

THESE de DOCTORAT de l'UNIVERSITE PARIS 6

Spécialité : Mécanique énergétique (UFR 923)

Acoustique, Traitement du signal et Informatique Appliqués à la Musique (ATIAM)
présentée :

par Mr Laurent CERVEAU

pour obtenir le grade de DOCTEUR de l'université PARIS 6

Sujet de la thèse : Couplage temps réel d'outils d'acoustique prévisionnelle et de dispositif d'auralisation

soutenue le 22 décembre 1999

devant le jury composé de :

- Mr Mendel KLEINER, Rapporteur
- Mr Jacques LEMORDANT, Rapporteur
- Mr Claude PUECH, Examineur
- Mr Jean-Dominique POLACK, Examineur
- Mr René CAUSSE, Examineur
- Mr Olivier WARUSFEL, Examineur

*A Jean et à Louis qui, pour des raisons différentes,
auraient sûrement aimé connaître l'existence de cette thèse*

Remerciements

J'avais, au début de la rédaction de cette thèse, accordé un soin particulier à la page de remerciements. À dire vrai, j'avais même commencé par cela, et y avais fait le tour d'un certain nombre de personnes, collègues, amis, connaissances, rencontrés lors de ces dernières années.

Maintenant que ce travail est achevé, je vais revenir à une idée plus sobre. Que toutes les personnes non remerciées ne soient pas vexées (on ne sait jamais cela peut arriver), telle n'est pas mon intention.

Je remercie donc Balthazar qui, sans ce geste spontané de ma part, ne sera jamais remercié dans aucune thèse écrite sur cette planète. Balthazar est très sympathique, gros, trouillard, tout noir tel son homonyme Roi Mage et totalement stupide. Balthazar est le chat de ma mère, et quelque part représente pour moi un certain art de vivre.

Je voudrais remercier Olivier Warusfel, qui a été mon directeur de thèse pendant ces années. Qu'il me soit permis ici de rendre hommage autant à ses qualités (à mon sens exceptionnelles) d'acousticien, qu'à ses qualités humaines. Afin d'éviter à l'avance toute tentative de censure de sa part sur ces quelques lignes (la modestie faisant partie de ses qualités), je me permettrai de citer le début des remerciements de sa propre thèse :

"Si je ne craignais de leur attribuer maladroitement les faiblesses du travail présenté dans ce mémoire, j'interviendrais volontiers les noms de l'auteur présumé de la thèse et ceux des personnes citées dans cette page de remerciements, rétablissant ainsi une paternité usurpée."

Il m'arrive de penser des choses semblables à son égard.

Enfin je ne saurais clore ce chapitre, sans remercier ma femme.

Table des matières

1	Introduction	3
2	Couplage temps-réel d'outils d'acoustique prévisionnelle et de dispositifs d'auralisation	9
2.1	L'effet de salle	10
2.1.1	La réponse impulsionnelle (RIF)	10
2.1.2	Morphologie d'une réponse impulsionnelle	14
2.2	Acoustique prévisionnelle	19
2.2.1	Méthodes employées en acoustique prévisionnelle	20
2.2.2	Auralisation	28
2.2.3	Autres contextes	31
2.3	Représentation et manipulation de l'effet de salle	33
2.3.1	Représentation bas-niveau	33
2.3.2	Manipulation de l'effet de salle	35
2.3.3	Critères d'évaluation des modalités de représentation d'une scène sonore	40
2.4	Conclusion	43
2.4.1	Synoptique d'un système de manipulation de l'effet de salle	43
2.4.2	Présentation des chapitres suivants	45
3	Description d'une scène sonore virtuelle	47
3.1	Description de l'effet de salle	48
3.1.1	Scène sonore virtuelle	48
3.2	Construction d'une scène sonore	50
3.2.1	Approche géométrique	50
3.2.2	Approche perceptive	52
3.2.3	Approche statistique : stratégie de remplacement d'une interface géométrique	57
3.2.4	Retour sur la directivité des sources	60

3.3	Examen de langages existant	63
3.3.1	Le langage VRML 97	64
3.3.2	La norme MPEG4	65
3.3.3	MPEG4 Version 1 : la partie audio 3D	66
3.3.4	Bilan des propositions pour la Version 2	67
4	Etude de différents moteurs de rendu	71
4.1	Spatialisateur de l'Ircam	72
4.1.1	Description de l'environnement MAX	72
4.1.2	Description du Spatialisateur	73
4.1.3	Scalabilité	74
4.1.4	Commandes du Spatialisateur	75
4.2	Environnemental Audio de Creative Labs	77
4.2.1	Contexte	77
4.2.2	Description de l'environnement	78
4.2.3	Interfaces de programmation	78
4.2.4	Correspondance avec une représentation bas-niveau	80
4.3	Apple SoundSprocket	80
4.3.1	Contexte	80
4.3.2	Description de l'environnement SoundSprocket	81
4.3.3	Interfaces de programmation	81
4.3.4	Conclusion	84
5	Représentation bas-niveau de l'effet de salle	85
5.1	Conversion vers une représentation morphologique	86
5.1.1	La courbe de décroissance intégrée	86
5.1.2	Estimation de l'EDC en bandes étroites	88
5.1.3	L'analyse par relief de décroissance	90
5.1.4	Approfondissement sur l'analyse de la réverbération tardive	96
5.1.5	Exemple d'analyse d'une réponse impulsionnelle	98
5.1.6	Applications pratiques	100
5.2	Implantation informatique	103
5.2.1	Librairie EDR-1.5	103
5.2.2	Applications : Réglages du Spatialisateur	104
5.3	Resynthèse arbitraire d'une reponse impulsionnelle	106
5.3.1	Contributions discrètes	107
5.3.2	Synthèse de la réverbération tardive	108
5.3.3	Le cluster	108

6	Simulation géométrique	109
6.1	Approche générale	110
6.1.1	Cadre de l'acoustique prévisionnelle	110
6.1.2	Approche paramétrique	112
6.1.3	Prise en compte des résultats de la physique	112
6.1.4	Prise en compte des résultats de la psychoacoustique	114
6.1.5	Intérêts attendus de la paramétrisation	115
6.2	Calcul du son direct	115
6.2.1	Calcul théorique	115
6.2.2	Absorption du milieu	116
6.2.3	Occlusion	117
6.2.4	Implantation	117
6.3	Calcul des contributions discrètes	117
6.3.1	Méthodes utilisées en acoustique prévisionnelle	118
6.3.2	Formulation mathématique	122
6.3.3	Comparaison des méthodes	124
6.4	Calcul des contributions diffuses	127
6.4.1	Fondements théoriques	128
6.4.2	Mode de propagation	132
6.5	Passage entre les deux processus	133
6.5.1	Modèle son direct-reverbération	133
6.5.2	Présence de premières réflexions et du cluster	134
7	Architecture d'un système de contrôle de l'effet de salle	137
7.1	Présentation générale	138
7.1.1	Caractéristiques	138
7.1.2	Organisation	139
7.1.3	Prise en compte du contexte	140
7.2	La manipulation du canal acoustique	141
7.2.1	Objets en présence	141
7.2.2	Utilisation pratique	142
7.2.3	Exemple : traitement des données d'une interface géométrique	143
7.3	Manipulation géométrique de l'effet de salle	146
7.3.1	Description des modules de commande géométrique	146
7.3.2	Stockage des informations	147
7.3.3	Construction des modules de commande géométrique	147
8	Conclusion	151

Chapitre 1

Introduction

L'évolution des capacités de traitement informatique a permis l'accès de techniques, jusque-là réservées à des experts, à un public plus large. Par public, on entend non pas l'utilisateur final d'un produit, mais les personnes contribuant à son élaboration. A titre d'exemple, les domaines de la réalité virtuelle et les jeux informatiques permettent la rencontre de différentes professions. Une des composantes d'un univers virtuel, le graphisme et notamment le graphisme en 3D, représente bien ce point de vue. Les intervenants sont le graphiste, l'informaticien, l'animateur, le designer. Un problème qui se pose alors est celui de l'établissement d'un langage commun permettant à ces différents corps de métier de travailler ensemble, sans équivoque sur la nature des informations communiquées. La définition d'un vocabulaire en est souvent la première étape : dans le cas du graphisme 3D, la mise au point de celui-ci est grandement facilitée par la nature même du matériau, c'est-à-dire sa composante visuelle ; il est en effet aisé de désigner "du doigt" des objets et de leur attribuer un nom.

En ce qui concerne le son 3D et la simulation d'un environnement audio des problématiques semblables apparaissent. Par environnement audio, il sera entendu un univers virtuel en trois dimensions et possédant des propriétés acoustiques spécifiques. La nature et le contenu des messages audio diffusés ne fera pas ici l'objet d'une étude. Contrairement au cas du graphisme, une difficulté supplémentaire est celle de la désignation et de la caractérisation des composantes d'un tel environnement. La mise en place d'un vocabulaire commun vient alors de la confrontation entre les différents intervenants. L'arrangeur¹ est souvent proche du milieu musical et utilise un vocabulaire imagé pour décrire des propriétés acoustiques (son chaud, réverbération métallique...); l'acousticien a une vision plus proche de la configuration physique du système ayant engendré cet univers et des phénomènes de propagation des ondes ; le psychoacousticien s'intéresse à la manière dont l'oreille perçoit les informations qui lui sont transmises ; l'informaticien et le traiteur de signaux apporte une vue sur la vitesse d'exécution des algorithmes utilisés et sur l'architecture du système global. Si de plus, des intervenants externes sont présents, comme par exemple un graphiste travaillant sur le même projet, un point de vue totalement étranger doit être pris en compte. La clarté et la justesse de l'expression de ce point de vue avec le vocabulaire du langage proposé pourra servir de test à la validation de ce dernier.

Cette phase d'élaboration de la composante audio doit, de plus, obéir à différentes contraintes. Tout d'abord, un environnement virtuel possède généralement plusieurs composantes, la plus importante, en terme de coûts de calcul, étant la composante graphique. La part de puissance allouée au rendu sonore ne doit en général pas dépasser 10% de la puissance totale disponible. Le choix des algorithmes utilisés pour le rendu de cet univers doit donc considérer ces contraintes. Certaines situations réclament éga-

1. En anglais, le terme d'usage courant est "Sound designer".

lement qu'il n'y ait pas d'incohérences perceptives entre la partie graphique et la partie sonore, tout en permettant la réalisation d'effets totalement fictifs et cependant plausibles. Enfin il convient de séparer les phases d'utilisation de cet univers, de celle de son élaboration. Si la phase d'utilisation réclame une exécution en temps réel d'algorithmes spécifiques, et ce à partir de données stockées sur l'ordinateur ou précalculées lors d'une phase d'initialisation, la phase d'élaboration fait intervenir un processus supplémentaire d'essais successifs permettant d'aboutir à la réalisation de l'idée originale en un minimum de temps. A la notion de temps d'exécution s'ajoute celle d'un système contrôlable par un ensemble de paramètres intuitivement manipulables. La mise au point d'un vocabulaire commun doit alors s'établir en tenant compte des contraintes informatiques tout en offrant à l'arrangeur une compréhension et une souplesse dans l'utilisation des fonctions disponibles.

L'objet "élémentaire" autour duquel s'organise cette démarche est l'effet de salle. Celui-ci comporte aussi bien des informations sur la localisation du son, que sur la transformation qu'il subit lors de sa propagation dans un espace. Il peut être l'objet de plusieurs opérations : l'enregistrement, la modélisation, la modification éventuelle et la simulation². L'application d'une ou plusieurs de ces techniques touche des domaines divers : jeux vidéos, téléconférence, applications musicales, réalité virtuelle, auralisation. Le principe de la démarche reste le même : un signal dénué d'information d'espace est traité afin de lui rajouter un effet de salle. Cette opération globale peut être appelée "spatialisation du son". Si, dans un premier temps, on recherche simplement à appliquer un effet de salle à un signal initial, un deuxième souci peut être celui de présenter à un utilisateur un ensemble de contrôles permettant de régler un dispositif de simulation. Différentes approches sont possibles pour la réalisation d'un environnement de simulation audio.

