arrivée, même sur une seule trame temporelle. Dans le cadre de l’analyse d’une DRIR, cette précision laisse alors présager une estimation sur la trame représentant au mieux le son direct, et compte-tenu de la discussion ci-dessus celle-ci peut être reliée à la valeur maximale de la cohérence (puisque le son direct devrait, par définition, être le plus directif de la DRIR). C’est donc de cette manière que le processus est intégré à l’algorithme complet de calcul de la diffusivité.

Finalement, en appliquant l’estimation de direction d’arrivée à la base de données entière des DRIR enregistrées par l’équipe EAC pour chaque haut-parleur de l’Espace de Projection (ESPRO) de l’IRCAM, on trouve une erreur moyenne de $8,65^\circ$ en azimuth et de $3,93^\circ$ en élévation, confirmant à la fois la performance globale de l’algorithme et le fait que le calcul de l’azimut y est la plus importante source d’erreur.

3.1.2 Algorithme de détermination du temps de mélange

L’algorithme responsable de la recherche du point temporel précis caractérisant au mieux le temps de mélange (t_{mix}) est schématisé à la figure 13 et les résultats de son application à la

4. Cette dernière moyenne inclut une étape supplémentaire de régularisation dans le cas où le calcul est faussé par des points de part et d’autre de la limite azimutale $\theta = \pm 180^\circ$ (voir la figure 31, annexe A.2.2). Elle est aussi pondérée en fonction de la distance euclidienne entre chaque point et la moyenne "de base".



(a) Détection d'une trajectoire azimutale

(b) Erreur en fonction de la fréquence

FIGURE 12 – Détection de direction d'arrivée d'un processus dynamique (trajectoire azimutale) et analyse de l'erreur en fonction de la fréquence (moyenne sur tous les haut-parleur du Studio 4)

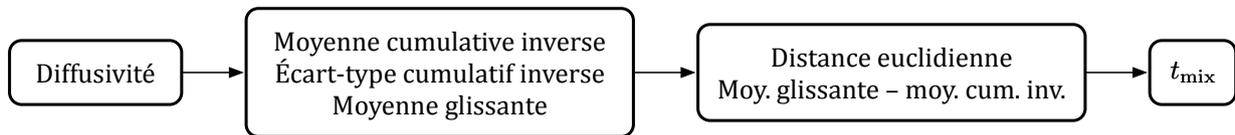


FIGURE 13 – Schématisation de l'algorithme de détermination du temps de mélange

DRIR mesurée à partir du haut-parleur n°36 de l'ESPRO (Espace de Projection) à l'IRCAM sont présentés à la figure 14. Les profils de ces courbes de diffusivité sont représentatifs de la grande majorité des DRIR analysées durant ce travail : un plateau est rapidement atteint à une valeur maximale après une première période d'instabilité due aux réflexions précoces. De ce fait, il semble judicieux d'essayer de définir le temps de mélange par rapport au début de la période stable, en trouvant une manière de caractériser le moment où le plateau est atteint.

Deux choses compliquent cependant l'analyse directe de la diffusivité : la "stabilité" perçue à vue d'œil n'est bien sûr qu'une tendance au sein de données plus ou moins bruitées, et le tout début du plateau peut être lui-même entrecoupé par l'effet d'une forte réflexion discrète mais assez tardive. Ce genre de baisse soudaine mais ponctuelle de la diffusivité est difficile à différencier du bruit global, caractérisé par la "largeur de bande" du plateau. Cette largeur est elle-même variante selon la stabilité numérique des valeurs des harmoniques sphériques à la direction d'arrivée du son direct (voir le paragraphe sur la détection de direction d'arrivée ci-dessus)⁵. Mais la prise en compte de ce type de réflexion est critique pour la fidélité du processus d'hybridation (voir la section 3.1.3 ci-dessous), car si le temps de mélange est défini avant la réflexion, celle-ci disparaît de la réverbération et la qualité perçue est fortement dénaturée.

En conséquence, l'algorithme proposé ici utilise trois critères descriptifs calculés sur la diffusivité donnée :

5. La largeur de bande du plateau varie aussi selon la plage de fréquences sur laquelle la diffusivité est moyennée ; au court de ce travail, des bornes fréquentielles ont été déterminées à la fois empiriquement et tenant compte de certaines considérations techniques au niveau de l'encodage en harmoniques sphériques des signaux sortant de l'Eigenmike. La diffusivité est donc moyennée entre 1000 Hz et 10 kHz.

1. Une moyenne cumulante "inverse", c'est-à-dire où chaque point représente la moyenne calculée à partir de ce point-là jusqu'à la fin,
2. Un écart-type cumulant inverse, analogue de la moyenne décrite ci-dessus,
3. Une moyenne glissante à trois points, réitérée autant de fois qu'il y a de points dans la diffusivité analysée.

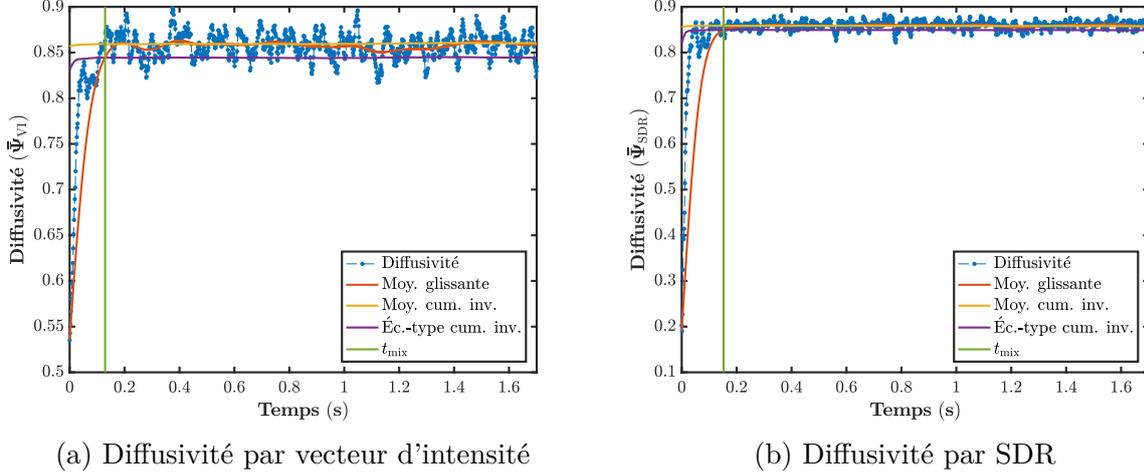


FIGURE 14 – Détermination du temps de mélange sur les deux mesures de diffusivité de la DRIR du haut-parleur n°36 de l'ESPRO

Ces critères permettent alors de définir le temps de mélange, t_{mix} , comme le premier point où la distance euclidienne (l_2) entre la moyenne glissante et la moyenne cumulée inverse est plus petite ou égale à l'écart-type cumulatif inverse. Sur les courbes de la figure 14 sont affichées, respectivement, la moyenne cumulée inverse, la moyenne cumulée inverse moins l'écart-type cumulant inverse, et la moyenne glissante, explicitant la définition donnée du point correspondant au temps de mélange.

Afin de tester sa robustesse et sa régularité, ainsi que de commencer à placer et à comparer les résultats obtenus de par son utilisation, l'algorithme est appliqué à la base de données entière des DRIR mesurées par l'équipe EAC dans l'ESPRO en configuration diffusante⁶. Pour une position de microphone (Eigenmike), cela veut dire les DRIR mesurées à partir de chacun des 75 haut-parleurs qui entourent la salle, ainsi que pour 3 haut-parleurs "flottants" placés dans l'espace (voir le schéma 32 à l'annexe A.3.1), pour un total de 78 haut-parleurs/DRIR. On travaille ici sur la position de microphone 2, mais des tests ont aussi été tournés sur la position 1, la configuration absorbante (positions de microphone 1 et 2), et les DRIR d'autres salles (la Kraftzentrale de Duisburg, la salle Fleuret au CNSM de Paris, la cathédrale San Lorenzo de Venise); les résultats de l'analyse de la Kraftzentrale sont présentés à l'annexe A.3.2.

La figure 15 compare alors les résultats obtenus à la fois par diffusivité, avec l'algorithme décrit ci-dessus, et par les méthodes statistiques présentées à la section 2.2.1 telles qu'implémentées dans le Spat~, avec comme référence la mesure de Polack [4] calculée pour l'ESPRO ($\sqrt{V} \simeq 65,2$ ms). Les critères statistiques (EDP [24] et kurtosis [25]), définis

6. Les murs et le plafond de l'ESPRO sont composés de panneaux acoustiques capable de trois comportements différents : diffusant, absorbant, et réfléchissant. La configuration "diffusante" dont il s'agit ici est celle où tous les panneaux sont en position diffusante.

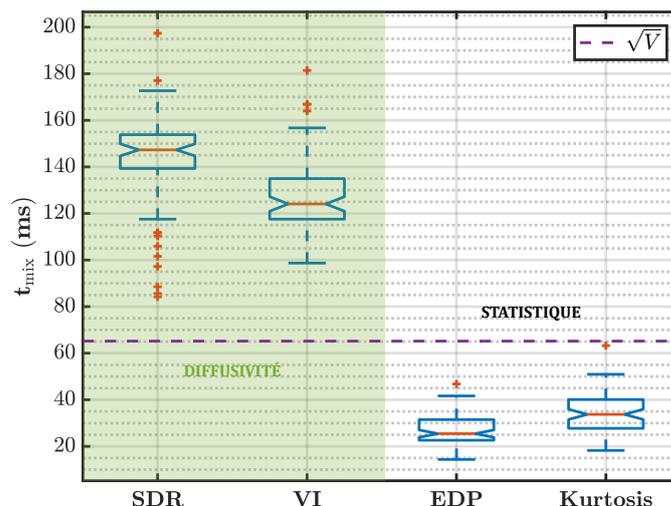


FIGURE 15 – Comparaison des différentes méthodes de caractérisation du temps de mélange telles qu’appliquées à la base de données entière des DRIR mesurées dans l’ESPRO en configuration diffusante : à partir de la diffusivité, par SDR et par vecteur d’intensité (VI), et à partir d’analyses statistiques, par profil de densité d’échos (EDP [24]) et par kurtosis [25], et avec la mesure de Polack [4] $\sqrt{V} \simeq 65,2$ ms pour l’ESPRO

pour des réponses impulsionnelles monophoniques, sont calculés sur chaque composante en harmoniques sphériques de la DRIR analysée, puis moyennés sur l’ensemble de ces composantes.

La détermination du temps de mélange à partir de la diffusivité donne globalement des valeurs plus élevées que les mesures statistiques, dû à la façon choisie de le caractériser dans l’algorithme développé. Certaines premières versions produisaient d’ailleurs des valeurs plus proches du critère de Polack, mais celles-ci étaient justement proies au problème des fortes réflexions tardives décrit ci-dessus. La véritable différence découle du fait que la définition "traditionnelle" du temps de mélange au sens de Schroeder et Polack est celle du moment à partir duquel le modèle stochastique commence à devenir valable, par rapport à des seuils de densités d’échos et de modes, tandis que dans le cadre présent du moteur de réverbération hybride il est nécessaire de représenter le moment à partir duquel le modèle stochastique peut être utilisé exclusivement, afin de décrire entièrement le contenu de la réponse impulsionnelle.

En comparant les deux méthodes basées sur la diffusivité, il semblerait que le SDR soit un peu plus robuste que le vecteur d’intensité, malgré le plus grand nombre de résultats au-delà de la moyenne $\pm 2,7\sigma$ (la définition par défaut des valeurs aberrantes dans la fonction `boxplot` de MATLAB, voir la figure 15; σ est l’écart-type), ce qui suit la tendance des courbes de diffusivité basées sur le SDR à être plus "stables" (globalement moins bruitées, voir la figure 14), sauf pour les sus-dites valeurs extrêmes d’angles de direction d’arrivée (ce qui peut éventuellement expliquer les résultats aberrants sur la figure 15). Compte-tenu de ces caractéristiques, il paraît alors judicieux d’utiliser une combinaison des deux mesures dans la définition finale du point de transition telle qu’utilisée au sein du réverbérateur hybride (voir ci-dessous).