- Historiquement la première démarche est fondée sur les notions d'enregistrement et restitution. La mesure d'un effet de salle est réalisée et sa reproduction effectuée en utilisant des techniques à base de convolution. Un ensemble de configurations peut être stocké dans une base de données de signaux, photographies exactes de la situation enregistrée. Aucune limitation n'intervient sur la nature des situations décrites et stockées. Il est de plus possible d'effectuer un ensemble de transformations arbitraires sur les éléments de cette base : par exemple réalisation de morphing entre plusieurs éléments, ou création de signaux hybrides. Les résultats de ces opérations n'ont alors généralement aucune validité au regard de la notion d'enregistrement d'une situation réelle, et ne peuvent être considérés comme décrivant des situations intermédiaires du point de vue physique. Les limites d'une telle approche apparaissent alors : l'utilisateur a le choix entre un ensemble de confi-

2. On peut à ce titre effectuer une analogie avec les notions d'analyse/transformation/synthèse.

gurations figées qui ne s'appuient sur aucun modèle. Il lui est alors impossible de prédire le résultat d'une interpolation entre les éléments stockés et d'organiser celle-ci de manière modulaire.

- Une approche, qualifiée de géométrique, provient du domaine de l'acoustique architecturale ; elle décrit une situation par la configuration physique du système l'ayant engendrée et un ensemble de lois régissant l'évolution des phénomènes qui s'y déroulent. Elle semble de prime abord complète, et possède une légitimité par rapport aux champs de recherche de l'acoustique des salles, principalement l'acoustique prévisionnelle. Ceci permet donc *a priori* la réutilisation immédiate des algorithmes utilisés dans ce domaine. Un examen plus approfondi de cette approche révèle cependant des lacunes. Si en effet certains paramètres issus de la géométrie sont facilement manipulables (distance, volume), d'autres le sont moins, comme par exemple les coefficients caractéristiques des propriétés acoustiques des matériaux (absorption, diffusion). Les résultats auditifs d'une modification de leurs valeurs sont loin d'être intuitivement prévisibles. Quant à la réutilisation d'algorithmes provenant de l'acoustique des salles, elle se heurte principalement à deux problèmes : la vitesse d'exécution de ces algorithmes, qui n'ont généralement pas été conçus pour une utilisation en temps réel, et le fait que les données obtenues en sortie de ces algorithmes n'apportent pas toujours une information directement pertinente en terme de résultat auditif.

- Un autre approche s'intéresse à une description paramétrique d'un effet de salle par un ensemble d'indices perceptifs, eux-mêmes calculés à partir de données objectives. L'environnement virtuel est décrit par la donnée d'un ou plusieurs jeux de paramètres, qui constituent un "modèle" de la perception de l'espace. Par nature, cette approche permet une compréhension intuitive des phénomènes ayant lieu dans l'univers virtuel, et ce après apprentissage auditif de l'interface de contrôle. Cependant, les études sur la perception, qui sont à la base d'un tel modèle, possèdent souvent un cadre limité quant aux types d'environnements et salles considérés, et certaines configurations réelles ne peuvent être ainsi modélisées. En effet, ces études sont souvent basées sur un ensemble de tests psychoacoustiques construits avec des stimuli balisant de manière partielle l'espace du possible. Dans le domaine de l'acoustique, elles ont principalement eu lieu pour des salles de concert, auditoriums et opéras. Parmi les espaces étrangers à cette modélisation, on peut citer les espaces de petite taille, les espaces ouverts et certaines configurations particulières. Ainsi, le cas d'un environnement type "théâtre antique" : la présence de gradins régulièrement espacés se traduira par l'apparition d'une série résonances, dont l'organisation fréquentielle est liée à l'espacement entre chaque gradin, et

dont les effets n'entrent généralement pas dans le cadre des études perceptives consacrées à la caractérisation acoustique des salles. Enfin, ce genre d'approche étant récente, peu usitée, et nécessitant un apprentissage particulier, elle peut avoir des difficultés à coexister avec une approche géométrique.

L'objet de cette thèse est d'examiner ces différentes approches concernant la simulation d'un environnement virtuel audio, et ce dans le contexte d'une exécution en temps réel. Le but en est multiple :

- dégager des critères d'appréciation des différents modèles de l'effet de salle et évaluer les différentes approches en regard de ces critères,
- examiner l'influence et la duplication du choix d'un modèle sur le langage de description, le type d'interface utilisé, et le moteur de rendu sonore utilisé,
- considérer les possibilités d'échange entre les différents modèles afin que différentes interfaces puissent coexister et piloter en cohérence une scène sonore.

Le premier chapitre reviendra sur le titre de ce mémoire et permettra de délimiter les champs étudiés. Il comprend une description de l'effet de salle selon différents points de vue, et un état de l'art quant à ses contextes d'utilisation actuels. On explicitera notamment le choix qui a été fait de recourir à une description morphologique de la réponse impulsionnelle d'une salle. Celle-ci permet d'offrir un cadre général auquel peuvent se rattacher différents types d'interface ainsi que de moteurs de rendu. On exposera ses qualités principales, ainsi que ses limitations tant structurelles, que factuelles. Il sera alors présenté un schéma général rassemblant différents cas de traitement de l'effet de salle.

Le chapitre deux sera consacré aux problématiques de l'établissement d'un langage de description de l'effet de salle. On proposera des critères d'évaluation d'un tel langage et procédera à la description de deux démarches différentes pouvant permettre son établissement. Un examen de langages existant sera, de plus, effectué en regard des critères d'évaluation établis. Le chapitre trois emprunte une démarche analogue afin d'évaluer les possibilités de commande de différents moteurs de rendu audio. Cet examen sera effectué sur un moteur développé à l'IRCAM, appelé Spatialisateur, ainsi que sur des moteurs commerciaux.

Ayant vu au chapitre 1 l'importance d'une représentation morphologique de l'effet de salle, le chapitre quatre présentera ses liens avec une réponse impulsionnelle complète et des méthodes d'analyse permettant d'en extraire les paramètres. Le chapitre cinq s'intéressera à une manipulation de l'effet de salle par une interface géométrique et proposera une approche permettant d'effectuer, en temps réel, une commande de cette représentation morphologique. Cette approche s'appuie autant sur les résultats de l'acoustique prévisionnelle que sur ceux de la psychoacoustique.

Enfin la partie six proposera la structure d'un environnement informatique complet de manipulation d'un environnement audio virtuel qui vise l'indépendance et de l'interface utilisée et du moteur de rendu.

Chapitre 2

Couplage temps-réel d'outils d'acoustique prévisionnelle et de dispositifs d'auralisation

Afin d'établir de façon précise le cadre de cette étude, ce chapitre propose d'en définir et d'en décrire l'objet principal, c'est-à-dire l'effet de salle. Puis, il sera présenté un état de l'art des méthodes et applications relatives à son traitement (reconstruction et restitution), et ce dans des domaines proches du sujet, c'est à dire l'acoustique prévisionnelle et l'auralisation. Enfin, la prise en compte de la composante principale de cette thèse, c'est à dire la manipulation en temps réel d'un environnement de simulation, permettra de dégager un plan de travail, qui sera développé dans les chapitres suivants.

2.1 L'effet de salle

2.1.1 La réponse impulsionnelle (RIF)

Définition

La formalisation des phénomènes acoustiques fait intervenir les lois de propagation des ondes dans un milieu, ainsi que leurs interactions avec différents obstacles qui donnent lieu à des phénomènes de réflexion, diffusion et diffraction. Une hypothèse couramment utilisée lors de cette formalisation est celle de la linéarité des phénomènes étudiés. De plus, les variations de certains paramètres en fonction des conditions de température et d'hygrométrie sont considérées comme lentes par rapport à celle des phénomènes acoustiques.

Compte tenu de ces hypothèses, il est possible de caractériser la transformation subie par un signal émis, lors de sa propagation dans une salle, par la réponse impulsionnelle $h(t)$ du système formé par une source, en un endroit donné, la salle, et un récepteur également dans une position donnée. Cette transformation décrit un "canal acoustique" (*figure 2.1*).

Mathématiquement, la relation entre le signal, $x(t)$, émis en un point source, à celui, $y(t)$, capté en un point récepteur, est une opération de convolution, soit :

$$y(t) = x * h(t) \tag{2.1}$$

qui peut s'écrire, pour $h(t)$ causale :

$$y(t) = \int_0^{+\infty} h(\tau)x(t - \tau)d\tau, \text{ en temps continu} \tag{2.2}$$

ou

$$y(n) = \sum_{p=0}^{+\infty} h(p)x(n - p), \text{ en temps discret} \tag{2.3}$$

Cette réponse apparaît comme un intermédiaire privilégié dans la représentation et la manipulation d'un effet de salle par un utilisateur (acousticien, traiteur de signal,

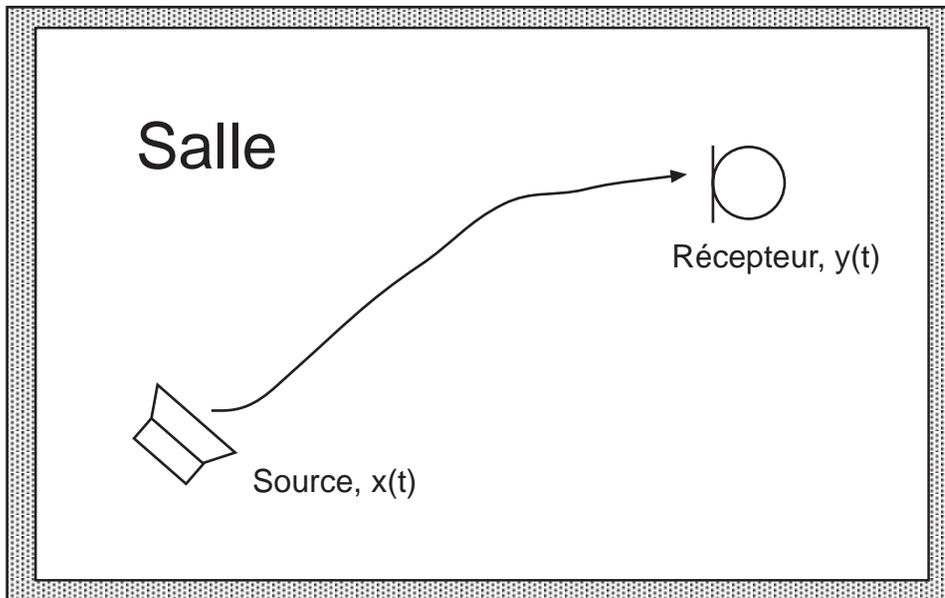


FIG. 2.1 – *Canal acoustique*

psychoacousticien, informaticien...) et ce, dans la mesure où elle contient toutes les informations sur le canal acoustique considéré.

On remarque qu'à un canal acoustique est associé un unique triplet $\{source, salle, récepteur\}$. Une salle entière ne sera donc complètement représentée que par un ensemble, *a priori* infini, de réponses, correspondant à autant de positions et d'orientations pour la source et le récepteur. Dans la pratique, un quadrillage de la salle est effectué, pour la source et le récepteur, permettant ainsi d'obtenir une distribution de réponses impulsionnelles¹ Etant donné que la mesure d'une réponse impulsionnelle nécessite l'emploi de transducteurs, on notera qu'un même canal peut être caractérisé par un ensemble de réponses impulsionnelles, correspondant aux orientations et caractéristiques de directivité de ceux-ci.

Obtention d'une réponse impulsionnelle

L'acquisition d'une ou plusieurs réponses impulsionnelles a principalement lieu selon deux modes.

1. Le choix d'un pas pour ce quadrillage reste un problème délicat. Il peut ou non tenir compte des particularités de chaque salle (géométrie, propriétés de mélange), et faire intervenir des considérations sur les seuils différentiels du système auditif. Enfin, la nécessité de considérer des ensembles de positions et d'orientations pour la source et le récepteur requiert une somme importante de réponses pour caractériser une salle de manière exhaustive.

Mesure

Il s'agit d'une situation d'enregistrement, faisant intervenir une salle réelle, ou éventuellement une maquette². Différentes méthodes sont utilisées [Sch79a], [Fos86], [JGS83], [DH89] :

- *mécaniquement* : par tir au pistolet, ou éclatement d'un ballon rempli d'air. Le dispositif mécanique est placé au point source, et émet un signal assimilable à une impulsion. Un dispositif d'enregistrement, au point récepteur, capte le signal transformé, soit la réponse impulsionnelle.
- *numériquement* : un haut-parleur émet un signal engendré par un ordinateur, qui est ensuite capté par un microphone. Ce signal est choisi de manière à ce que sa donnée, ainsi que celle de l'enregistrement effectué au point récepteur, permettent de retrouver la réponse impulsionnelle, après traitement informatique. Parmi les méthodes et signaux couramment utilisés, on peut citer : le générateur d'impulsion, les séquences pseudo-aléatoires (séquences de longueur maximale), les codes de Golay, et le balayage de fréquences.

Le choix préférentiel d'une de ces méthodes par rapport à une autre dépend des exigences requises sur le rapport signal à bruit, ainsi que des contraintes de reproductibilité sur une mesure, voire de robustesse par rapport à certaines conditions pratiques [No99],[POL88b],[Van94].

La réponse obtenue contient des informations sur l'ensemble du canal acoustique et intègre par conséquent simultanément les contributions des transducteurs en plus de celles de la salle que l'on cherche à mesurer. On verra, lors d'une description plus approfondie de la réponse impulsionnelle, quelle est la nature exacte de cette influence.

Prédiction

Une autre méthode pour obtenir une réponse impulsionnelle consiste à utiliser un environnement informatique afin de la simuler ; cette technique est principalement adaptée aux phases de conception d'un espace. Les caractéristiques physiques (géométrie et matériaux) d'une salle, ainsi que celles des sources et récepteurs sont saisies dans un logiciel, puis un ensemble d'algorithmes simulant les phénomènes de propagation acoustique est exécuté. La fidélité du résultat par rapport à une situation réelle est liée au niveau de finesse de ces algorithmes dans la prise en compte des phénomènes présents, ainsi qu'au niveau de détail utilisé pour décrire la géométrie du lieu.

2. Dans ce cas, la bande de fréquences utilisée ainsi que le milieu de propagation sont modifiés de manière à s'adapter à l'échelle de la maquette. Cette situation est *a priori* réservée à la conception d'espaces architecturaux. Moyennant certaines précautions et post-traitements, il est possible d'utiliser ces réponses pour effectuer une auralisation de la maquette. Cette méthode tient lieu également de simulation.

Exploitation d'une réponse impulsionnelle

Caractérisation

Cette démarche concerne principalement l'acousticien dans le cadre de projets architecturaux. Afin d'évaluer la qualité acoustique d'un canal (ou d'un ensemble de canaux), un ensemble d'indices représentatifs est calculé à partir d'une réponse impulsionnelle. En effet, si celle-ci contient par nature toutes les informations nécessaires à cette caractérisation, la donnée directe de tous ses échantillons représente un ensemble vaste et difficile à interpréter. Une représentation par un ensemble d'indices s'avère alors plus souple et manipulable.

Le nombre et la définition de ces indices a été l'objet de nombreuses études dans le domaine de l'acoustique des salles ; on peut considérer les travaux de Sabine, à la fin du siècle dernier, et ceux de Beranek en 1962, comme point de départ de ce type de recherche [Ber62]. Une des difficultés de cette démarche réside dans l'adoption d'un double point de vue, considérant à la fois l'aspect perceptif de la qualité acoustique d'une salle, conjointement à sa caractérisation à partir de données objectives (énergies, durées), mesurables sur une réponse impulsionnelle. Deux problèmes sont à résoudre : la recherche de grandeurs traduisant au mieux la qualité perceptive d'un événement sonore et exprimables en fonction de données objectives, et la détermination du nombre minimum d'indices permettant de caractériser l'effet de salle de manière pertinente.

Sans présenter l'ensemble des indices (perceptifs et/ou objectifs) issus de ces recherches, on peut néanmoins en donner une première classification. On peut ainsi distinguer :

- les indices monoraux, énergétiques ou temporels, liés à la distribution temporelle de l'énergie dans une réponse,
- les indices multicanaux, mesurés sur un ensemble de réponses³, permettant de traduire la répartition spatiale de l'effet de salle.

Auralisation

Compte tenu de la relation de convolution gouvernant un canal acoustique, il est alors possible d'appliquer celle-ci à un signal quelconque, pour en effectuer une simulation auditive. Une telle démarche a dépassé le cadre de l'acoustique architecturale, où elle a généralement lieu après prédiction d'une réponse. Dans le domaine de la prise de son, il est ainsi possible de rajouter *a posteriori* un effet de salle à un enregistrement, afin de travailler son esthétique. Les domaines de la réalité virtuelle (multimédia, télécommunication, jeux) utilisent également ces techniques afin d'augmenter le réalisme lors du

3. Si l'on s'intéresse à la caractérisation de phénomènes se déroulant aux oreilles de l'auditeur deux réponses sont généralement utilisées ; on parle alors d'indices binauraux. Si l'on cherche plutôt à décrire le champ sonore présent autour de l'auditeur, le nombre de réponses peut être supérieur.

rendu d'un monde virtuel.

Deux composantes de l'effet de salle doivent être reproduites lors de l'auralisation d'un canal : la localisation, ou positionnement de la source par rapport au récepteur, et l'effet de salle proprement dit. La reproduction de la localisation ne peut cependant s'effectuer avec la seule donnée d'une réponse impulsionnelle monophonique, en raison de la nature du système auditif.⁴

Suivant la technique utilisée pour la diffusion spatialisée du message sonore, le signal subit un encodage spécifique, nécessitant alors un nombre de réponses impulsionnelles au minimum égal à 2. Ainsi les techniques stéréophoniques, binaurales et transaurales réclament deux signaux, les techniques Ambisonics quatre⁵ [Beg94], [Mal90], [Ger77].

2.1.2 Morphologie d'une réponse impulsionnelle

Description

L'examen d'un échogramme, évolution temporelle de l'énergie d'une réponse impulsionnelle (ETC : Energy time curve) permet d'y distinguer différentes sections temporelles (*figure 2.2*). Plusieurs angles de vue sont possibles. Si l'on considère le dispositif générateur du signal (salle, transducteurs) ainsi que les lois de propagation des ondes, il s'agit d'un point de vue "physique". Sous l'angle "signal", on observe l'enregistrement de l'évolution de la pression au point récepteur, dans ses dimensions temporelles, fréquentielle, et directionnelle (nécessitant pour cette dernière plusieurs réponses impulsionnelles) . Si l'on s'intéresse à la sensibilité du système auditif aux différentes composantes de la réponse, on examine l'angle "perceptif".

On donnera ici les principales caractéristiques, selon ces trois angles, des différentes sections d'une réponse impulsionnelle.

Le son direct, OD

Il correspond à la propagation en champ libre entre la source et le récepteur, et est représenté par les premiers pics de l'échogramme. Idéalement, en négligeant l'absorption de l'air, la présence éventuelle d'obstacles, et pour une source ponctuelle à réponse parfaite, le son direct se réduit à une impulsion de Dirac dont l'amplitude dépend de la distance entre la source et le récepteur. La prise en compte du milieu de propagation,

4. En effet, le système auditif localise une source par la donnée de deux signaux, arrivant respectivement à l'oreille gauche et à l'oreille droite. Ce sont les différences de retard et de gain en fonction de la fréquence, entre ces deux signaux (HRTF), qui permettent d'effectuer la localisation d'un son.

5. D'autres techniques existent (Pan-pot d'intensité, Surround) pour la diffusion d'un signal en trois dimensions, mais elles ne sont pas basées sur un encodage accessible par un dispositif microphoniques. Elles ne sont donc pas accessibles en situation d'enregistrement, et sont réservées au cadre de la post-production ou de la synthèse.

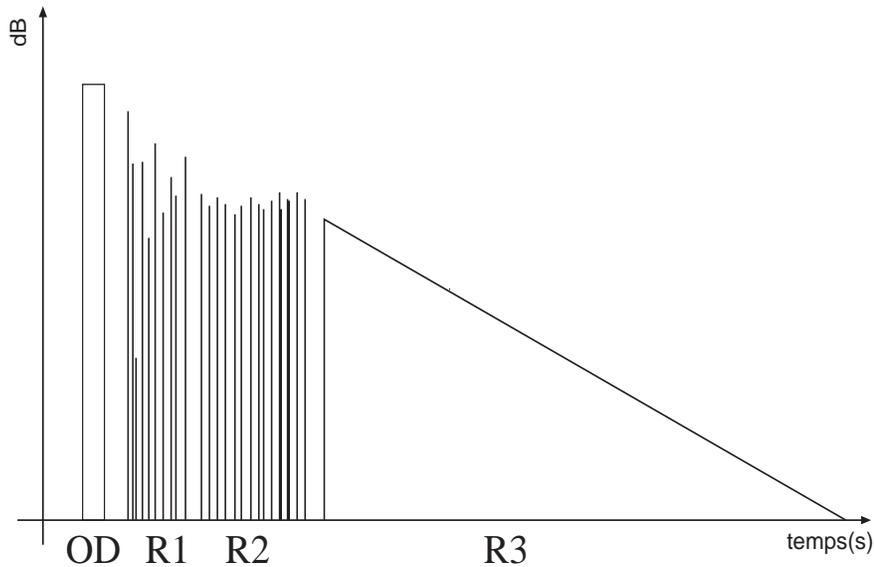


FIG. 2.2 – *Forme de l'échogramme d'une réponse impulsionnelle*

des caractéristiques de directivité des transducteurs, et éventuellement des conditions de champ proche conduit à un étalement temporel de ce pic, et à une modification de son contenu fréquentiel. L'effet de cette directivité se manifeste à la fois par une modification d'amplitude du son direct, selon l'orientation relative source-récepteur, ainsi que par des modifications sur le spectre du signal injecté dans la source. Dans ce cas il se peut que certaines informations concernant le canal acoustique ne soient pas directement lisibles sur un échogramme ; l'accès à ces informations peut éventuellement s'effectuer par une opération de déconvolution de la réponse par des signaux caractéristiques des éléments de la chaîne de mesure. En anticipant sur la suite de cette description, il faut cependant noter qu'une déconvolution de la réponse impulsionnelle n'est possible que si la source et le récepteur sont de type omnidirectionnel⁶.

D'un point de vue perceptif, le son direct contribue en particulier à la détermination du niveau sonore perçu⁷. En raison de l'effet de précedence (ou loi du premier front d'onde), sa direction de provenance détermine la localisation de la source.