3.1.3 Ajustement de l'algorithme d'hybridation

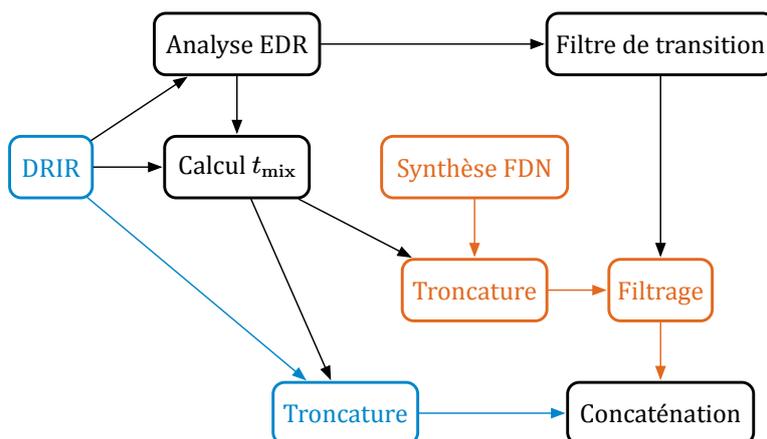
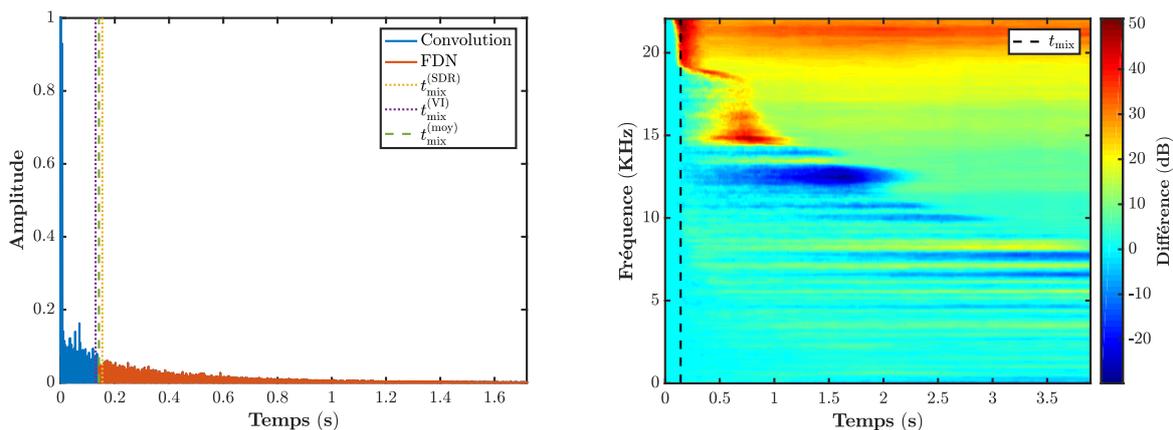


FIGURE 16 – Schématisation simplifiée de l'algorithme d'hybridation tel qu'implémenté en MATLAB avec la détermination automatique du temps de mélange t_{mix}

Comme cela a été évoqué en première partie de la section 2, un certain travail a déjà été entamé au sein de l'équipe EAC afin d'étendre l'algorithme d'hybridation (voir la figure 2) à des réponses impulsionnelles directionnelles (DRIR). L'implémentation existe pour l'instant en MATLAB, et donc non en temps-réel, avec un réglage manuel et empirique du temps de mélange; dans le cadre de ce stage, il s'agit alors de remplacer ce paramétrage fixe par le calcul automatique décrit ci-dessus (section 3.1.2). L'algorithme résultant de cet ajustement est schématisé à la figure 16 (de manière très simplifiée), où les couleurs suivent celles de la figure 17 pour les parties "convolutives" (la DRIR d'origine) et de synthèse par FDN. En réalité, ce processus hors temps-réel génère une DRIR hybride produite par concaténation de la DRIR d'origine, entre $t = 0$ et $t = t_{\text{mix}}$, avec le résultat d'une synthèse par FDN (en passant par les modules *early* et *cluster* également mentionnés à la section 2) en réponse à un signal impulsionnel idéal, à partir de $t = t_{\text{mix}}$.



(a) Échogramme de la DRIR hybride

(b) Différence en EDR avec la DRIR d'origine

FIGURE 17 – Réponse impulsionnelle hybride générée par l'algorithme d'hybridation ajusté, implémenté en MATLAB avec une détermination automatique du temps de mélange

La figure 17a montre l'échogramme de la composante omnidirectionnelle de la réponse impulsionnelle hybride produite à partir de la DRIR mesurée du haut-parleur n°36 de

l'ESPRO (le même qu'à la figure 14), tandis que la différence entre les EDR (profils de décroissance énergétique, voir la section 2.1.2) de la DRIR hybride et celle d'origine est montrée à la figure 17b. L'échogramme (figure 17a) explicite la concaténation des deux parties (réelle et synthétique) de la DRIR hybride, avec en tant que choix de point de mélange la moyenne entre les mesures par vecteur d'intensité et SDR décrites ci-dessus (section 3.1.2; le choix de moyenner est fait par souci de robustesse). Le temps de mélange semble bien intervenir après les principales réflexions précoces visibles sur l'échogramme et la synthèse par FDN continue la décroissance exponentielle que l'on voit apparaître après le son direct. La différence entre les EDR ($10 \log_{10}[\text{EDR}_{\text{hyb}}] - 10 \log_{10}[\text{EDR}_{\text{mes}}]$, figure 17b) confirme qu'avant le temps de mélange (ligne hachée noire), la DRIR hybride est identique à celle d'origine. Les principales différences après le temps de mélange sont dues à des aspects techniques de l'analyse de la DRIR mesurée et de la synthèse par FDN (telles qu'implémentées pour l'instant dans les bibliothèques MATLAB de l'équipe EAC), notamment par rapport à la détermination du T_{60} et à la définition des bandes de fréquences du FDN (voir la figure 35 à l'annexe A.4.1)⁷. Des petites erreurs dans le profil énergétique et le temps de réverbération au moment de la transition s'accroissent alors dans le temps, mais les différences les plus importantes sont néanmoins globalement contenues à des plages de fréquences peu importantes perceptivement (aux alentours de 15 KHz et au-delà); un résumé de l'erreur dans le processus d'hybridation est présenté à la table 1 pour la même DRIR du haut-parleur n°36 de l'ESPRO.

Différence moyenne globale	9,86 dB
Différence moyenne sur les bandes de fréquences	4,16 dB
Différence maximale globale	51 dB

TABLE 1 – Erreur dans l'hybridation de la DRIR du haut-parleur n°36 de l'ESPRO en termes des différences entre les EDR de la DRIR hybride et celle d'origine

Évaluation perceptible à l'écoute Un patch Max/MSP permet d'évaluer rapidement les résultats de l'hybridation en utilisant l'objet de convolution temps-réel du `Spat~` (`spat5.conv~`) et en passant de la DRIR hybride à celle d'origine (voir la figure 36 à l'annexe A.4.2). Quoique non contrôlée perceptivement et dotée d'un côté subjectif non négligeable (voir la section 4), cette petite expérience donne malgré tout un aperçu utile de la performance de l'algorithme d'hybridation, surtout par rapport au nombre de bandes de fréquences et de canaux de FDN nécessaires pour une synthèse fidèle à la réverbération mesurée, ainsi que l'ordre de grandeur du temps de mélange; en effet, l'utilisation de la moyenne entre les deux mesures par diffusivité semble donner régulièrement une DRIR qui préserve le caractère essentiel des réflexions précoces.

3.2 Synthèse de la réverbération des espaces couplés

Comme présenté à la section 2, le deuxième objectif de ce stage est d'étendre les capacités de la synthèse par FDN utilisée dans le moteur de réverbération hybride à des profils

7. La cause est essentiellement liée à l'ordre relativement faible des filtres utilisés pour créer les bandes de fréquences du FDN (par souci d'implémentation temps-réel en Max/MSP à l'origine) conjugué au nombre limité de bandes sur lesquelles le temps de réverbération doit être réduit (16 bandes "logarithmiquement" égales entre 20 Hz et 20 KHz dans l'exemple présenté ici).

de décroissance plus complexes, et notamment, dans un premier temps, les décroissances dites en "double-pente", caractéristiques des espaces composés de deux volumes couplés (voir la figure 7). Afin de déterminer la bonne paramétrisation du FDN, le processus de synthèse nécessite au préalable une analyse correspondante, à l'image du cas "simple" utilisé jusqu'ici dans les algorithmes d'hybridation. Un cadre est donc développé d'abord à partir des outils déjà mis en place par l'équipe EAC, puis en utilisant les méthodes présentées aux sections 2.1.4 et 2.2.2, tout en suivant la théorie de base proposée par Cremer et Müller [9]. Ce modèle théorique est ensuite adapté afin de paramétrer une synthèse par FDN équivalente en utilisant les résultats de l'analyse ; le processus complet est validé en analysant des signaux de synthèse aux paramètres connus. Il est à noter que l'on s'occupe ici d'une réverbération entièrement basée sur la modélisation stochastique, représentant donc la partie tardive de la réponse impulsionnelle, sans se soucier des réflexions précoces qui ne sont pas synthétisées ni modélisées dans le cadre du réverbérateur hybride.

3.2.1 Implémentation des méthodes d'analyse

Afin de s'attaquer à l'analyse de la réverbération des volumes couplés, le travail présenté ici se base sur les techniques déjà mises en place par l'équipe EAC, brièvement décrites à la section 2.1.2, et qui consistent essentiellement à extraire les principaux paramètres du modèle stochastique à partir de l'EDR de la réponse impulsionnelle. Plus précisément, ce sont le temps de réverbération (T_{60}), le profil énergétique à $t = 0$ (P_0), et le point temporel (t_{lim}) où l'EDC de chaque canal fréquentiel passe sous le seuil de bruit correspondant (P_{noise}) qui sont caractérisés. Ces techniques sont étendues en intégrant des méthodes d'analyse plus adaptées à la réverbération à pentes multiples (voir encore les sections 2.1.4 et 2.2.2).

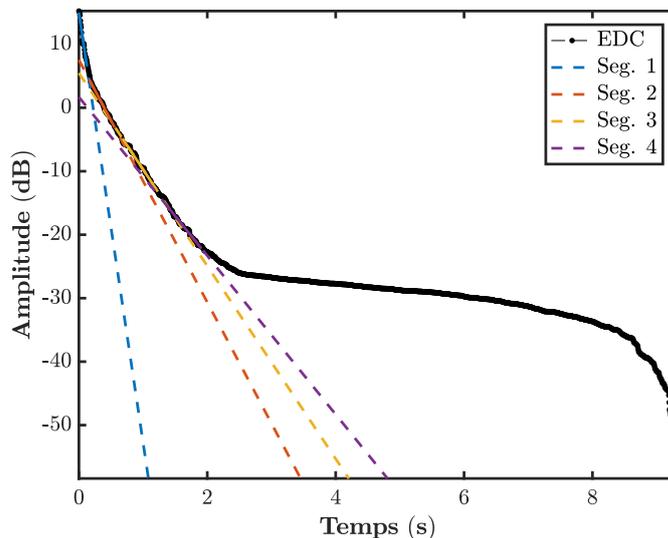


FIGURE 18 – Segmentation par régression linéaire locale selon Luizard et Katz [11] ; le seuil de bruit est évité en fixant un T_{60} maximal par segment égal à la longueur temporelle de l'EDC

Détection par segmentation Les outils d'analyse de l'équipe EAC se basent sur une segmentation de l'EDC à chaque fréquence discrète par régression linéaire locale, soumise à un critère de déviation maximale ; grossièrement, si la régression linéaire du segment

en cours dévie de l'EDC analysée au-delà du critère, un nouveau segment est entamé, et ainsi de suite sur toute la courbe. Les segments sont ensuite évalués par rapport à d'autres critères de régularisation afin de déterminer les paramètres cités ci-dessus, ou sinon de marquer leur valeur à définir par interpolation une fois toutes les fréquences analysées. Dans le cas des pentes multiples, les segments peuvent être traités en fixant d'abord le nombre de pentes recherchées, puis en essayant de trouver la combinaison approchant au mieux l'EDC analysée. Si ce genre d'algorithme peut se montrer performant sur certaines EDC, les résultats obtenus sur toutes les fréquences d'un EDR en utilisant un seul ensemble de critères de recherche ne semblent pas assez réguliers ou robustes pour pouvoir justifier leur utilisation dans le cadre d'analyse-synthèse souhaité⁸.

Ce type d'analyse est finalement assez similaire à la méthode proposée par Luizard et Katz [11] pour les EDC à pentes multiples ; le résultat de l'implémentation de leur algorithme appliqué à l'EDC (canal 32, $f = 668$ Hz) d'un signal synthétisé par double FDN (voir ci-dessous, section 3.2.2) est présenté à la figure 18, où seulement les segments détectés avant le seuil de bruit sont conservés. Celui-ci est évité en fixant un temps de réverbération (T_{60}) maximal par segment égal à la longueur temporelle totale de l'EDC, supposant qu'une décroissance détectée de 60 dB ou moins sur toute cette longueur correspondrait à une courbe trop bruitée pour pouvoir en tirer une analyse correcte.

Les pentes ainsi trouvées peuvent ensuite être combinées de plusieurs manières afin de les réduire au nombre recherché. La plus performante de celles-ci semble être de fusionner successivement les pentes les plus similaires, avec une nouvelle régression linéaire sur la section de l'EDC recouvrant les segments concernés pour une meilleure estimation des paramètres (voir la figure 37 à l'annexe B.1.1). Cependant, ce processus souffre également des soucis de régularité et de robustesse mentionnés ci-dessus, y inclus les problèmes posés si un nombre trop faible de segments est détecté dès le départ.

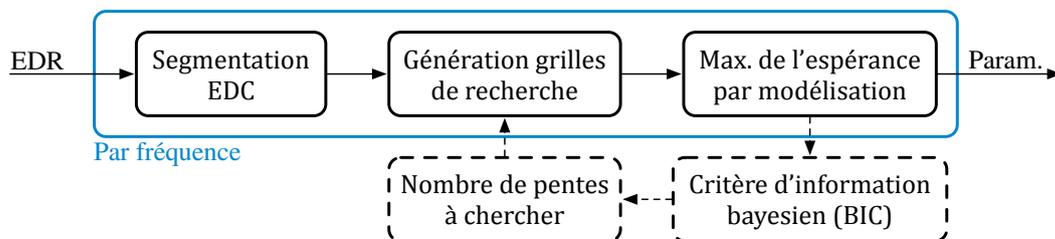


FIGURE 19 – Schématisation de l'algorithme d'analyse multi-pente par recherche bayésienne ; les parties hachées sont incluses dans le cas où le nombre de pentes recherchées est inconnu

Inférence bayésienne par modélisation Pour palier aux limites observées dans l'implémentation des techniques par segmentation décrites ci-dessus, une recherche par inférence bayésienne, suivant le travail de Xiang *et al.* [23] [10] et plus étroitement liée à la théorie de Cremer et Müller (voir la section 2.1.4), est finalement développée sur les bases du travail précédent ; sa structure générale est schématisée à la figure 19.

La mise en place du calcul direct du postérieur bayésien décrit à l'équation 21 sur l'espace entier des paramètres démontre que la génération des grilles de valeurs définissant cet

8. Et ce en plus d'une complexité algorithmique plutôt excessive compte tenu des traitements en question.

espace est primordiale et critique à la bonne performance d'un algorithme basé sur ce genre d'inférence. L'idée est alors d'utiliser une segmentation de l'EDC telle que celle proposée par Luizard et Katz [11] afin de guider la création de ces grilles de valeurs et assurer une certaine régularité au niveau des résultats. Cependant, le coût de calcul d'un algorithme de "force brute" opérant sur tout l'espace des paramètres devient rapidement astronomique pour des pentes multiples, supposant la génération de grilles de valeurs assez fines pour assurer une bonne précision perceptive. Pour y remédier, le calcul complet peut être remplacé par un algorithme de maximisation de l'espérance (*Expectation-Maximization*, EM ; voir la schématisation à la figure 20), et c'est donc ce qui est proposé ici⁹.

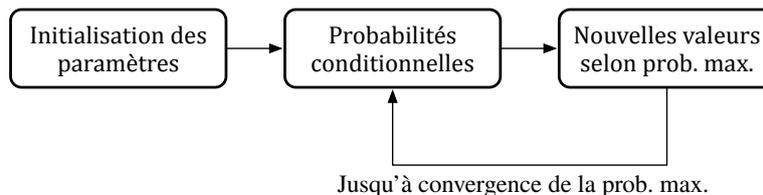


FIGURE 20 – Schématisation de l'algorithme de maximisation de l'espérance ; les probabilités conditionnelles sont calculées en utilisant la modélisation décrite par Xiang *et al.*

La différence principale entre l'algorithme de force brute et la maximisation de l'espérance est que le premier permet d'accéder au maximum global du postérieur (supposant la bonne définition de l'espace des paramètres, incluant les valeurs menant au maximum global), alors que le second garantit seulement l'atteinte d'un maximum local. Afin d'assurer que les résultats tendront vers le maximum global, l'EM (maximisation de l'espérance) est répétée autant de fois que le nombre total de paramètres à fixer (autrement dit, la dimensionnalité totale de l'espace des paramètres). Conjugué à la définition des grilles de valeurs informée par la segmentation de l'EDC, cela permet quand même de minimiser le coût de calcul de l'algorithme tout en produisant des résultats réguliers et robuste sur l'ensemble des fréquences de l'EDR.

La figure 21 compare l'EDR d'origine (figure 21a), tel que calculé directement sur un signal de synthèse, avec sa modélisation reconstruite à partir des valeurs des paramètres donnés par la recherche bayésienne décrite ci-dessus (figure 21b) ; un récapitulatif de l'analyse de l'erreur entre les deux se trouve à la table 2. Les signaux sont synthétisés en MATLAB avec deux FDN en suivant la théorie de Cremer et Müller (voir la section 3.2.2 ci-dessous), avec des T_{60} et des P_0 constants sur les bandes de fréquences des FDN ainsi qu'un bruit de fond continu à environ -72 dB. La figure 22 montre alors l'allure du temps de réverbération (figure 22a) et du profil énergétique à $t = 0$ (figure 22b) extraits de l'analyse du même signal qu'à la figure 21, c'est-à-dire synthétisé avec des T_{60} de 0,7 s et 4,1 s respectivement pour un $\Delta P_0 = P_0^{(1)} - P_0^{(2)} \simeq 18$ dB. Il est à noter que les FDN sont paramétrés de la même manière qu'à la section 3.1.3, et l'on retrouve donc les mêmes artefacts dûs à la définition des bandes de fréquences. Mais l'implémentation du calcul d'une moyenne sur les fréquences pondérée selon le standard ITU-R 468 (techniquement ITU-R BS.468-4), plus pertinent perceptivement, permet de comparer les résultats par rapport aux valeurs utilisées pour la synthèse.

9. Xiang *et al.* contournent ce même problème avec des techniques d'échantillonnage de Monte-Carlo, et notamment l'échantillonnage de Gibbs, mais il est à noter que leur but est quand même d'obtenir un aperçu global de la probabilité *a posteriori*, alors que dans le cas présent on cherche "simplement" les valeurs optimales des paramètres.

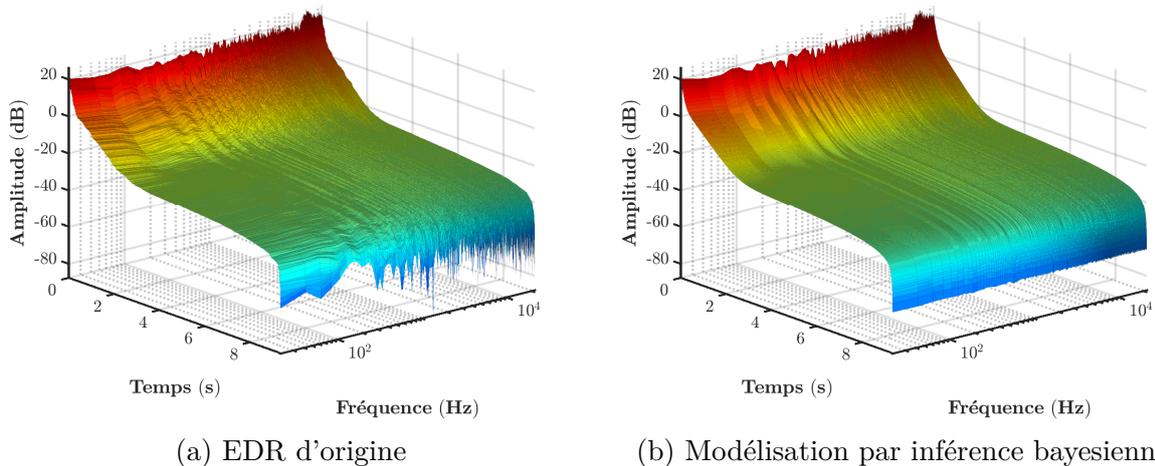


FIGURE 21 – Profils de décroissance énergétique (EDR) : calculé directement sur le signal de synthèse et reconstruit à partir des résultats de recherche par inférence bayésienne

Erreur moyenne EDR modélisé	0,326 dB par trame temporelle
Erreur maximale EDR modélisé	0,656 dB/trame à 3,92 kHz
Erreur $T_{60}^{(1)}$ (moy. R468)	4,3%
Erreur $T_{60}^{(2)}$ (moy. R468)	3,89%
Erreur ΔP_0 (moy. R468)	38,8%

TABLE 2 – Erreur dans l'analyse de l'EDR du signal de synthèse par rapport aux valeurs théoriques connues des paramètres telles que données par la théorie de Cremer et Müller

Finalement, afin de réduire la variance dans les résultats et régulariser encore un peu plus les valeurs obtenues, un algorithme de lissage fréquentiel logarithmique par noyau (*logarithmic kernel smoothing*)¹⁰ est implémenté et appliqué à l'ensemble des paramètres. Globalement, ce processus donne non seulement les paramètres nécessaires pour recréer la réverbération à pentes multiples avec un nombre équivalent de FDN, mais aussi, en théorie, les ingrédients nécessaires pour remonter à une caractérisation physique du système couplé (voir la prochaine section 3.2.2 ainsi que la 4.2 ci-dessous).

Il est aussi à noter que la recherche bayésienne peut être ainsi implémentée de manière "informée", en fixant un nombre de pentes connu, ou de manière "non-informée" en calculant un critère d'information bayésien (*Bayesian Information Criterion*, BIC) (tel que définit par Xiang *et al.* [10]) pour un nombre de pentes cherchées augmentant à chaque itération. Si le BIC commence à chuter à l'itération $N + 1$, on peut alors considérer que N est le nombre de pentes le plus approprié, et l'on peut extraire les paramètres correspondants. La version non-informée de l'analyse requiert donc les parties hachées du schéma présenté à la figure 19.

10. Tel qu'implémenté ici, le lissage est obtenu en calculant une moyenne glissante pondérée par une fenêtre choisie dont la taille augmente d'un échantillon à chaque fois que l'indice présent est égal au carré du nombre de fois que la taille a été augmentée; c'est-à-dire que le premier point a une fenêtre de taille 1, les points 2 à 4 ont une fenêtre de taille $2^2 = 4$, les points 5 à 16 ont une fenêtre de taille 16, et ainsi de suite jusqu'à ce que la fenêtre devienne "trop grande" par rapport au nombre de points restants, à quel point celle-ci est alors simplement prise à la taille maximale possible jusqu'au point final.

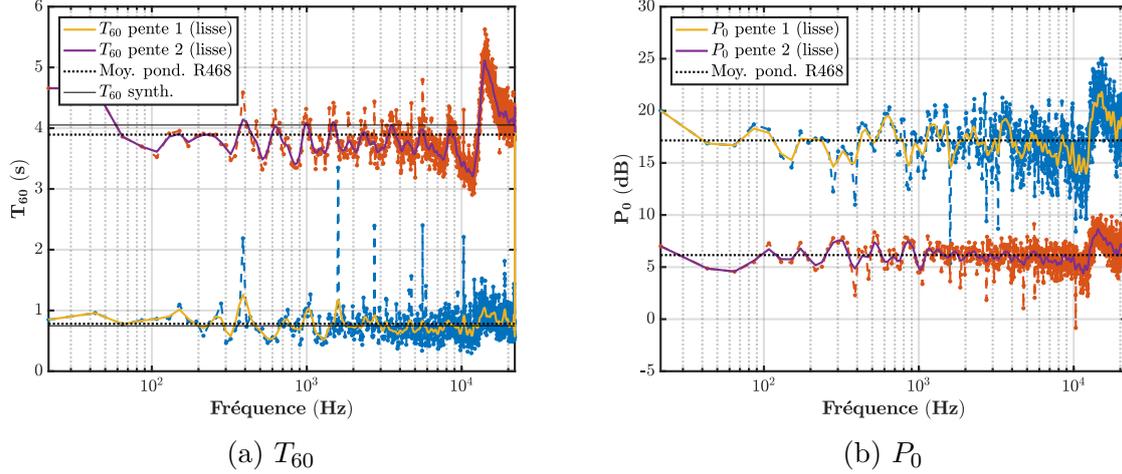


FIGURE 22 – Extraction du temps de réverbération (T_{60}) et du profil énergétique à $t = 0$ (P_0) à partir des résultats de recherche par inférence bayésienne

3.2.2 Synthèse par double FDN

Synthétiser la réverbération (stochastique, donc tardive) de deux espaces couplés avec deux FDN correspondants nécessite un ajustement de la théorie de Cremer et Müller [9] afin de "traduire" les considérations physiques de leur modèle en paramètres de contrôle. Dans un premier temps, ceci est fait directement en MATLAB et en suivant de près la description de Cremer et Müller. Les relations entre les différents paramètres sont ensuite plus rigoureusement caractérisées afin de mener à une implémentation temps-réel en Max/MSP où le transfert d'énergie acoustique est directement simulé en faisant passer les deux FDN l'un dans l'autre.