La réverbération tardive, R3

Elle correspond à la partie tardive de la réponse impulsionnelle. Physiquement le

6. Par exemple, si le microphone placé au point récepteur est de type figure 8, l'onde directe sera annulée pour une source placée dans le plan médian du microphone. De même, si le spectre émis par la source diffère selon les directions, chaque réflexion devrait être déconvolué par une fonction de transfert différente

7. De manière simplifiée le niveau perçu peut être séparé en deux composantes : la "présence de la source" et la "présence de la salle". L'onde directe contribue à la "présence de la source".

champ tardif tend vers les conditions de champ diffus, c'est-à-dire un champ où, en tout point de la salle, les ondes arrivent avec des directions équiprobables et une même intensité. Dans ces conditions Sabine montre que le flux d'énergie traversant les parois de la salle est proportionnel à l'énergie résidant dans la salle. Cette propriété se traduit, en régime libre, par une décroissance exponentielle caractérisable par le temps de réverbération. En pratique ces conditions de champ diffus ne sont qu'approchées.

Sur le plan du signal, les propriétés statistiques du signal réverbéré ont été étudiées notamment par Schroeder et Polack. Compte tenu de la densification progressive des modes de résonance de la salle en fonction du carré de la fréquence, on peut considérer qu'en haute fréquence la réponse fréquentielle observée par un microphone dans la salle résulte de la superposition des contributions d'un grand nombre de modes. L'amplitude et la phase de chaque mode varient en fonction de la position du micro, et Schroeder montre que parties réelle et imaginaire suivent une loi de Gauss de même variance en fonction de la position. De manière duale, pour une position de micro donnée et une excitation stationnaire, la réponse fréquentielle observée par le récepteur répond à une loi identique en fonction de la fréquence. Cette propriété est vérifiée au delà d'une fréquence limite appelée fréquence de Schroeder (*chapitre 5*).

D'un point de vue temporel, au sein de la réverbération tardive, le microphone subit, à chaque instant les contributions simultanées de plusieurs réflexions conduisant également à un comportement gaussien de la réponse temporelle (plus exactement la réponse est alors le produit d'un bruit gaussien et d'une enveloppe de décroissance exponentielle. Polack établit un lien entre cette propriété temporelle et les capacités de "mélange" spatial de la salle. Selon la géométrie de la salle et le caractère diffusant des parois, la distribution des vecteurs d'intensité s'uniformise plus ou moins rapidement et influe sur le temps au delà duquel on peut considérer les conditions d'établissement de ces propriétés statistiques.

En première approximation, le niveau du champ réverbéré est identique en tout point de la salle, et son énergie décroît exponentiellement au cours du temps (ou linéairement lors d'une expression logarithmique). Cette réverbération est caractérisée par un niveau global d'énergie, $R3$, ainsi qu'un paramètre de décroissance temporelle, T_r , ces deux quantités variant avec la fréquence.

Il sera montré au chapitre 5 que les variations fréquentielles du temps de réverbération dépendent uniquement des caractéristiques de la salle, tandis que celles de l'énergie $R3$, sont reliées également aux transducteurs, et plus particulièrement à leurs fonctions de transfert en champ diffus, ce qui traduit une autre influence de la directivité des transducteurs sur la mesure d'une réponse impulsionnelle.

La réverbération est perçue comme un continuum sonore, entourant l'auditeur, et possédant une coloration fréquentielle spécifique ; en outre, ce continuum sera d'autant plus

audible que les événements qui précèdent ce phénomène de réverbération ont une contribution énergétique faible. Si l'on adopte une modélisation simple de l'effet de salle (son direct+réverbération), la perception de ces deux composantes, est reliée à leur niveau relatif (appelé parfois *Dir/Rev*).

Lors de la mesure d'une réponse impulsionnelle *in situ*, la fin de la réverbération est généralement noyée dans du bruit. Celui-ci peut avoir une influence néfaste lors d'une utilisation ultérieure de la réponse impulsionnelle. Par exemple, lors d'une auralisation par convolution, une augmentation du temps de réverbération perçu est remarquée ; des biais lors de l'estimation d'indices objectifs et perceptifs peuvent également apparaître.

Les premières réflexions, $R1$

Il s'agit des contributions correspondant aux premières réflexions sur les parois de la salle, subissant ainsi une altération de leur amplitude et de leur contenu fréquentiel.

Ces échos sont caractérisés par leur amplitude, leur spectre, leur instant d'arrivée, et leur direction (non lisible sur une réponse monophonique).

Les effets perceptifs de ces réflexions sont multiples ; principalement, elles contribuent à une augmentation de la puissance subjective du son direct ("présence de la source"). Elles agissent également sur les sensations de largeur apparente de la source et d'impression d'espace. Chacune de ces deux notions amène à considérer des pondérations spatiales et temporelles de ces réflexions. Si elles arrivent de manière précoce et dans un angle solide voisin de la source, elles contribuent à augmenter la largeur apparente de la source, tandis que, plus tardives et provenant de directions latérales, elles contribuent à l'impression d'espace. Les réflexions provenant de l'arrière auront plutôt tendance à être intégrées au champ tardif [Leh75], [Bar71], [Bar74], [BM81], [Kah95].

Les réflexions tardives, $R2$

Cette section se justifie essentiellement sur le plan perceptif ; elle peut être considérée comme un intermédiaire entre les premières réflexions et la réverbération tardive. La densité temporelle des réflexions ne permet plus de les séparer les unes des autres, mais l'évolution temporelle de l'énergie n'a pas encore adopté un profil exponentiellement décroissant, caractéristique de la réverbération tardive. On peut le caractériser par un niveau d'énergie globale, $R2(f)$.

Perceptivement, son énergie est intégrée conditionnellement au son précoce et peut contribuer à la présence de la source. Conjointement, la présence d'une valeur forte de $R2$ s'accompagne d'une modification de la réverbérance perçue, surtout pendant le déroulement du discours dans la salle ("running reverberance" par opposition à la "stopped reverberance" entendue en fin de discours⁸). Une description plus détaillée de

8. Cette modification peut se formaliser par le calcul d'un temps de réverbération précoce, EdT

ces phénomènes⁹ peut être trouvée dans [Kah95].

Validation de ce modèle

Cette segmentation fait apparaître le problème des valeurs à attribuer aux limites temporelles de chacune des parties. Deux points de vue sont possibles.

D'un point de vue physique, ces frontières temporelles correspondent à des modifications sur le régime de succession d'échos dans la réponse : on sépare ainsi le son direct, les réflexions discrètes (early), et la réverbération tardive dont l'énergie décroît de manière exponentielle. Ces temps sont reliés aux dimensions de la salle, ainsi qu'à la répartition géométrique des parois, et aux propriétés acoustiques des parois. Chaque canal acoustique doit *a priori* être considéré dans sa particularité. Certaines propriétés du phénomène de réverbération permettent cependant de s'affranchir de cette description géométrique fine, pour estimer le temps d'arrivée de la réverbération, en fonction de quantités globales (volume, libre parcours moyen. . .) (*cf:2.2.1*).

Une autre approche provient des études portant sur la caractérisation objective et perceptive de l'effet de salle. Les bornes temporelles choisies permettent d'offrir une base d'indices perceptifs, indépendant les uns des autres, et expliquant au mieux la qualité de la salle et¹⁰ (on reviendra au chapitre 2 sur une description plus précise de la démarche ayant permis d'établir ces limites). Dans le moteur de rendu utilisé pour cette étude, des valeurs par défaut, résultats de recherches psychoacoustiques et expérimentales menées au laboratoire de l'IRCAM, leur sont attribuées [JGS83].

$$\left\{ \begin{array}{ll} 10 \text{ ou } 20 \text{ ms} & \text{début des premières réflexions } R1(f) \\ 40 \text{ ms} & \text{début du cluster } R2(f) \\ 80 \text{ ms ou } ; 100 \text{ ms} & \text{début de la réverbération } R3(f) \end{array} \right.$$

On remarquera que ces valeurs ne sont *a priori* valides que pour des espaces semblables à ceux considérés lors de ces recherches sur la perception de l'effet de salle, c'est-à-dire les salles de concerts, opéras, auditoriums.

(early decay time), mesuré sur le début de la courbe de décroissance intégrée (voir chapitre 3). Alors que le temps de réverbération classique est évalué sur un intervalle correspondant à une décroissance de -5 à -35 dB, l'EdT est évalué sur un intervalle correspondant à une décroissance de 0 à -15 dB. La présence d'une forte contribution de *R2* aura tendance à diminuer l'EdT.

9. Sans entrer maintenant dans les détails de son fonctionnement, on peut noter que le moteur de rendu Spatialisateur en propose une implantation simplifiée.

10. La détermination de ces valeurs se fait par expérimentation. Différents jeux de valeurs sont testés; celui retenu permet d'obtenir une base de facteurs perceptifs la plus explicative. Le fait que ces valeurs correspondent à des modifications dans la nature de la réponse impulsionnelle est une constatation *a posteriori*.

Limitations

D'autres types d'espaces sortent cependant de ce cadre d'étude. On peut ainsi citer les petits espaces et les espaces ouverts ou extérieurs (*figure 2.3*). Parmi les déviations observables, on peut pévoir :

- Non vérification des hypothèses sur l'établissement d'un champ diffus,
- Non respect des limites temporelles établies précédemment concernant le sectionnement d'une réponse impulsionnelle,
- Recouvrement de ces mêmes sections temporelles (par exemple, présence d'échos discrets de fort niveau hors des limites de $R1$),
- Absence éventuelle de certaines composantes de la réponse impulsionnelle (par exemple, absence de réverbération).

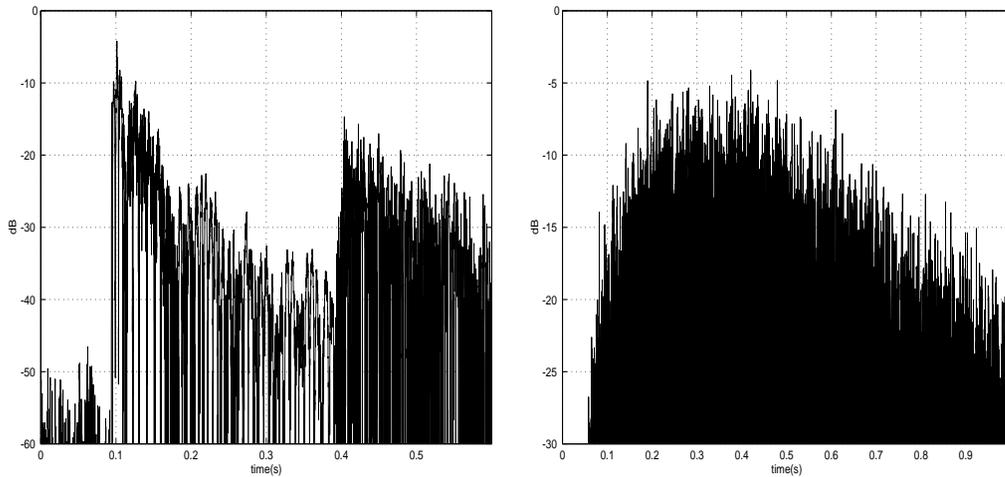


FIG. 2.3 – *Exemples de réponse impulsionnelle hors du cadre d'étude : la réponse de gauche correspond à un espace ouvert où, une réflexion arrive à 300 ms après le son direct et une première série de réflexions diffuses, et où un second régime de réverbération commence; la réponse de droite correspond à un escalier (absence de son direct, et surtout croissance lente de l'énergie du signal de 0 à 300 ms, conduisant à une valeur négative de l'EdT.)*

2.2 Acoustique prévisionnelle

Le contexte de l'acoustique des salles fait généralement intervenir deux types d'opération : la prédiction et l'auralisation. Différents algorithmes sont employés pour réaliser

ces opérations.