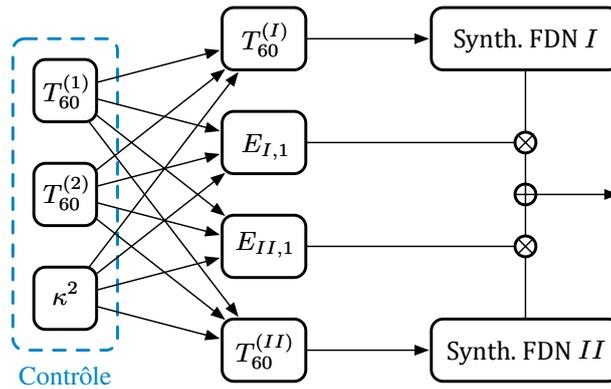


FIGURE 23 – Schématisation de l'algorithme de synthèse de réverbération couplée par double FDN tel qu'implémenté en MATLAB à partir de la théorie de Cremer et Müller [9]

Modèle informé physiquement Le modèle de la réverbération des volumes couplés proposé par Cremer et Müller [9] permet de décrire la décroissance énergétique de chacun des deux espaces en fonction des paramètres physiques du système, en particulier les volumes V_1 et V_2 , les "surfaces équivalentes d'absorption" $A_{1,0}$ et $A_{2,0}$, et la surface de couplage $S_{1,2}$ (voir la section 2.1.4). Dans le contexte du travail présent, il semble cependant intéressant d'essayer de réduire ces caractéristiques à des paramètres de contrôle plus pertinents,

idéalement les temps de réverbération T_{60} des deux espaces ainsi qu'un seul coefficient de couplage, en l'occurrence la moyenne géométrique κ^2 introduite par Cremer et Müller.

Cette simplification est possible en supposant, comme le font Cremer et Müller, que la condition initiale $E_{0,2}$ dans l'espace V_2 couplé au volume V_1 contenant la source peut être décrite par $E_{0,2} = \kappa^2 E_{0,1}$. Alors les amplitudes des deux processus de réverbération, pris du point de vue d'un récepteur dans le même espace que la source et tels qu'affectés par le couplage, peuvent être réécrites à partir des équations 17 et 18 :

$$\begin{cases} E_{I,1} = \frac{E_{0,1}[1 - \kappa^2/(1 - \delta_{II}/\delta_1)]}{1 - \kappa^2/(1 - \delta_I/\delta_2)(1 - \delta_{II}/\delta_1)} \\ E_{II,1} = \frac{E_{0,1}\kappa^2[1 - 1/(1 - \delta_I/\delta_2)]}{1 - \delta_{II}/\delta_1 - \kappa^2/(1 - \delta_I/\delta_2)} \end{cases} \quad (22)$$

Les facteurs de décroissance δ_1 et δ_2 sont reliés aux $T_{60}^{(1)}$ et $T_{60}^{(2)}$ par $\delta_i = 3 \ln(10)/T_{60}^{(i)}$ et l'équation 15 montre que les facteurs ajustés δ_I et δ_{II} dépendent alors seulement des temps de réverbération de base et du facteur de couplage moyen κ^2 ; les temps de réverbération ajustés $T_{60}^{(I)}$ et $T_{60}^{(II)}$ sont ensuite simplement obtenus en inversant la relation ci-dessus : $T_{60}^{(j)} = 3 \ln(10)/\delta_j$. Ces valeurs ajustées paramètrent finalement les deux FDN représentant les deux processus de réverbération du système couplé (voir l'équation 16) qu'il suffit de sommer en multipliant chacun par son amplitude respective selon l'équation 22 (traduites d'énergie en signal par leur racine carrée).

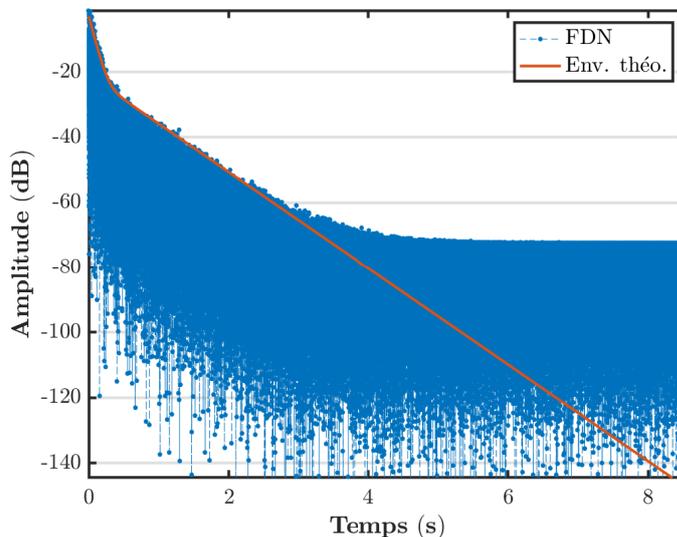


FIGURE 24 – Synthèse d'un processus de réverbération à double pente par deux FDN paramétrés et sommés selon la théorie de Cremer et Müller [9]

La figure 24 montre le résultat d'une synthèse ainsi réalisée, avec l'enveloppe théorique calculée selon l'équation 16 superposée à l'amplitude du signal en decibels. Cet exemple est celui analysé aux figures 18, 21, et 22, avec $T_{60}^{(1)} = 0,75$ s, $T_{60}^{(2)} = 4$ s, $\kappa^2 \simeq 0,01$, et un bruit de fond constant à environ -72 dB, ce qui donne un rapport $E_{I,1}/E_{II,1} \simeq 18$ dB (correspondant au ΔP_0 du côté de l'analyse, voir la section 3.2.1 ci-dessus).

Analyse paramétrique du modèle Comme l'indiquent Cremer et Müller [9], un système de deux volumes couplés ne donne pas nécessairement lieu à une réverbération en double-pente : en fonction des paramètres du modèle, les relations entre les différentes amplitudes $E_{I,1}$, $E_{II,1}$, $E_{I,2}$ et $E_{II,2}$ produisent divers profils de décroissance. Puisque le présent travail se focalise sur les combinaisons de paramètres générant les double-pentes en particulier, il est important d'approfondir quelque peu la caractérisation de leur comportement et d'en tirer des renseignements capables de guider à la fois la paramétrisation de la synthèse et l'évaluation des résultats d'analyse.

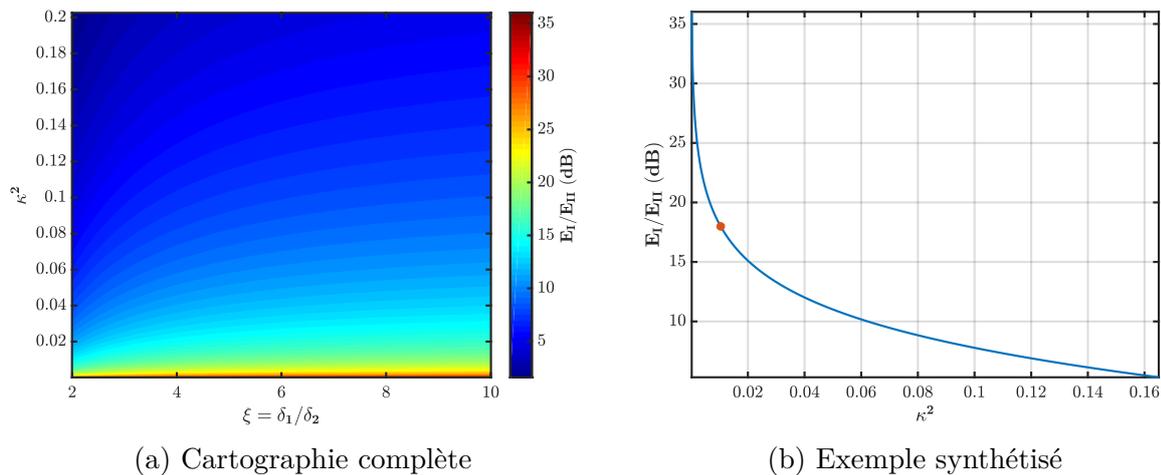


FIGURE 25 – Analyse de l'évolution du rapport d'amplitude E_I/E_{II} en fonction des paramètres de couplage ; à droite, le point orange indique la valeur utilisée pour l'exemple de la figure 24

Cremer et Müller démontrent que seul un récepteur dans le même espace que la source, et dans le cas où ce premier volume est moins réverbérant que celui auquel il est couplé, peut percevoir une réverbération à deux décroissances distinctes [9]. Il est donc seulement nécessaire d'utiliser la première expression pour E_I à l'équation 16, et l'on peut se passer de l'indice $.1$ dans l'écriture notamment du rapport des amplitudes E_I/E_{II} . En définissant ensuite un rapport entre les facteurs de décroissance séparés, $\xi = \delta_1/\delta_2$, il est possible d'exprimer E_I/E_{II} uniquement en fonction de celui-ci et de κ^2 (et des facteurs de décroissance couplés δ_I et δ_{II} calculés à partir de ξ et κ^2). La figure 25a montre la cartographie de cette relation pour ξ allant de 2 à 10 ; la borne supérieure de κ^2 est déterminée en considérant qu'un rapport E_I/E_{II} de moins de 6 dB devient pratiquement imperceptible en tant que double-pente (voir la figure 38 à l'annexe B.2).

Finalement, la figure 25b montre la courbe du rapport E_I/E_{II} pour l'exemple de synthèse présenté à la figure 24 (et analysé aux figures 18, 21, et 22), en fonction des temps de réverbération $T_{60}^{(1)}$ et $T_{60}^{(2)}$ réglés en paramètres de contrôle. Cette courbe permet ainsi de viser un rapport E_I/E_{II} précis pour la synthèse (18 dB dans ce cas), ce qui peut être beaucoup plus intuitif que de fixer une valeur de κ^2 sans réelle référence. Globalement, plus le couplage est faible, moins il y a de transmission de l'énergie sonore entre les deux espaces, et plus le rapport E_I/E_{II} augmente. Aux deux extrêmes, il semble y avoir des limites perceptives : soit le rapport est trop élevé et la réverbération du deuxième volume tombe sous le seuil de bruit et devient inaudible, soit le rapport est trop faible et il devient impossible de discerner entre les deux réverbérations.

4 Travail ultérieur

Le travail accompli au long de ce stage, décrit à la section 3 précédente, représente une certaine partie des objectifs de recherche nécessaires au développement complet du réverbérateur hybride spatialisé mis en œuvre par l'équipe EAC [1]. Cependant, compte tenu du temps alloué et des ressources disponibles (accès au studio, etc.), plusieurs aspects ont dû être relégués à un approfondissement futur, qui pourra vraisemblablement s'inscrire dans le cadre d'une thèse de doctorat à venir sur la manipulation des réponses impulsionnelles spatialisées de manière plus générale, encadrée à nouveau par Markus Noisternig (et en collaboration avec Thibaut Carpentier). Les points identifiés les plus immédiats sont esquissés ici afin de donner une idée des prochaines étapes possibles, et une discussion plus globale des perspectives d'étude conséquentes est aussi proposée pour cadrer l'évolution du réverbérateur hybride par rapport aux axes de recherche à la fois de la thèse et de l'équipe EAC.

4.1 Temps de mélange

Si les développements relatifs à la caractérisation automatique du temps de mélange d'une réponse impulsionnelle directionnelle (DRIR), décrits à la section 3.1, vont une certaine distance dans la quête des objectifs énoncés en introduction, il reste néanmoins plusieurs points à approfondir ; particulièrement par rapport à la validité perceptive finale des travaux entrepris, ainsi que leur stabilité et leur robustesse algorithmique.

4.1.1 Validation perceptive

Perceptivement, il serait nécessaire d'établir si le temps de mélange déterminé par l'algorithme proposé correspond bien à une appréciation intuitive de ce paramètre, à tel point qu'il pourrait mener à une reproduction hybride indiscernable de la DRIR d'origine. Cela peut passer par différents types d'expériences, en cherchant notamment à comparer la caractérisation par diffusivité à une mesure "naturelle" correspondante, mais aussi en regardant "simplement" la capacité à discerner entre une DRIR hybride générée selon la méthode de la section 3.1.3 et celle d'origine (ou un son plus complexe convolué avec). Si des études liées à ce sujet existent dans la littérature, il semblerait que l'idée du réverbérateur hybride n'ait pas encore été exploitée afin de caractériser là où un utilisateur placerait instinctivement le temps de mélange.

Le paradigme du moteur de réverbération hybride permet donc de comparer le réglage perceptif du temps de mélange en tant que paramètre de contrôle avec la détermination automatique proposée ici, moyennant la constance et la précision de tous les autres paramètres (ce qui devrait être possible en tenant compte de certains aspects techniques de l'analyse de la DRIR et de la synthèse par FDN). Ceci donne non seulement accès à une évaluation critique de la performance des algorithmes développés au long de ce travail, mais aussi un aperçu des caractéristiques perceptives de cette propriété souvent que très superficiellement comprise.

4.1.2 Moyennage des mesures de diffusivité

Suivant dans la veine de ces considérations perceptives, il semblerait pertinent d'en prendre en compte certains aspects dans la manière dont les mesures de diffusivité sont moyennées sur les fréquences. Tel qu'implémenté ici, l'algorithme utilise simplement une moyenne sur une plage de fréquences jugée empiriquement "stable" par rapport aux caractéristiques de l'encodage des DRIR en Ambisonics (voir la note à la section 3.1.2). Non seulement serait-il utile de caractériser plus rigoureusement l'étendue de cette plage en fonction, par exemple, de l'antenne de microphones utilisée et de l'ordre d'encodage choisi, mais "l'importance" relative des différentes fréquences pourrait aussi être prise en compte par une pondération perceptive correspondante (comme par exemple la norme ITU-R 468 utilisée à la section 3.2.1).

4.1.3 Stabilité de la diffusivité par SDR

Une des causes du plus grand nombre de résultats aberrants pour le temps de mélange par diffusivité à base de SDR que celui à base de vecteur d'intensité, tel que calculé sur toutes les DRIR de l'ESPRO en configuration diffusante et présenté à la figure 15, peut être reliée à la sensibilité de ce premier à la valeur de la direction d'arrivée sur laquelle il dépend (voir les sections 2.1.3 et 3.1.1). Afin d'obtenir une application plus robuste, il serait donc nécessaire de cerner la source précise de cette instabilité numérique autour des valeurs d'azimut ou d'élévation approchant zéro et d'y remédier en définissant des cas limites donnant lieu à des méthodes de calcul alternatives, ce qui demande au préalable une étude numérique rigoureuse du calcul du SDR selon Jarret *et al.* [6].

4.1.4 Théorie et complexité algorithmique

Pour finir d'évaluer la performance et la solidité de l'algorithme proposé, il faudra également développer une description théorique complète en termes informatiques, ainsi qu'une analyse de sa complexité calculatoire. Jusqu'ici, les seules mesures de stabilité et de robustesse ont été limitées à une appréciation entièrement qualitative de l'application de l'algorithme à des bases de données plus ou moins extensives, menant encore une fois à des résultats tels que la figure 15. Avec une description mathématique approfondie et détaillée des différentes étapes, ces propriétés pourraient être plus rigoureusement investiguées, et par conséquent une estimation de la complexité des calculs serait aussi accessible.

4.1.5 Diffusivité par covariance

Malgré une première ébauche d'implémentation de la diffusivité calculée à partir de la matrice de covariance, telle que proposée par Epain et Jin [22] [8] et présentée à la section 2.1.3, il reste un certain travail à accomplir avant de pouvoir l'intégrer à l'algorithme de détermination du temps de mélange et d'en comparer les résultats aux deux autres mesures (SDR et vecteur d'intensité). Leur méthode semble notamment assez sensible aux propriétés d'encodage et aux paramètres de traitement de la DRIR, et en particulier le facteur d'oubli β (voir l'équation 12) qui, avec la taille de la fenêtre choisie, détermine l'échelle et la résolution temporelle de la mesure ; il serait intéressant d'obtenir au final une

précision comparable au SDR et au vecteur d'intensité afin de pouvoir les évaluer ensemble de manière égalitaire.

4.2 FDN couplés

L'analyse-synthèse de la réverbération des espaces couplés représente un sujet entièrement plus étendu que celui de la caractérisation du temps de mélange, et offre donc une série autant plus vaste de pistes potentielles à poursuivre. Au niveau de la synthèse, le travail de ce stage a surtout été basé sur l'adaptation du modèle théorique de Cremer et Müller [9] pour en tirer une paramétrisation équivalente des FDN ; cependant, le même lien est encore incomplet du côté de l'analyse, et il serait intéressant de pouvoir remonter de l'analyse d'une DRIR aux paramètres de couplage des volumes constituants. Cela dit, le travail théorique manque lui-même d'une certaine généralité dans le sens où le modèle de Cremer et Müller se limite au couplage de deux espaces : serait-il possible d'étendre leur approche à un nombre arbitraire de couplages ? De plus, leur modèle ne prend pas en compte les caractéristiques fréquentielles de la réverbération, comme le fait notamment l'EDR.

La relation entre les deux sous-sujets abordés dans ce stage peut paraître peu évidente par moments, mais ce travail sur une synthèse par FDN capable de prendre en compte les espaces couplés s'inscrit tout autant dans l'optique du développement de la réverbération hybride. Mais l'hybridation d'une DRIR à pentes multiples relève encore d'autres points importants à traiter, notamment par rapport à la définition (et par conséquent la détermination) du temps de mélange dans une salle composée de plusieurs volumes et donc non "typiquement mélangeante" au sens de Schroeder et Polack. Finalement, quelques aspects plus techniques de l'analyse-synthèse restent aussi à aborder, tels qu'une évaluation quantitative des résultats obtenus par le patch Max/MSP présenté à la section 3.2.2 (figure 26) ou une caractérisation de la complexité des algorithmes développés, entre autres.

4.2.1 Détermination des paramètres de couplage

À la section 3.2.2, la réverbération en double-pente caractéristique de deux espaces couplés est synthétisée avec deux FDN paramétrés selon la théorie de Cremer et Müller ; auparavant, à la section 3.2.1, un cadre d'analyse permettant d'extraire ces principaux paramètres est développé, mais pour l'instant celui-ci ne donne pas accès aux propriétés physiques du modèle. La détermination du facteur de couplage κ^2 serait particulièrement utile, permettant éventuellement de remonter ensuite aux temps de réverbération "séparés" $T_{60}^{(1)}$ et $T_{60}^{(2)}$. Malheureusement, puisque la théorie de Cremer et Müller définit les temps de réverbération couplés $T_{60}^{(I)}$ et $T_{60}^{(II)}$ comme les valeurs propres du système d'équations 14 (selon l'équation 15), il n'existe aucune formulation analytique capable d'exprimer $T_{60}^{(1)}$, $T_{60}^{(2)}$, et κ^2 en fonction de $T_{60}^{(I)}$, $T_{60}^{(II)}$, E_I , et E_{II} (i.e. les résultats actuels de l'analyse multi-pente décrite à la section 3.2.1). Il serait donc nécessaire de mettre en place une approche algorithmique, en exploitant par exemple une recherche de paramètres basée sur une connaissance théorique du modèle. Néanmoins, une telle détermination de ces paramètres offrirait une évaluation réellement complète des résultats de la synthèse et de l'analyse appliquée à celle-ci.

4.2.2 Extensions de la théorie de couplage

Dans la continuité du développement de ce lien entre la théorie de Cremer et Müller et l'analyse par EDR ainsi que la synthèse par FDN, deux autres ajustements majeurs seraient également à apporter à leur modélisation. Premièrement, il faudrait pouvoir la reformuler en tenant compte d'une évolution fréquentielle quelconque, ce qui a été fait implicitement au long du travail présenté à la section 3.2 mais qui aurait besoin d'être plus rigoureusement explicité afin de procéder aux développements décrits ci-dessus (section 4.2.1). Le processus d'analyse décrit à la section 3.2.1 étant théoriquement capable de chercher un nombre arbitraire de pentes, il serait ensuite intéressant de faire évoluer le cadre théorique de la même manière afin de garder la duplicité entre les deux. Cela demanderait à la fois une définition ainsi qu'une résolution algorithmique du système d'équations décrivant les transferts d'énergie acoustique entre les espaces couplés, comme le fait l'équation 14 pour le cas de deux volumes.

4.2.3 Hybridation multi-pente

Comme cela a été noté ci-dessus, le travail accompli sur l'analyse et la synthèse de la réverbération des espaces couplés s'inscrit dans le développement continu du réverbérateur hybride présenté en introduction ; il est donc nécessaire d'adapter les bases théoriques du processus d'hybridation afin de le rendre applicable aux pentes multiples. Les conséquences sur la définition du temps de mélange sont particulièrement importantes, puisque celui-ci est traditionnellement décrit pour des volumes "simples" : est-il encore valable de chercher un seul paramètre pour deux espaces couplés ? L'utilisation d'une mesure de la diffusivité paraît alors encore plus judicieux car celles-ci font intrinsèquement le lien avec la modélisation stochastique du signal et ne passent pas par une description physique des salles, facilitant la connexion avec la théorie de Cremer et Müller. Par conséquent, cela demandera en plus une manière d'évaluer la performance d'une détermination automatique du temps de mélange sur des DRIR à plusieurs pentes.

4.2.4 Vérification des résultats temps-réel

Si la paramétrisation du patch Max/MSP présenté à la section 3.2.2 (figure 26) semble donner, à l'écoute, des résultats plutôt cohérents, il est impératif de faire une comparaison rigoureuse de la réverbération produite par rapport aux réponses impulsionnelles générées en MATLAB (avec l'algorithme décrit en première partie de la même section 3.2.2). Puisque les paramètres ne sont pas fixés de la même manière dans les deux cas, cela nécessitera la définition d'un point commun autour duquel la comparaison pourra être basée ; dans ce but, il serait éventuellement judicieux d'aligner plus étroitement les deux méthodes de contrôle de la synthèse. Cette comparaison devrait être à la fois qualitative (prenant donc en compte certaines caractéristiques perceptives) et analytique par rapport aux principaux paramètres, c'est-à-dire à nouveau les temps de réverbération $T_{60}^{(I)}$ et $T_{60}^{(II)}$ et les amplitudes respectives E_I et E_{II} . Il est à noter que le patch Max/MSP de la figure 26 contient déjà un mécanisme d'enregistrement de réponse impulsionnelle monophonique (ou omnidirectionnelle) prévu à cette fin.

4.2.5 Complexité algorithmique de l’analyse

À l’instar des algorithmes développés durant la première partie de ce stage (sur la détermination du temps de mélange), ceux mis en place pour analyser la réverbération multi-pente nécessiteront également une caractérisation de leur complexité calculatoire, ce qui requiert de nouveau une formalisation théorique complète. Celle-ci existe pour certains aspects du processus d’analyse, notamment pour les calculs liés à la théorie de Cremer et Müller et la recherche bayésienne (voir la section 2.1.4), mais manque encore pour la fonction de segmentation et la couche de maximisation de l’espérance (EM) tels qu’ils ont été adaptés au cadre d’analyse en question. Il serait ensuite intéressant de considérer la complexité additionnelle du cas de la recherche d’un nombre de pentes inconnu, éventuellement en fonction du nombre réel de pentes, en sachant qu’une analyse non-informée de ce genre implique, pour N pentes réelles, des recherches jusqu’à $N + 1$ (voir la section 3.2.1).

4.2.6 Améliorations et considérations techniques

Au-delà des points élucidés ci-dessus, quelques détails plus techniques restent encore à régler au sein des développements réalisés pendant cette deuxième partie de stage. Deux d’entre eux relèvent d’une importance particulière : l’extraction des amplitudes de synthèse E_I et E_{II} à partir de l’analyse à pentes multiples, ce qui demande une équivalence qui n’est pas encore au point avec les valeurs de P_0 sortant de l’algorithme de recherche, et le souci de l’ordre des filtres utilisés pour séparer les bandes de fréquences des FDN, tel que noté à la section 3.1.2. La cause du premier point reste à déterminer précisément, mais les résultats présentés à la table 2 et à la figure 22 suggèrent une réduction de la différence ΔP_0 par un facteur d’à peu près 2 à un certain moment au cours des manipulations (comparé au rapport E_I/E_{II} en dB) ; de plus, l’analyse se faisant sur des courbes de Schroeder intégrées tandis que la synthèse est définie directement par rapport à la somme des décroissances couplées, il serait à vérifier que ces deux descriptions des amplitudes peuvent déjà être théoriquement équivalentes à la base.