2.2.1 Méthodes employées en acoustique prévisionnelle

Une première manière d'obtenir une réponse impulsionnelle consiste à résoudre les équations physiques de propagation des ondes et ce de manière numérique (méthode des éléments finis, méthode des différences finies). En raison de la puissance et du temps de calcul requis par ces méthodes, elles ne seront pas considérées dans le cadre de cette étude.

On présentera ici des méthodes employées couramment dans le cadre de l'acoustique des salles, et basées sur une approche géométrique des phénomènes étudiés.

Méthode des rayons

Historiquement, elle fut la première à être utilisée. Une impulsion est émise en un point de la salle, et son énergie partagée en N rayons, véhiculant chacun une portion d'onde plane. La direction du rayon est perpendiculaire au front de l'onde véhiculée. Le parcours de ces rayons est calculé, en supposant qu'il est rectiligne, et qu'il subit l'absorption par l'air. La rencontre d'un obstacle induit une réflexion, selon un modèle spéculaire, basé sur la loi de Descartes (*figure 2.4*). L'utilisation de ce modèle suppose que les parois sont infinies et planes.

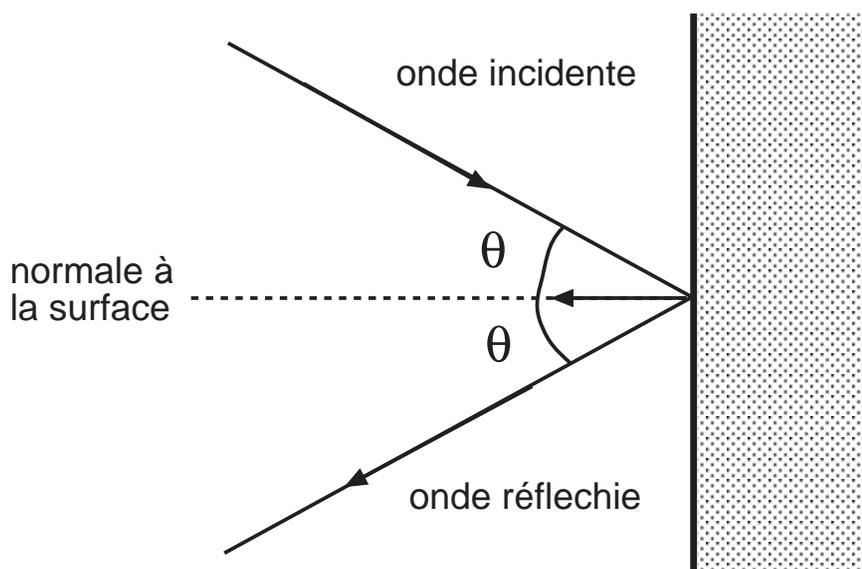


FIG. 2.4 – Réflexion d'une onde sur une paroi - loi de Descartes

L'atténuation et le déphasage subis par l'onde lors de sa réflexion sont déterminés par le facteur de réflexion $R(f, \theta)$, nombre complexe variant selon la fréquence et l'angle

d'incidence. L'énergie véhiculée par cette réflexion est proportionnelle à $(1 - \alpha)$, α étant le coefficient d'absorption de la paroi ; son lien avec le facteur de réflexion s'exprime par :

$$\alpha(f, \theta) = 1 - |R(f, \theta)|^2 \quad (2.4)$$

Bien que cette formule indique une modification de la valeur de α selon la direction d'incidence, on utilise en pratique un coefficient évalué sous hypothèse d'incidence aléatoire (α Sabine).

Le récepteur est modélisé par un secteur de forme simple et de petite dimension. Chaque fois que ce secteur est traversé par un rayon, l'énergie du rayon, ainsi que son instant de passage sont mémorisés et portés sur l'échogramme du point récepteur (*figure 2.5*). La procédure est stoppée lorsque l'énergie portée par un rayon devient trop faible, ou que la distance entre rayons voisins devient supérieure aux dimensions du récepteur.

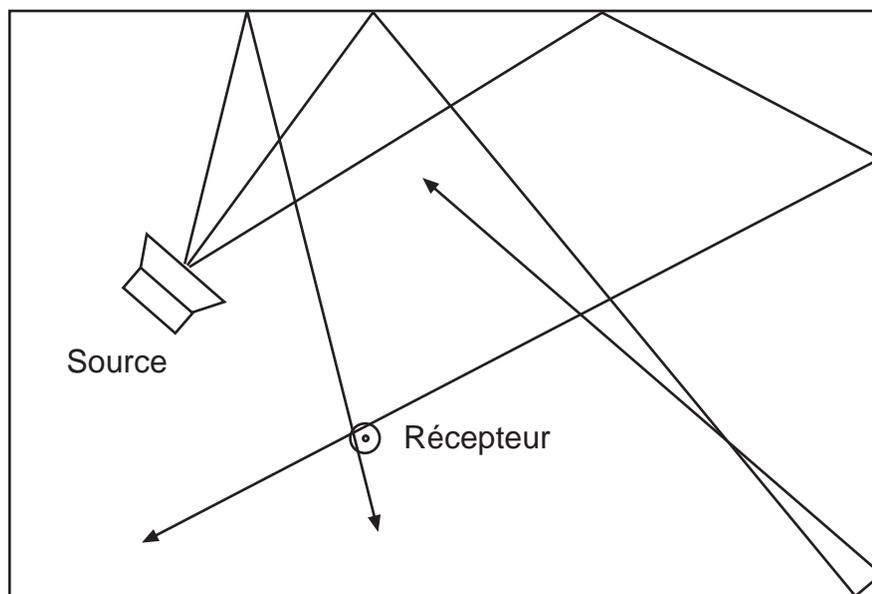


FIG. 2.5 – *Méthode des rayons*

L'inconvénient majeur de cette méthode est qu'elle ne prend généralement pas en compte la nature ondulatoire des phénomènes acoustiques et ne tient donc pas compte des interférences présentes dans la salle. De plus, le modèle de réflexion des ondes sur les parois n'est qu'une approximation de la réalité, et notamment néglige les phénomènes de diffusion. Enfin les phénomènes de diffraction ne sont pas pris en compte par l'application directe de cette méthode et doivent faire l'objet de traitements complémentaires.

Le chapitre 5 présentera de manière plus détaillée des variantes de la méthode des rayons permettant d'en pallier certaines limitations.

Méthode des sources-images

Elle est également basée sur une approche géométrique des phénomènes étudiés. Les hypothèses nécessaires à son application, et les phénomènes modélisés sont semblables à ceux d'une méthode des rayons.

On considère que chaque paroi est le plan d'une symétrie donnant naissance à une sources-images (*figure 2.6*). Ce processus est ensuite itéré à partir des sources-images construites. Une sources-images d'ordre P (obtenue par P symétries successives) correspond au trajet d'un rayon ayant subi P réflexions sur les parois de la salle.

Le signal issu d'une sources-images subit une atténuation en $1/r^2$ de l'énergie, ainsi qu'une atténuation due aux absorptions dans l'air et à la traversée de chaque paroi. La réponse présente au récepteur est obtenue en sommant les contributions de toutes les sources virtuelles, sous réserve qu'elles soient visibles par le récepteur. Dans le cas d'une salle de forme parallélépipédique, il est possible de construire un réseau de sources-images selon une succession de symétries par rapport aux parois. Du fait de la forme de la salle, ces opérations sont peu coûteuses en puissance de calcul. Le cas d'une salle de forme quelconque sera examiné au chapitre 4.

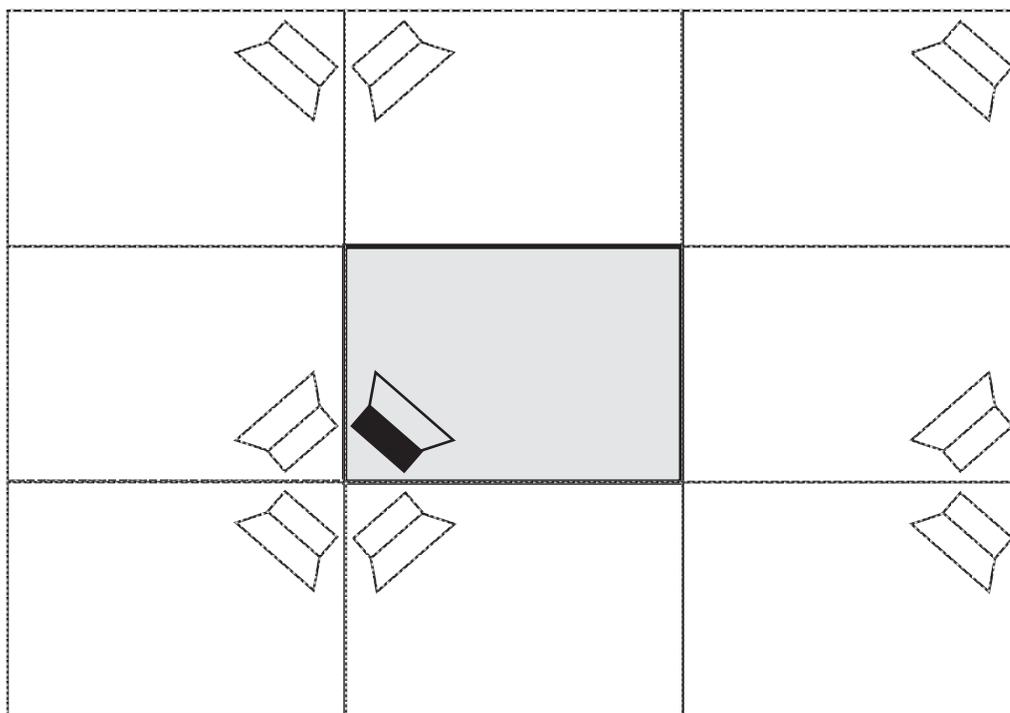


FIG. 2.6 – Méthode des sources-images (salle parallélépipédique)

Lien entre les méthodes basées sur l'hypothèse de réflexions spéculaires

Ces deux méthodes, rayons et sources-images sont par nature très proches l'une de l'autre. En effet, par définition, à une sources-images d'ordre P correspond un rayon ayant effectué P réflexions. Ce principe est du reste parfois utilisé afin de corriger certaines imprécisions pouvant survenir lors de l'utilisation d'une méthode des rayons. La principale différence entre ces deux méthodes tient surtout dans la façon dont elles considèrent le rôle du récepteur. En effet, si les deux méthodes mettent en œuvre un algorithme indépendant de la position du récepteur (tir de rayons, placement des sources-images), l'évaluation des contributions réelles apportées au récepteur prend des formes différentes :

- dans le cas d'une méthode des rayons, elle implique un test de proximité pour chaque rayon tiré et à chaque ordre de réflexion,
- dans le cas d'une méthode des sources-images, le placement de celles-ci revient à éliminer implicitement un certain nombre de rayons. En effet, à un ensemble de surfaces coplanaires ne correspond qu'une source-image et donc un seul rayon. Pour un ordre faible, le nombre de sources-images considéré est alors inférieur à celui de rayons. Lorsque l'ordre augmente, ce nombre croît théoriquement de manière exponentielle, mais reste en pratique limité (*chapitre 6*).

Si l'on considère le cas d'un déplacement du récepteur, il est alors plus aisé de conserver en mémoire les sources-images pour recalculer leur contribution au récepteur, comparativement à complet de rayons.

Dalenbäck qualifie cette différence en parlant de méthode de "point à point", contrairement à une méthode des rayons qualifiée de "point à volume" [Dal92].