Comme cela a donc été noté à la section 3.1.2, le problème des filtres définissant les bandes de fréquences des FDN est principalement dû à l’ordre utilisé pour leur synthèse au sein des outils MATLAB de l’équipe EAC, qui est pour l’instant un paramètre non ajustable. Cette limitation découle du fait que ceux-ci sont basés sur les versions développés pour la synthèse en temps-réel en Max/MSP ; il s’agirait alors de mettre en place une implémentation capable d’un ordre arbitraire, ce qui permettrait d’éviter les imprécisions autour des limites des bandes de fréquences (voir la figure 35 à l’annexe A.4.1). Il serait ainsi possible de définir un nombre de bandes plus élevé que les 16 utilisées au cours de ce travail, et donc par conséquence s’approcher d’une précision arbitraire au niveau de la synthèse par FDN. À noter qu’un tel développement permettrait aussi une amélioration de l’analyse des DRIR hybrides générées selon l’algorithme présenté à la section 3.1.3.

4.3 Perspectives générales

Globalement, le développement du moteur de réverbération hybride s’inscrit dans la branche de recherche de l’équipe EAC relative à la caractérisation physique des espaces acoustiques, même si le travail lui-même emprunte des autres axes (techniques de spatialisation du son par décomposition en harmoniques sphériques, perception cognitive du son spatialisé, etc.).

La thèse de doctorat qui fera suite au stage présenté ici propose d’agir dans sa continuité, mais en élargissant le cadre de recherche à la manipulation des réponses impulsionnelles tridimensionnelles de manière plus générale. Ainsi, la quasi-totalité des points détaillés ci-dessus pourront y être abordé, et de plus car certains apporteront des améliorations au-delà de la réverbération hybride.

Réciproquement, certains sous-sujets de la thèse pourront aider à étendre encore plus les capacités du réverbérateur hybride. Par exemple, il est proposé de travailler sur des manipulations directes des premières réflexions selon une représentation par décomposition en ondes planes (*Plane Wave Decomposition*, PWD), une extension en quelque sorte de la spatialisation en harmoniques sphériques. La partie convolutive de la réverbération hybride pourrait donc être elle-même manipulée aussi profondément que la synthèse par FDN, offrant au final une emprise presque totale sur les caractéristiques de la réverbération. Le sujet de thèse prévoit ensuite le développement d’un modèle perceptif à plusieurs couches permettant un contrôle plus ou moins haut-niveau de tous ces traitements et facilitant à la fois leur utilisation et leur évaluation.

5 Conclusion

Les principaux objectifs de ce stage, détaillés aux deux premières sections de ce rapport, sont réunis autour de deux points focaux au sein du développement plus général du moteur de réverbération hybride de l'équipe EAC (Espaces Acoustiques et Cognitifs) du STMS à l'IRCAM : la caractérisation et la détermination automatique du temps de mélange selon une mesure de la diffusivité du champ sonore décrit par une réponse impulsionnelle directionnelle (DRIR), et la mise en place d'un cadre d'analyse et de synthèse par réseaux de retards rebouclés (FDN) de la réverbération en double-pente des espaces couplés. Ce travail est mené dans le contexte du son spatialisé selon la décomposition en harmoniques sphériques, base du système de spatialisation Ambisonics. La réverbération hybride est basée sur une modélisation de la réponse impulsionnelle directionnelle d'une salle en deux parties principales (voir la figure 1) : une première contenant les réflexions précoces, discrètes et difficiles à synthétiser de manière fidèle, et une "queue" tardive pour laquelle un modèle stochastique synthétisable par FDN est valable. Le réverbérateur hybride exploite donc une convolution directe de la DRIR pour la partie précoce et une synthèse par FDN pour la deuxième partie stochastique ; ce moteur de réverbération peut être implémenté en temps-réel avec un algorithme de convolution efficace, ou sinon "offline" (hors temps-réel) en générant une DRIR hybride théoriquement équivalente à celle d'origine.

En première partie du stage (section 3.1), différentes mesures de la diffusivité sont implémentées et évaluées, sur des signaux types synthétiques ainsi que sur des DRIR, afin d'être validées comme bases pour le développement d'un algorithme de détermination automatique du temps de mélange. Le processus d'hybridation développé par l'équipe EAC [1] est alors ajusté de manière correspondante, et les DRIR hybrides ainsi générées sont quantitativement comparées à celles d'origine, évaluant à la fois la caractérisation du temps de mélange mise en place ici ainsi que la performance globale de l'hybridation. L'utilisation d'un critère basé sur des mesures de la diffusivité semble donner une caractérisation généralement stable et robuste du temps de mélange, moyennant certains détails plus techniques que fondamentaux qui restent néanmoins à corriger ultérieurement. De plus, des campagnes de tests perceptifs sont à prévoir afin de réellement valider la pertinence de cette méthode de détermination ; l'ensemble de ces travaux encore à venir sont rassemblés à la section 4.

À la section 3.2, l'extension de la synthèse par FDN à la réverbération en double-pente caractéristique des espaces couplés, point focal de la seconde partie du stage, est présentée avec le cadre d'analyse correspondant développé en parallèle. La description théorique et la paramétrisation conséquentes de la synthèse sont basées sur un modèle de couplage proposé par Cremer et Müller [9], tandis que le processus d'analyse combine une segmentation du profil de décroissance énergétique (EDR) avec une recherche par probabilité bayésienne selon la méthode de Xiang *et al.* [10] et une optimisation par maximisation de l'espérance (EM). Les paramètres extraits par cette analyse sont ensuite reliés au modèle de Cremer et Müller afin de produire une chaîne complète d'analyse-synthèse ; dans ce travail, le but étant de vérifier et d'évaluer la performance de ce processus, l'analyse est utilisée pour confirmer les résultats d'une synthèse paramétrée par des valeurs connues. Dans une application idéale, les FDN seraient à l'inverse paramétrés automatiquement par les résultats de l'analyse, à l'image de leur utilisation au sein de l'algorithme d'hybridation.

Ce travail de stage sera finalement étendu à une thèse de doctorat inscrite dans la continuité des recherches abordées, en se tournant plus généralement vers l'analyse, la synthèse, et

la manipulation temps-fréquence-espace complète des réponses impulsionnelles directionnelles. Celle-ci permettra accessoirement de traiter une grande partie des détails encore à régler suite aux développements réalisés au cours ce stage, mais elle représente surtout une étape de plus le long de l'axe de recherche entamé par l'équipe EAC relatif à la caractérisation physique des espaces acoustiques par leur réverbération. La plupart des méthodes et des processus utilisés ou mis en place dans ce stage seront ainsi exploités ou étendus lors de la thèse, et des considérations perceptives de l'expérience individuelle d'un espace acoustique seront intégrées aux aspects techniques et physiques de la manipulation des signaux afin de créer un cadre complet non seulement d'analyse-synthèse mais aussi de contrôle accessible à plusieurs niveaux (du très bas niveau agissant directement sur les signaux à des caractéristiques perceptives très haut niveau). Le travail accompli lors de ce stage formera donc certainement une base fondamentale pour la grande majorité des recherches prévues au sein de la thèse à venir.

Références

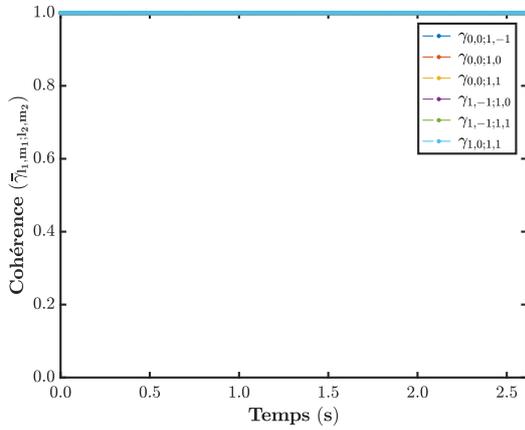
- [1] T. CARPENTIER, M. NOISTERNIG et O. WARUSFEL, « Hybrid Reverberation Processor With Perceptual Control », in *Proceedings of the 17th International Conference on Digital Audio Effects*, 2014.
- [2] M. R. SCHROEDER et K. H. KUTTRUFF, « On Frequency Response Curves in Rooms : Comparison of Experimental, Theoretical, and Monte Carlo Results for the Average Frequency Spacing between Maxima », *The Journal of the Acoustical Society of America*, 1962.
- [3] M. R. SCHROEDER, « Statistical Parameters of the Frequency Response Curves of Large Rooms », *Journal of the Audio Engineering Society*, 1987.
- [4] J.-D. POLACK, *La transmission de l'énergie sonore dans les salles*. Thèse doctorat, Université du Maine, 1988.
- [5] J.-M. JOT, L. CERVEAU et O. WARUSFEL, « Analysis and Synthesis of Room Reverberation Based on a Statistical Time-Frequency Model », in *Proceedings of the 103rd Audio Engineering Society Convention*, 1997.
- [6] D. P. JARRETT, O. THIERGART, E. A. P. HABETS et P. A. NAYLOR, « Coherence-Based Diffuseness Estimation in the Spherical Harmonic Domain », in *Proceedings of the 27th IEEE Convention of Electrical and Electronics Engineers in Israel*, 2012.
- [7] J. MERIMAA et V. PULKKI, « Spatial Impulse Response Rendering I : Analysis and Synthesis », *Journal of the Audio Engineering Society*, 2005.
- [8] N. EPAIN et C. T. JIN, « Spherical Harmonic Signal Covariance and Sound Field Diffuseness », *IEEE/ACM Transactions on Audio Speech and Language Processing*, 2016.
- [9] L. CREMER, H. A. MÜLLER et T. J. SCHULTZ, *Principles and Applications of Room Acoustics, vol. 1*. Barking, England : Applied Science Publishers, 1982.
- [10] N. XIANG, P. GOGGANS, T. JASA et P. ROBINSON, « Bayesian Characterization of Multiple-Slope Sound Energy Decays in Coupled-Volume Systems », *The Journal of the Acoustical Society of America*, 2011.
- [11] P. LUIZARD et B. F. G. KATZ, « Coupled Volume Multi-Slope Room Impulse Responses : A Quantitative Analysis Method », in *Proceedings of the Institute of Acoustics' 8th International Conference on Auditorium Acoustics*, 2011.
- [12] T. CARPENTIER, M. NOISTERNIG et O. WARUSFEL, « Twenty Years of IRCAM Spat : Looking Back, Looking Forward », in *Proceedings of the 41st International Computer Music Conference*, 2015.
- [13] W. G. GARDNER, « Efficient Convolution without Input-Output Delay », *Journal of the Audio Engineering Society*, 1995.
- [14] G. GARCÍA, « Optimal Filter Partition for Efficient Convolution with Short Input/Output Delay », in *Proceedings of the 113th Audio Engineering Society Convention*, 2002.
- [15] M. A. GERZON, « Periphony : With-Height Sound Reproduction », *Journal of the Audio Engineering Society*, 1973.
- [16] J. DANIEL, *Représentation de champs acoustiques, application à la transmission et à la reproduction de scènes sonores complexes dans un contexte multimédia*. Thèse doctorat, Université de Paris 6, 2001.

- [17] E. G. WILLIAMS, *Fourier Acoustics : Sound Radiation and Nearfield Acoustical Holography*. Cambridge, U.S.A. : Academic Press, 1999.
- [18] B. RAFAELY, « Analysis and Design of Spherical Microphone Arrays », *IEEE Transactions on Speech and Audio Processing*, 2005.
- [19] M. R. SCHROEDER, « New Method of Measuring Reverberation Time », *The Journal of the Acoustical Society of America*, 1965.
- [20] M. R. SCHROEDER, « Natural-Sounding Artificial Reverberation », *Journal of the Audio Engineering Society*, 1962.
- [21] J.-M. JOT et A. CHAIGNE, « Digital Delay Networks for Designing Artificial Reverberators », in *Proceedings of the 90th Audio Engineering Society Convention*, 1991.
- [22] N. EPAIN, T. NOOHI et C. T. JIN, « Sparse Recovery Method for Dereverberation », in *Proceedings of the 2014 REVERB Workshop*, 2014.
- [23] N. XIANG, T. JASA et Z. SU, « Schroeder Decay Decomposition for Sound-Energy Decay Analysis in Acoustic Coupled-Volume Systems », in *Proceedings of the 9th Western Pacific Acoustics Conference*, 2006.
- [24] J. S. ABEL et P. HUANG, « A Simple, Robust Measure of Reverberation Echo Density », in *Proceedings of the 121st Audio Engineering Society Convention*, 2006.
- [25] G. DEFRANCE et J.-D. POLACK, « Measuring the Mixing Time in Auditoria », in *Proceedings of Acoustics '08*, 2008.

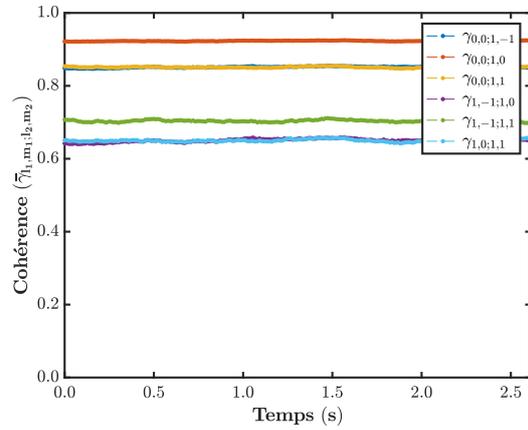
A Temps de mélange

A.1 Cohérence spatiale

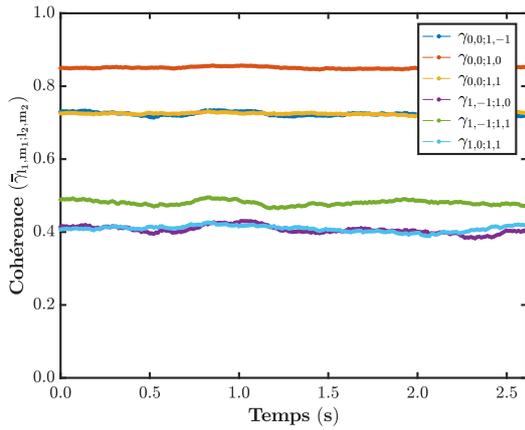
A.1.1 Signaux synthétiques stationnaires



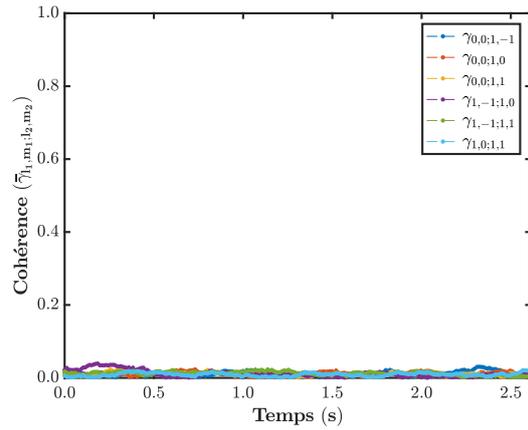
(a) Une source directionnelle, $\Omega_{\text{dir}} = (45^\circ, 45^\circ)$



(b) 4 sources centrées sur $\Omega_{\text{dir}} = (45^\circ, 45^\circ)$



(c) 25 sources centrées sur $\Omega_{\text{dir}} = (45^\circ, 45^\circ)$



(d) 49 sources également réparties sur les Ω_{dir}

FIGURE 27 – Les six cohérences inter-composantes du premier ordre, calculées sur des signaux de synthèse numérique stationnaires encodés au premier ordre, moyennées sur les fréquences

A.1.2 Comparaison des ordres d'analyse

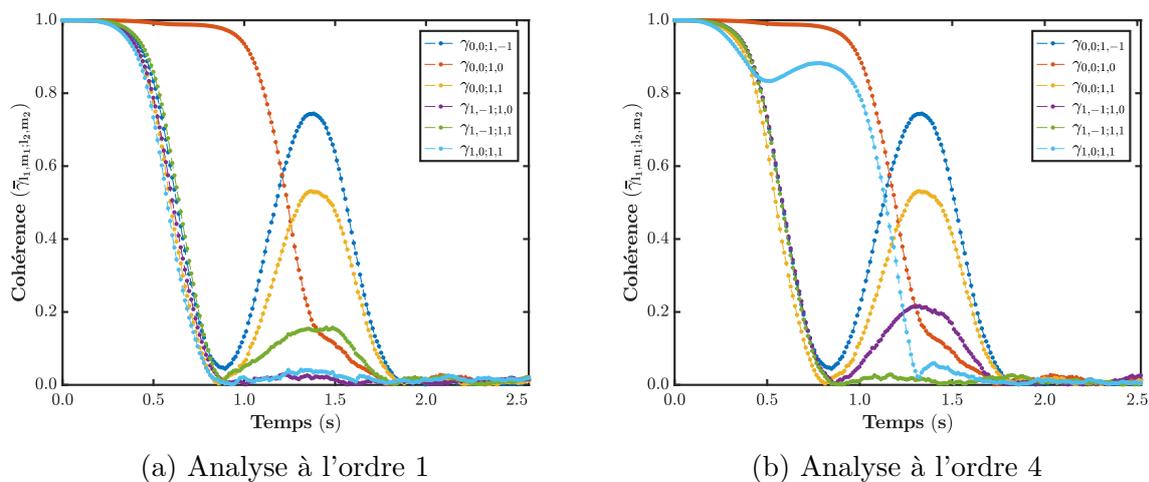


FIGURE 28 – Les six cohérences inter-composantes du premier ordre, calculées à l'ordre 1 (figure 28a, gauche) et à l'ordre 4 (figure 28b, droite) sur un signal de synthèse numérique encodé à l'ordre 4 ($\Omega_{\text{dir}} = (45^\circ, 45^\circ)$), moyennées sur les fréquences

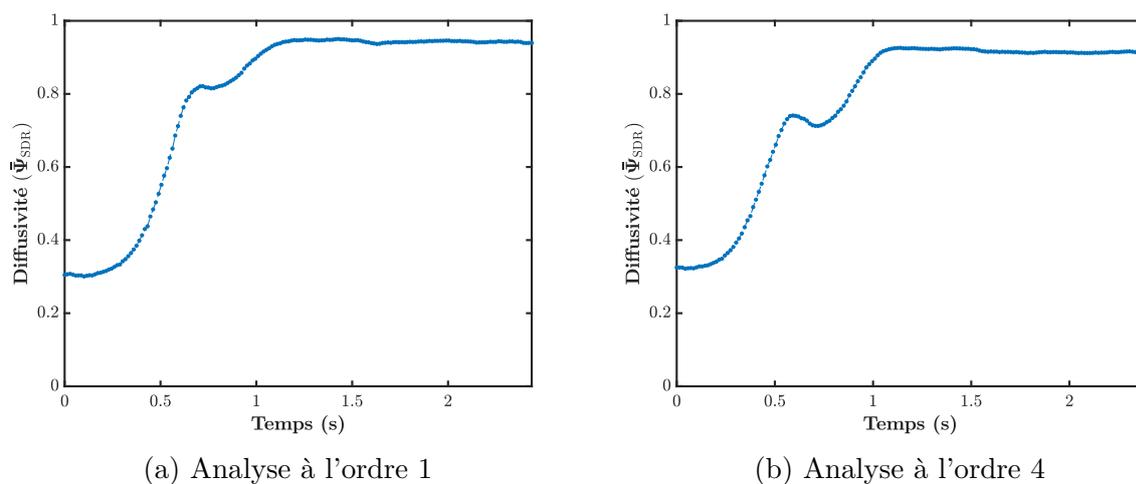


FIGURE 29 – Diffusivité par rapport signal-sur-diffus (SDR, $\bar{\Psi}_{\text{SDR}}$) calculée à l'ordre 1 (figure 29a, gauche) et à l'ordre 4 (figure 29b, droite) sur un signal de synthèse encodé à l'ordre 4 ($\Omega_{\text{dir}} = (45^\circ, 45^\circ)$) et enregistré par Eigenmike à partir de 24 haut-parleurs, moyennée sur les fréquences

A.2 Détection de direction d'arrivée

A.2.1 Cartographie de l'erreur

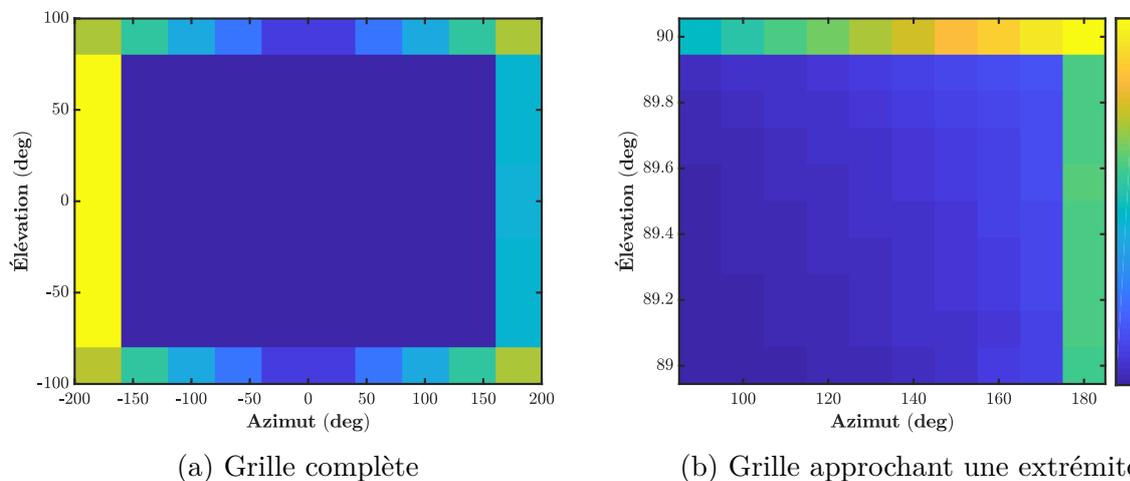


FIGURE 30 – Cartographie de l'erreur dans l'algorithme de détection tel qu'appliqué à une grille de signaux synthétiques numériques directs

A.2.2 Ajustement des valeurs autour de l'extrême azimutale

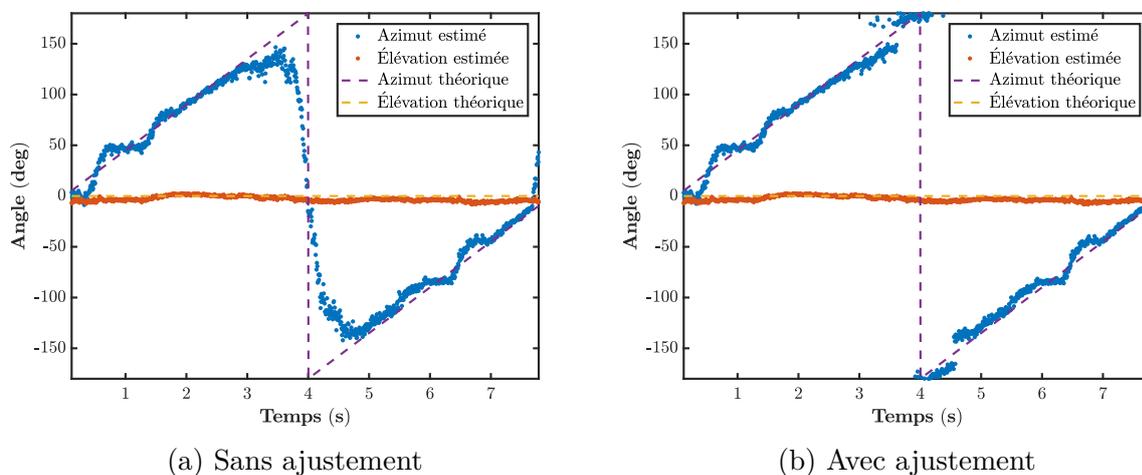


FIGURE 31 – Détection de direction d'arrivée d'une trajectoire azimutale mesurée par Eigenmike, sans (figure 31a, gauche) et avec (figure 31b, droite) l'application d'un algorithme d'ajustement des valeurs autour de l'extrémité azimutale $\pm 180^\circ$