Modèle de réflexion diffuse

Les hypothèses faites aux paragraphes précédents sur le comportement des ondes lors d'une rencontre avec une paroi modélisent un cas extrême du phénomène de réflexion. A l'opposé, une autre hypothèse considère que l'énergie arrivant sur les parois est réfléchi de manière diffuse, c'est-à-dire dans toutes les directions et de manière uniforme.

La formalisation de ce phénomène est connue sous le nom de Loi de Lambert. Elle se formule de la manière suivante. On suppose qu'une surface dS reçoit une énergie par unité de surface et de temps I , assimilable à une intensité. L'observation de l'intensité I_r réémise selon la direction θ à une distance r s'exprime par :

$$I_r(\theta) = \frac{1}{\pi r^2} (1 - \alpha) I dS \cos(\theta) \quad (2.5)$$

α est le coefficient d'absorption de la surface.

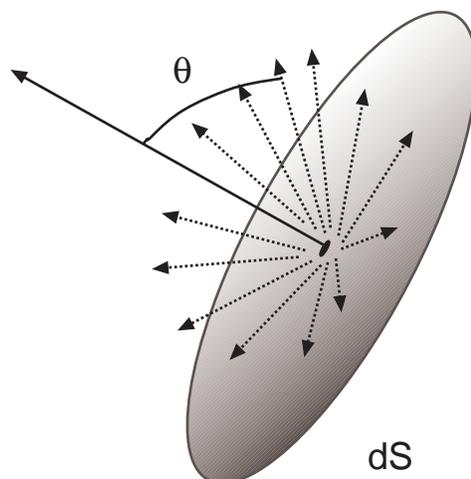


FIG. 2.7 – Modèle de réflexion diffuse selon la loi de Lambert

L'utilisation de cette loi permet d'utiliser un modèle de transfert de l'énergie sous forme diffuse entre les parois. Elle peut être également exploitée pour la simulation du champ tardif sous hypothèse de réflexion diffuse sur les parois. Le comportement réel des parois peut-être approché par une combinaison de ces deux modèles idéaux [CB99] ou encore faire appel à des modèles de radiosité à distribution directionnelle [Tsi98].

Simulation du champ tardif

Il a été vu en effet que la réverbération tardive correspond plutôt à une tendance vers l'établissement dans la salle d'un champ diffus, c'est-à-dire où toutes les directions de propagation sont équiprobables. Celui-ci s'établit d'autant plus vite que les parois de la salle renvoient l'énergie de manière diffuse et non spéculaire.

Différents modèles du champ tardif peuvent être décrit en exploitant différentes hypothèses: l'une considère l'établissement d'un champ diffus dans la salle (théorie de Sabine), tandis que l'autre suppose que l'échange d'énergie entre les parois s'effectue sous forme diffuse.

Théorie de Sabine

Cette théorie suppose qu'un champ diffus s'établit dans la salle; la distribution spatiale énergétique est alors uniforme. Après extinction des sources présentes dans une salle la densité volumique d'énergie suit une loi exponentiellement décroissante. On peut définir un temps au bout duquel cette énergie a décru de 60 dB, le temps de réverbération. La valeur de ce temps est donnée, selon Sabine par :

$$T_{60} = \frac{0.161V}{\sum_i \alpha_i A_i} \quad (2.6)$$

avec

$$\left\{ \begin{array}{l} V \text{ volume de la salle} \\ A_i \text{ aire de la surface } i \\ \alpha_i \text{ coefficient d'absorption de la surface } i, \text{ mesuré en champ diffus} \end{array} \right.$$

La valeur initiale de la densité d'énergie est donnée par

$$E_0 = \frac{T_{60}}{6 \ln(10)} \frac{W}{V} \quad (2.7)$$

où W est la puissance de la source.

Ce modèle n'est cependant valable qu'au delà d'un certain temps, le temps de mélange, à partir duquel l'hypothèse de répartition spatiale uniforme est valide. Sabine donne comme estimation de ce temps [Pie81] :

$$t_{mélange} = \frac{3 * l}{c} \quad (2.8)$$

avec

$$\left\{ \begin{array}{l} l \text{ grandeur "représentative" de la salle} \\ c \text{ vitesse du son (340 m/s)} \end{array} \right.$$

Cette grandeur représentative l peut être prise égale au libre parcours moyen \bar{l} , espérance de la longueur du trajet séparant deux réflexions, et s'exprimant par

$$\bar{l} = \frac{4V}{S} \quad (2.9)$$

ou S représente la surface de la salle.

Cette durée correspond en général à une surestimation, et ce pour deux raisons. En effet, étant donné que l'absorption des parois varie avec la fréquence, l'étalement temporel d'un écho croît à chaque réflexion ; d'autre part les parois de la salle étant également diffusantes, cela entraîne une augmentation de la densité temporelle d'échos.

On notera que cette théorie permet de calculer des paramètres décrivant le champ diffus et non un signal de réponse impulsionnel ou d'échogramme. Néanmoins ces paramètres peuvent servir à contrôler un processus de synthèse de réverbération tardive, sous réserve qu'il soit perceptivement pertinent (*chapitre 5*).

Radiosité

Contrairement au modèle décrit précédemment, cette méthode ne fait pas d'hypothèse sur la répartition spatiale du champ tardif dans la salle, mais suppose que les échanges d'énergie entre les parois s'effectuent de manière diffuse.

Un bilan de l'énergie reçue sur toutes les parois, supposées diffusantes, peut être effectué selon la loi de Lambert. Le récepteur est alors un observateur, dans la salle de ce processus d'échange d'énergie (*figure 2.8*).

La formalisation de ce principe conduit à la formulation de Kuttruff (*chapitre 6*).

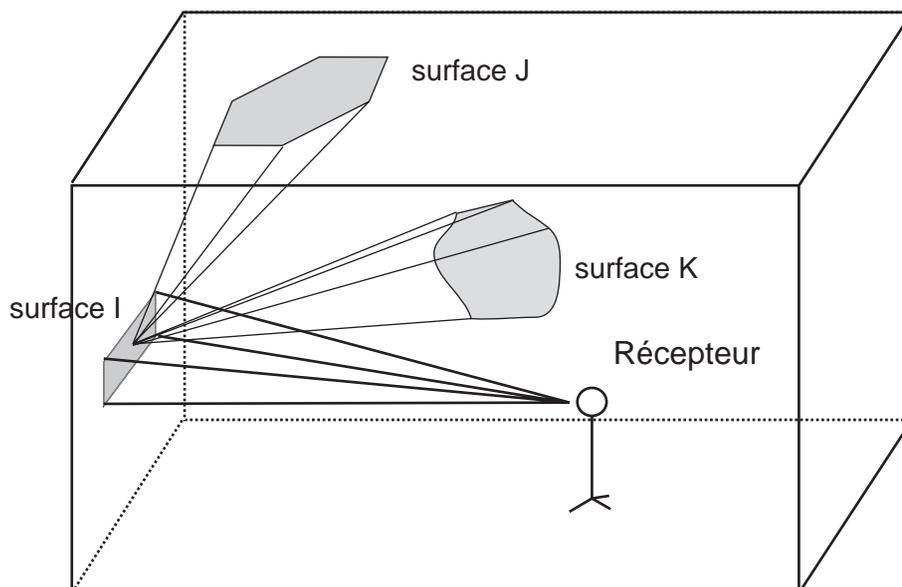


FIG. 2.8 – Evolution de l'énergie réverbérée entre les parois et le récepteur. Celui-ci est sensible à la contribution de la surface I, calculée à partir des contributions des autres surfaces.

Malcurt a utilisé cette méthode pour la simulation de la partie tardive d'une réponse impulsionnelle d'un logiciel de prédiction. Pour cela il a proposé une méthode de résolution numérique par discrétisation de la salle en surfaces élémentaires S_i [Ma186]. L'évolution du champ réverbéré s'exprime alors au moyen d'une matrice de passage A , dont les coefficients expriment le couplage entre les surfaces élémentaires S_i et d'une matrice diagonale des coefficients de réflexions α_i .

L'exploitation de cette matrice peut se faire soit pas à pas en progressant de manière temporelle et en parallèle sur toutes les surfaces, soit de manière synthétique en exploitant les propriétés de convergence pour certains régimes physiques :

- En régime stationnaire la distribution des intensités sur les parois ne dépend que de la matrice de passage et du vecteur d'intensité excitatoire¹¹, constitué par exemple des contributions des sources sonores présentes dans la salle ou étendu aux contributions des sources-images.

11. Il s'agit alors d'un régime forcé.

- En régime libre, et en supposant que l'on converge vers une décroissance exponentielle, la distribution des intensités et le temps de réverbération peuvent être approchés par une procédure itérative [Mal86],[CB99],[Kut95].

Selon le mode utilisé (progression temporelle ou synthétique), on parlera d'exploitation directe, ou indirecte.

Remarques : implantation

Grandeurs considérées

Les méthodes géométriques présentées précédemment font généralement abstraction de la nature ondulatoire du signal acoustique. Les calculs s'effectuent alors sur des grandeurs énergétiques, négligeant alors les phénomènes d'interférence, et la simulation produit des échogrammes et non des signaux de pression. Bien que cette démarche soit adaptée à l'évaluation d'indices acoustiques issus du signal simulé, elle n'est que partiellement adaptée à une utilisation ultérieure d'écoute par convolution. Une réponse doit alors être recalculée, de façon arbitraire, à partir de la racine carrée de l'échogramme obtenu. L'ajout d'une phase se fera en utilisant des techniques de traitement du signal. Il est cependant possible de calculer directement un champ de pression : soit par des méthodes géométriques (rayons, sources-images) en considérant la propagation en pression et en associant un filtre équivalent à chaque réflexion¹², soit par des méthodes de simulation numérique (différences finies, éléments finis). Mais dans le cas de salles (grands espaces), les puissances de calcul et l'espace mémoire nécessaires restent encore, pour ces dernières méthodes, au-delà des capacités des ordinateurs non dédiés.

Modèles hybrides

L'utilisation d'une méthode de prédiction particulière n'est adaptée qu'à la simulation d'une partie des phénomènes contribuant à l'effet de salle. Bien qu'en principe il ne soit pas possible d'associer directement un type d'hypothèse physique à une plage temporelle de la réponse impulsionnelle, les propriétés du champ tardif et des approximations effectuées sur la discrétisation des parois, rendent licites le transfert du processus de propagation spéculaire vers un modèle de radiosité.

Pour ce qui est du calcul du début de la réponse impulsionnelle, la méthode des rayons peut conduire à des imprécisions sur le niveau calculé (*chapitre 6*). Des techniques dites "hybrides" consistent à rechercher les sources-images correspondant à ces rayons et à recalculer le niveau correspondant au récepteur, permettant ainsi d'éviter les biais.

La prise en compte d'un modèle plus fin des réflexions sur les parois, comprenant une

12. En général ce filtre est déduit des caractéristiques données en bande de fréquences, et sous la forme d'un filtre à phase minimale.

part diffuse et une part spéculaire, permet en outre de prendre en compte les phénomènes de diffusion dès le début du calcul de la réponse impulsionnelle.

Les phénomènes physiques considérés jusqu'ici sont principalement la propagation, la réflexion et la diffusion. Une modélisation plus fine en prendra d'autres en considération ; ainsi la diffraction, qui permet de traiter les cas où la présence d'un bord diffractant sur le trajet du rayon entraîne une modification de son contenu fréquentiel. Il est ainsi possible de traiter les situations où un transducteur n'est pas directement visible par un autre (ou par une partie de la salle), et l'absorption en incidence rasante (seat-dip effect), qui considère les perturbations apportées par les rangées de sièges (celles-ci agissent comme un réseau diffractant et modifient la réponse en fréquence).

Enfin, il faut noter que les méthodes géométriques travaillent en général pour une bande de fréquence donnée (souvent une octave, ou un tiers d'octave). L'obtention des résultats pour un spectre plus large se fait soit en effectuant plusieurs simulations avec différentes valeurs de coefficients d'absorption et de diffusion, valides pour les bandes de fréquence considérées, soit en représentant l'action des parois sous forme de filtre.

2.2.2 Auralisation

L'auralisation d'un message par une réponse impulsionnelle peut s'effectuer de plusieurs manières. Si l'on utilise l'équation mathématique gouvernant le canal acoustique, il s'agit de réaliser la convolution entre deux signaux. Une autre approche consiste à réaliser un dispositif de traitement du signal permettant de recréer un effet de salle, aux propriétés perceptives identiques à la réponse originale.

Systemes basés sur la convolution

Principes

Cette première approche découle directement de la théorie des systèmes linéaires : l'auralisation d'un signal quelconque s'effectue par convolution avec la réponse du canal considéré.

Une contrainte majeure de cette méthode a longtemps été le coût de calcul, lors de l'application directe de l'équation de convolution. Le traitement d'une réponse de longueur N , nécessite en effet N additions-multiplications. Des techniques de convolution rapide basées sur la transformée de Fourier rapide, ont permis de faire passer l'ordre de ce coût à $4 * \log_2(2N)$ (pour des FFT sur $2N$ points), mais en introduisant une latence en échantillons proportionnelle à la taille de la transformée de Fourier. L'utilisation de méthodes hybrides, opérant un découpage progressif de la réponse impulsionnelle, couplée à l'augmentation de la puissance de calcul des processeurs, a permis néanmoins de

réaliser des convolutions en temps réel et à retard nul¹³ [Gar94].

D'autres problèmes subsistent cependant.

- utilisation dynamique : le passage d'une réponse à une autre s'effectue en général par interpolation sur l'ensemble des coefficients des réponses impulsionnelles. Il en résulte le problème de la mise à jour d'un vaste ensemble de paramètres, et celui de la validité, tant physique que perceptive, d'une configuration intermédiaire. En effet, la création d'un état intermédiaire, s'effectue généralement par interpolation entre deux jeux de coefficients correspondant à des configurations "réelles". Rien ne peut garantir que la loi d'interpolation utilisée (souvent monotone) permette d'obtenir des configurations correspondant à des mesures réelles. Par exemple, si l'on connaît les réponses impulsionnelles associées à deux positions de la source, rien ne permet de dire que la réponse correspondant à une position intermédiaire puisse se calculer par combinaison linéaire de ces deux signaux, liée notamment aux conditions de visibilité des sources ou des sources-images.
- dépendance au système de reproduction sonore : la convolution d'un signal avec une réponse permet de reproduire un signal monophonique, mais n'autorise pas de reproduction de la localisation. Un encodage multicanaux doit être utilisé. Deux schémas sont possibles : soit il correspond à cet encodage une technique de mesure (stéréophonie, binaural, Ambisonics), auquel cas on effectuera autant de convolutions que de canaux nécessaires, soit le principe d'encodage ne s'appuie pas sur un dispositif de prise de son (Surround, pan-pot d'intensité multicanaux) et il faudra créer artificiellement des signaux supplémentaires. Les besoins en puissance de calcul seront multipliés, selon le nombre de réponses utilisé. De plus, si l'on désire changer de système de reproduction sonore, il est nécessaire de recalculer un nouveau jeu de réponses, ou d'effectuer une transcription vers le nouveau codage utilisé.

Exemples

Cette approche par convolution a notamment été utilisée par la société LAKE DSP pour ses stations spécialisées Huron. Ces stations sont basées sur des cartes dédiées, chacune étant équipée de 4 processeurs traitement du signal Motorola (DSP 56002). Un algorithme adapté, et la présence de bus hautes performances permet d'effectuer ces opérations de convolution sur des réponses, et ce sans (ou avec très peu) de latence.

On peut également citer le système *Convolutron*, développé par la NASA et commercialisé par la société Crystal River. Ce système se présente sous la forme de cartes informatiques équipées de processeurs traitement du signal TMS 320.

13. Toutefois le coût de cette opération restera lié à la longueur de la RIF utilisée.

Synthétiseur générique d'effet de salle

Cette approche trouve ses racines dans les domaines de la prise de son et de la post-production ; l'adjonction de réverbération à des enregistrements permet d'en augmenter le naturel et le réalisme. Au départ naturels, avec l'utilisation de chambre d'échos, ces moyens sont par la suite devenus électromécaniques (réverbérateurs à plaques, à ressort), analogiques (lecteur de bandes magnétiques à plusieurs têtes de lecture), puis numériques.

L'avantage principal de ce type de synthétiseur tient au fait qu'il présente à l'utilisateur un ensemble de réglages constituant une modélisation paramétrique de l'effet de salle ; le contrôle de l'effet engendré est alors simplifié, mais la classe d'effets reproductibles est limitée aux champs d'action des paramètres choisis. On verra au paragraphe suivant qu'un autre désavantage est lié à l'architecture choisie pour une réalisation pratique.

Principe de réalisation

Les différentes parties de l'effet de salle sont généralement synthétisées séparément [Jot92]. La synthèse du son direct et des premières réflexions s'effectue par utilisation de filtres FIR permettant de contrôler l'amplitude et le temps d'arrivée de chaque écho. La synthèse de la réverbération s'effectue par la réalisation d'un "filtre réverbérant", dont la réponse impulsionnelle offre des propriétés semblables à celles d'un modèle statistique de la réverbération tardive. Les difficultés rencontrées lors de la réalisation d'un tel filtre sont de trois ordres :

- obtention simultanée d'une densité d'échos et d'une densité modale suffisante,
- éviter l'apparition de résonances et sonorités particulières qualifiées de "métalliques",
- contrôle indépendant du temps de réverbération et du niveau de l'énergie réverbérée selon la fréquence.

Différentes méthodes de réalisation ont été proposées. Schroeder a d'abord proposé une approche basée sur l'utilisation d'un filtre en peigne et d'un filtre passe-tout [Sch62]. Afin d'augmenter la densité d'échos et d'éviter des colorations fréquentielles, il proposa ensuite une amélioration de ce système en associant en parallèle des filtres en peigne, suivi de filtres passe-tout en série. Par ailleurs, l'introduction de fonctions de transfert passe-bas en lieu des gains de boucle des filtres en peigne, permet d'obtenir un temps de réverbération dépendant de la fréquence. Cette approche a été mise en pratique par Moorer.

Gerzon a proposé une généralisation du filtre passe-tout de Schroeder pour des réverbérateurs à plusieurs canaux d'entrée et de sortie [Ger76]. Stautner et Puckette ont proposé

en 1982 une généralisation multicanaux du filtre en peigne, permettant d'obtenir une densité d'échos suffisantes [SP87]. Jot a proposé une méthode générale d'élaboration des filtres réverbérants en partant de la structure proposée par Schroeder [Jot92].

Un désavantage de ce type de synthétiseur est la dépendance entre l'ensemble d'effets de salle synthétisables, et l'architecture du système. Par exemple, afin d'augmenter la densité décho pour la synthèse de la réverbération, le module de synthèse de la réverbération peut être alimenté par les signaux issus de la synthèse des premières réflexions. Si cette technique permet de faire des économies en termes de coûts de calcul, elle introduit néanmoins une dépendance entre la synthèse des différentes sections de l'effet de salle. Une conception différente, plus coûteuse, permettrait de pallier ce défaut. Ce type d'architecture empêche alors de synthétiser de manière contrôlée des réflexions hors des limites temporelles établies, limitant ainsi le type d'espaces décrits.

Exemples

A titre d'exemple de réalisation de tels synthétiseurs, on peut citer les modules de réverbération utilisés dans le monde de la musique et de la post-production (Yamaha, Roland, Lexicon). Cette approche a également été adoptée lors de la réalisation du Spatialisateur, par l'IRCAM et Espaces Nouveaux. Ce logiciel sera présenté plus en détail au chapitre 3 et en annexe.

2.2.3 Autres contextes

L'ensemble des méthodes présentées précédemment concerne principalement le contexte de l'acoustique des salles. La mise en œuvre des algorithmes de prédiction et auralisation fait alors l'objet d'implantation particulière, dans le cadre de logiciels dédiés.

Cependant, la reproduction d'un effet de salle peut être utilisée dans d'autres cadres que celui de l'acoustique architecturale. Il s'agit notamment des contextes de la musique et ceux de la réalité virtuelle : multimédia, télécommunication (la reproduction de la localisation et de l'effet de salle permet d'améliorer l'intelligibilité de vidéoconférences) et jeux.

L'examen de ces domaines (notamment le jeu) modifie le contexte d'utilisation des algorithmes présentés auparavant. En effet, alors que les contraintes, dans les domaines classiques de l'acoustique des salles, se portent sur la minimisation de l'erreur entre la simulation et la réalité, elles se portent alors sur le caractère plausible, éventuellement réaliste des espaces sonores créés. Une autre contrainte concerne, la possibilité d'évolution en temps réel de cet espace, particulièrement lors de l'interaction avec un utilisateur extérieur.

Un autre aspect important des univers virtuels est la présence, en général, d'une composante visuelle dynamique très souvent constituée de graphisme en trois dimensions.

La composante audio sert généralement à renforcer l'immersion dans cet univers virtuel graphique. Mais si sa contribution peut être déterminante, elle n'en reste pas moins au second plan en terme d'utilisation du processeur. En général, 80% de la puissance de calcul est dédiée au graphisme, 10% à l'audio et 10% au reste (gestion du scénario, des entrées sorties, composante réseau ...).

La mise en parallèle des méthodes graphiques et audio montre cependant de nombreuses similarités dans les méthodes et algorithmes utilisés, et il est possible d'établir une équivalence entre ces deux domaines ; un certain nombre de sources lumineuses sont placées dans un univers virtuel, et perçue par la caméra, qui joue le rôle du récepteur. Les algorithmes utilisés pour modéliser les phénomènes ondulatoires lumineux sont très proches de ceux présentés précédemment (ils portent parfois les mêmes noms : méthode des rayons, radiosité). La différence majeure entre les deux domaines, vient du fait que les phénomènes de propagation lumineux peuvent être considérés comme instantanés, contrairement au domaine de l'audio où la vitesse de propagation des ondes est nettement plus faible. Il peut alors s'avérer utile de mettre en commun un certain nombre d'algorithmes dont feront usage les moteurs audio et graphiques.

Dans ces contextes, l'écriture de moteurs de rendu audio peut se révéler relativement coûteuse et longue ; c'est pourquoi les fabricants de systèmes opérationnels, ou de cartes sons proposent aux programmeurs des bibliothèques informatiques spécialisées, (Application Programming Interface, API) permettant de diminuer le temps et les coûts de réalisation. La qualité du résultat repose alors sur celle de ces bibliothèques, dont la fabrication est souvent protégée. On peut citer dans cette catégorie :

- Microsoft, avec les bibliothèques DirectSound et DirectSound3D, composantes de l'ensemble DirectX, et qui présente notamment aux développeurs, une bibliothèque unifiée et indépendante du matériel, sur ordinateur compatible Windows™ PC.
- Creative Labs/E-mu, avec l'environnement EAX (Environmental Audio), extension de DirectSound3D. Cet environnement est notamment adapté aux ordinateurs compatibles PC équipés de cartes sons SoundBlaster Live, fabriquées par ... Creative Labs. La position dominante de Creative sur ce marché en fait cependant un quasi-standard.
- Aureal, concurrent direct de Creative Labs, avec l'environnement A3D. Il utilise la technologie WaveTracing développée par Aureal, et est particulièrement adapté aux ordinateurs compatibles PC, équipés d'une carte ... Aureal Vortex.
- Spatializer, société spécialisée dans les technologies d'auralisation binaurales (3-D Stereo™, N-2-2™, enCOMPASS™). Ces technologies ont été licenciées par de nombreuses sociétés d'informatique, dont notamment Compaq, Apple, Panasonic, JVC, et Motorola.