A.3 Bases de données de réponses impulsionnelles directionnelles mesurées

A.3.1 Espace de Projection (ESPRO), IRCAM

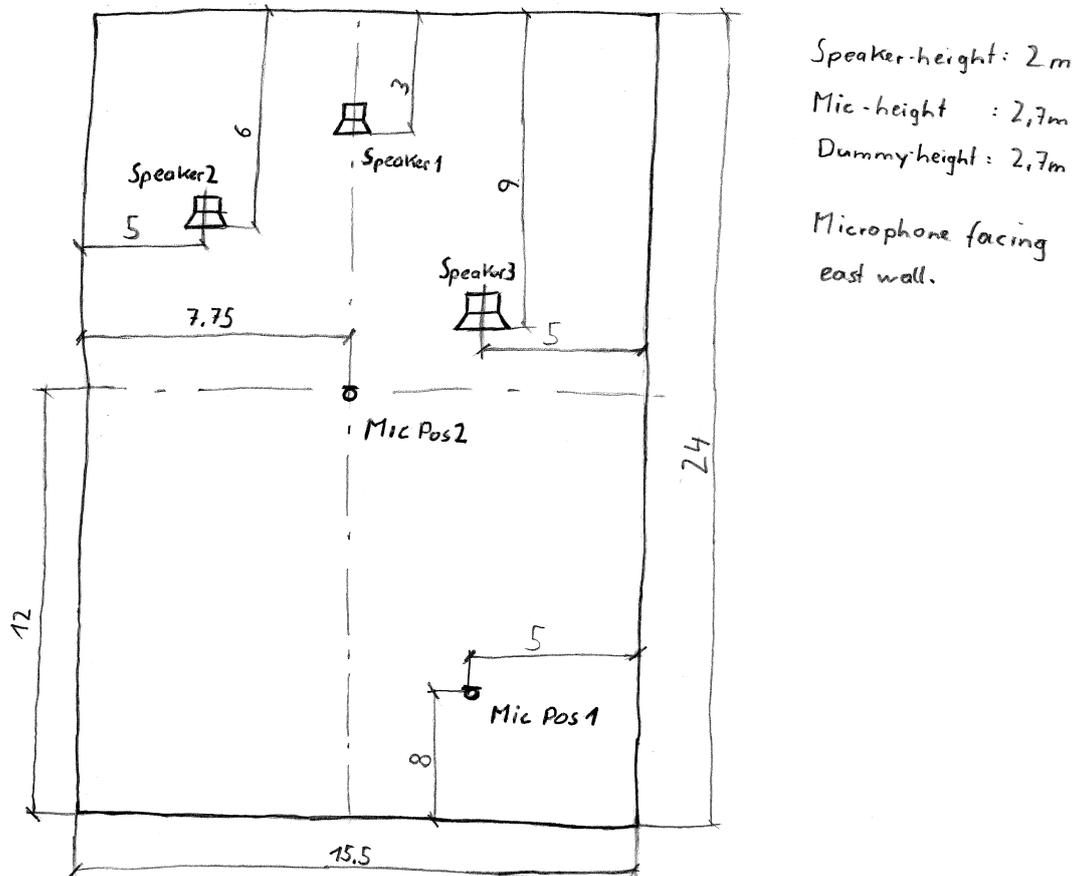


FIGURE 32 – Schéma du dispositif de mesure utilisé dans l'ESPRO pour la campagne d'enregistrements de l'équipe EAC en septembre 2013

A.3.2 Kraftzentrale, Duisburg

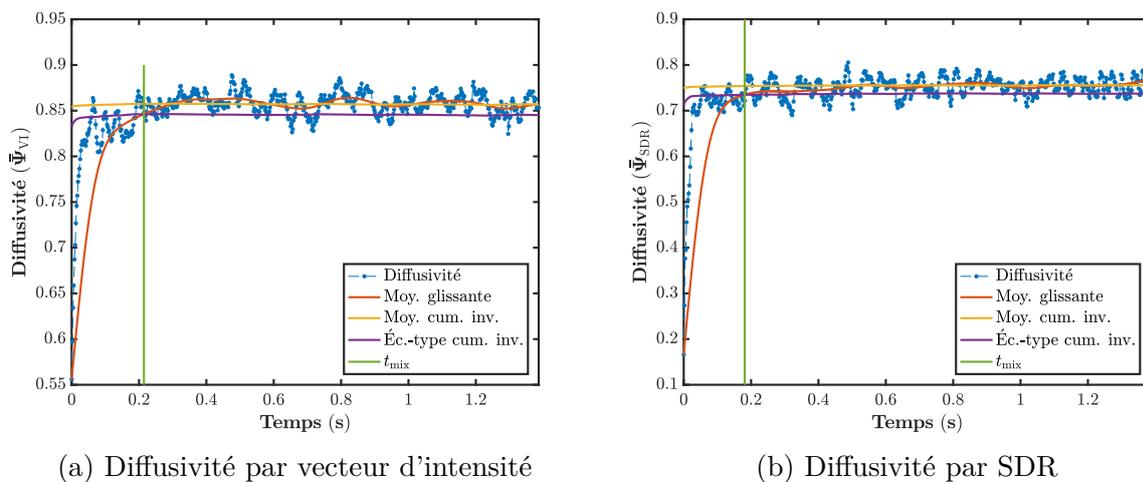


FIGURE 33 – Détermination du temps de mélange sur les deux mesures de diffusivité de la DRIR du haut-parleur n°7 de la Kraftzentrale

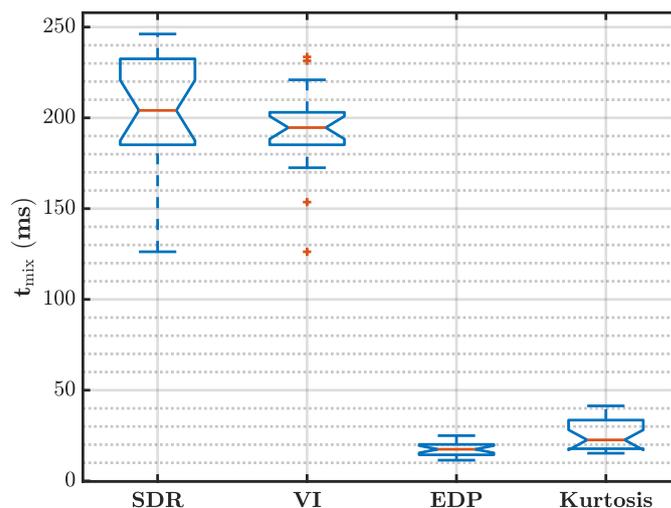


FIGURE 34 – Comparaison des différentes méthodes de caractérisation du temps de mélange telles qu'appliquées à la base de données entière des DRIR mesurées à la Kraftzentrale (20 haut-parleurs)

A.4 Hybridation

A.4.1 Comparaison des temps de réverbération

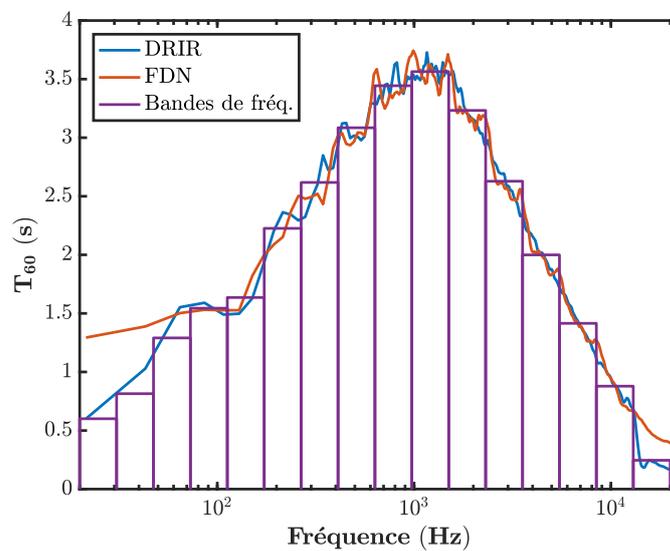


FIGURE 35 – Comparaison des temps de réverbération T_{60} de la DRIR d'origine, de sa moyenne sur les bandes de fréquences du FDN, et du FDN résultant

A.4.2 Patch Max/MSP d'évaluation à l'écoute

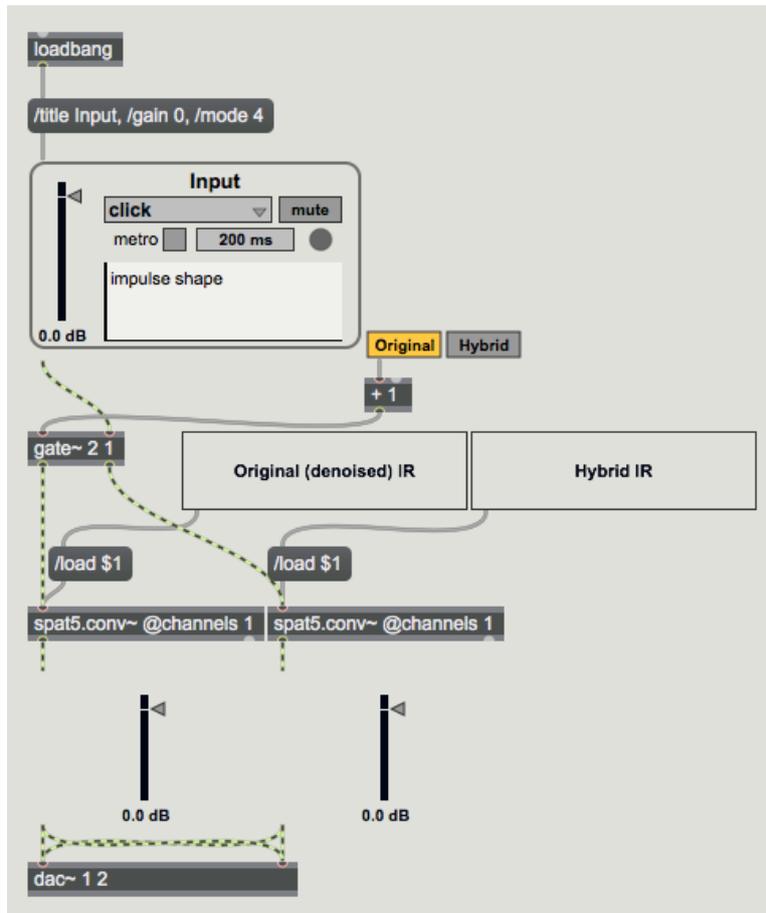
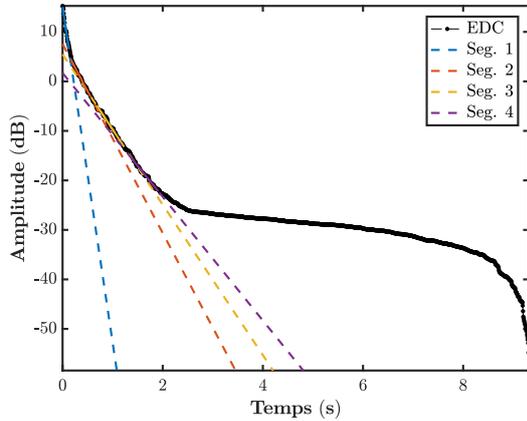


FIGURE 36 – Patch Max/MSP permettant une comparaison rapide à l'écoute de la réverbération d'origine et du résultat de l'hybridation

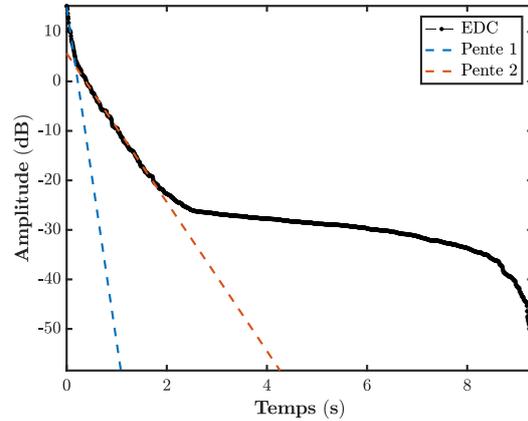
B FDN couplés

B.1 Analyse

B.1.1 Segmentation



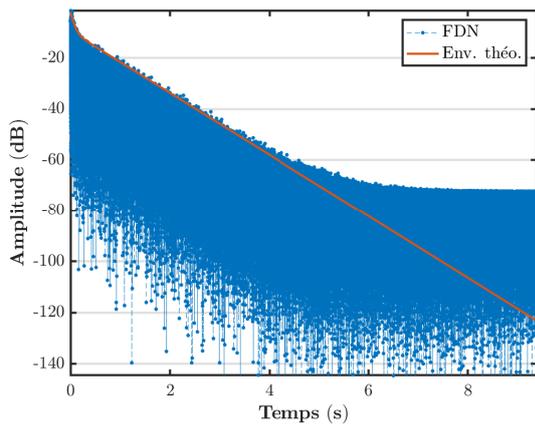
(a) Segmentation de l'EDC



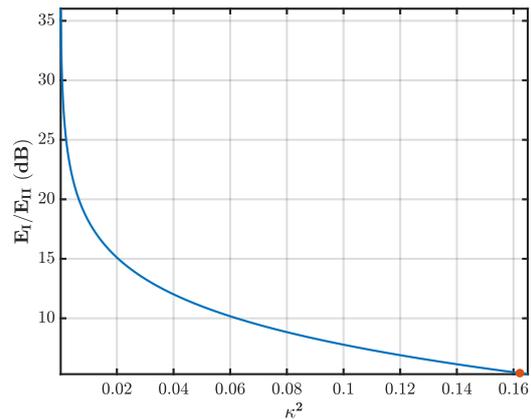
(b) Réduction à 2 pentes

FIGURE 37 – Segmentation de l'EDC par régression linéaire locale selon Luizard et Katz [11] et réduction conséquente au nombre de pentes le plus probable en fusionnant successivement les pentes les plus similaires

B.2 Synthèse



(a) Résultat de la synthèse



(b) Cartographie des paramètres

FIGURE 38 – Exemple d'une réverbération à double-pente synthétisée avec une valeur minimale du rapport E_I/E_{II} (5,4 dB)