- Apple, avec la librairie SoundSprocket destinée aux créateurs de jeu, et répondant aux problèmes de l'auralisation des sources, et de l'effet de salle. L'acquisition récente d'une licence auprès de Spatializer, pour la technologie enCOMPASS, fait preuve de l'intérêt nouveau que porte Apple à une technologie ayant peu évolué depuis son apparition, il y a environ 4 ans.

Un examen plus approfondi des librairies EAX et SoundSprocket sera effectué au chapitre 3.

2.3 Représentation et manipulation de l'effet de salle

Les paragraphes précédents ont permis d'exposer un certain nombre de traitements et méthodes concernant l'effet de salle et ce, plus particulièrement dans le cadre de l'acoustique des salles. Afin de se placer dans le contexte du temps-réel, cadre de ce mémoire, il convient d'effectuer une séparation de ces méthodes selon les contextes d'utilisation, et les types de problèmes liés à la notion de temps réel dans chacun de ces contextes. Pour d'effectuer cette catégorisation, on considère différents niveaux d'intervention sur l'effet de salle.

2.3.1 Représentation bas-niveau

Un premier niveau de représentation, que l'on qualifiera de "bas-niveau" est celui relatif au dispositif de traitement du signal utilisé pour l'auralisation. Cette appellation est justifiée par le fait que les paramètres d'une telle représentation sont destinés à régler un dispositif de traitement de signal, et n'ont pas forcément de signification par rapport à un utilisateur.

L'exploitation de cette représentation requiert des connaissances spécifiques provenant des domaines du traitement du signal et de l'acoustique.

Réponse impulsionnelle (RIF)

Sa légitimité, tant au niveau acoustique que traitement du signal, en fait un candidat naturel pour représenter un effet de salle. Elle est principalement utilisée conjointement à des moteurs de rendu basés sur la convolution ; les principaux problèmes rencontrés sont ceux du calcul en temps-réel et sans retard de cette opération, et du rafraichissement des coefficients de la RIF. On rappelle de plus que l'information apportée par chaque échantillon de la réponse, n'est pas forcément perceptivement pertinente.

Description morphologique

Elle prend en compte les résultats de la psychoacoustique pour décrire un canal acoustique, par la forme globale de sa réponse impulsionnelle et non par la donnée de chacun de ses échantillons. Elle est *a priori* plus adaptée à l'utilisation conjointe avec un synthétiseur générique d'effet de salle.

Une telle description s'appuie sur un découpage temporel de la réponse, selon les limites évoquées précédemment.

La description fournit comprend :

- des limites temporelles pour la segmentation en onde directe, premières réflexions, cluster, réverbération,
- des coefficients énergétiques pour chacune des plages de la réponse impulsionnelle correspondante. En reprenant le formalisme décrit au paragraphe 2.1.2 ils seront notés OD , $R1$, $R2$, $R3$. L'utilisation systématique de ces quatre coefficients dépendra des capacités du moteur de rendu, ainsi que du contexte d'utilisation¹⁴,
- Un taux de décroissance pour la partie tardive de la réponse, caractérisé par le Tr ,
- Des informations relatives à la localisation du son (azimuth, élévation).

Afin de décrire des espaces sortant du cadre psychoacoustique sous-jacent à ce type de représentation, il est nécessaire d'étendre cette représentation, en autorisant, par exemple, la présence d'événements acoustiques en dehors de toute limite temporelle (par exemple, réflexions spéculaires arrivant hors des limites de $R1$). Il faut alors étendre le schéma précédent, et ce suivant les capacités du dispositif d'auralisation, et les exigences de réalisme du dispositif mis en place. Des premières améliorations peuvent constituer en :

- une prise en compte des variations fréquentielles : OD , $R1$, $R2$, $R3$ et Tr sont donnés pour plusieurs bandes. En général, leur nombre est de 3.
- une représentation plus précise des premières réflexions (la donnée de $R1$ seul induit un mode de resynthèse arbitraire), et ce en raison de leur importance sur la perception. Une telle description peut comprendre :
 - le nombre de réflexions considéré, N_{R1} .

14. Par exemple, si l'on dispose d'une puissance de calcul réduite, on pourra effectuer la synthèse d'un effet de salle simplifié (son direct + réverbération), mais cependant plausible.

- pour chaque réflexion : un temps d'arrivé, une direction, une fonction de transfert représentant l'ensemble des transformations subies par cette réflexion.
- des valeurs de densité d'échos et de modes nécessaires à un rendu naturel de la réverbération
- la possibilité de placer des événements sonores indépendamment des limites temporelles (réflexion isolée).

Passage entre les représentations bas-niveau

Le passage d'une représentation à une autre n'est pas bijectif. En effet, si dans le sens Réponse \rightarrow Représentation Morphologique, il s'effectue par une série de méthodes de traitement du signal (*cf.* 5.1), permettant d'extraire des indices d'une réponse, il n'est pas possible d'associer une réponse unique à un ensemble d'indices donné. Une resynthèse arbitraire est alors nécessaire, devant obéir à la contrainte de similarité perceptive. Selon le degré de finesse présent dans la description morphologique, cette part d'arbitraire sera plus ou moins élevée. Par exemple, dans le cas où les premières réflexions sont représentées par un coefficient énergétique unique, $R1$, différents problèmes sont à résoudre pour la synthèse des premières réflexions : le choix arbitraire d'un nombre d'échos, la répartition de l'énergie entre ces réflexions, leur répartition temporelle selon une loi arbitraire (régulièrement, aléatoirement...), l'attribution de directions de provenance de ces réflexions (autour de la source, latéralement, aléatoirement).

Les méthodes de traitement du signal utilisées nécessitent une connaissance préalable du domaine d'étude, c'est-à-dire l'acoustique des salles, permettant ainsi de cerner le champ de méthodes applicables, autant pour l'analyse que pour la resynthèse (*figure 2.9*). Ces méthodes seront détaillées au chapitre 4.

2.3.2 Manipulation de l'effet de salle

Les représentations bas-niveau évoquées précédemment concernent principalement une description de l'effet de salle par rapport à un dispositif de traitement du signal et n'ont généralement pas de sens dans le cadre d'une utilisation pratique par une personne non experte dans ce domaine. Sans entrer dans une description exhaustive de toutes ces utilisations, il est néanmoins possible d'en effectuer des regroupements selon différents axes, non forcément orthogonaux.

On peut distinguer ainsi :

- les utilisations unidirectionnelles par opposition aux utilisations bidirectionnelles : par exemple, si l'on désire représenter une situation de promenade dans un monde

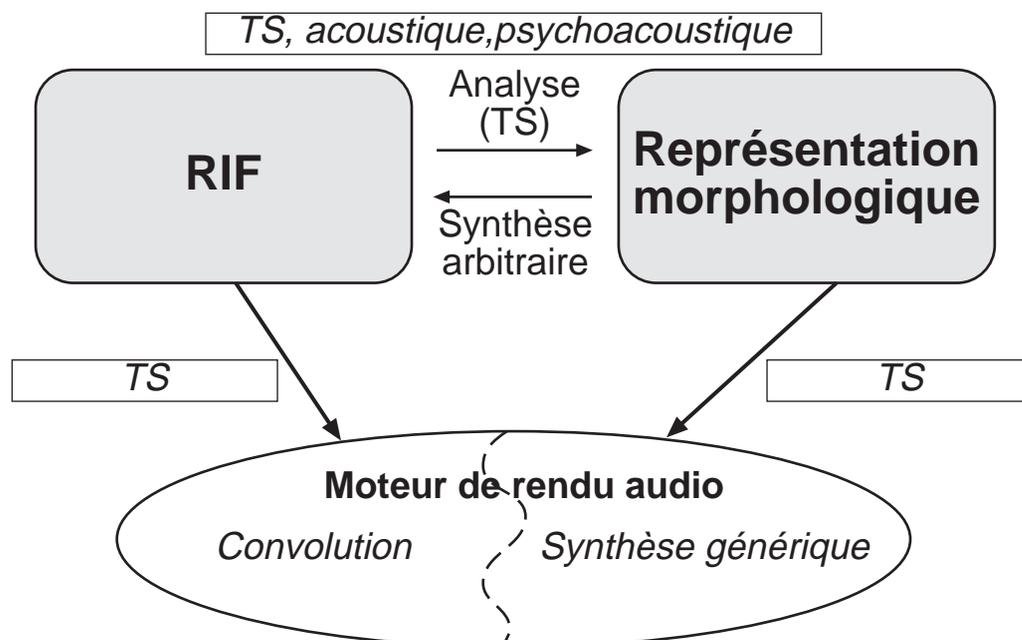


FIG. 2.9 – Représentations bas niveau de l'effet de salle. Ce schéma indique les liens entre ces représentations, ainsi que les domaines d'expertise concernés (cadres rectangulaires), lors de la réalisation de ces liens (TS : traitement du signal).

virtuel, il convient de séparer l'utilisation par un utilisateur final de l'élaboration de ce monde¹⁵,

- les utilisations temps réel par opposition aux utilisations temps différé : typiquement dans un contexte d'acoustique architecturale, le temps réel n'est pas une caractéristique recherchée *a priori*, alors que la promenade dans un monde virtuel le réclame,
- les utilisations nécessitant un rendu réaliste par opposition à celles nécessitant un rendu plausible. L'exemple précédent peut être à nouveau utilisé. Un contexte d'acoustique architecturale, spécialement lors de la phase de conception d'un espace, nécessite des simulations précises et proches de la réalité. Lors de la promenade dans un monde virtuel, c'est la sensation d'être dans l'espace créé qui prime ; si de plus une composante visuelle est présente, la composante audio doit servir à renforcer celle-ci et il est nécessaire d'éviter d'éventuelles incohérences perceptives entre ces deux sens.

Il convient, de plus, d'introduire la notion de *scène sonore virtuelle* qui regroupe un ou

¹⁵. Typiquement cette dernière phase est composée d'une succession d'essais jusqu'à l'obtention du résultat désiré. On parle en anglais de *sound design* ou *sound authoring*.

plusieurs triplets source-salle-récepteur, chacun possédant une représentation bas-niveau destinée au dispositif de rendu sonore. Elle peut également comprendre un ensemble de règles relatives à l'évolution de ces triplets. On appellera *interface de manipulation* le lieu de l'interaction de cette scène sonore avec un utilisateur. Une telle interface doit effectuer la traduction des actions de l'utilisateur vis-à-vis d'une représentation bas-niveau, et ce par un ensemble de méthodes issues de l'acoustique des salles. Le type d'utilisation, et les contraintes associées, permettront d'effectuer le choix de ces méthodes.

Propositions d'interface de manipulation

Interface géométrique

Elle représente le dispositif physique engendrant un effet de salle. La salle est discrétisée sous forme de facettes positionnées dans l'espace, auxquelles sont associés des coefficients d'absorption et de diffusion. Les valeurs de ces coefficients peuvent varier avec la fréquence, et sont donc données pour différentes bandes. Les sources et récepteurs sont décrits par leur position, leur orientation, ainsi que des caractéristiques de directivité.

De par sa composante visuelle intrinsèque, cette interface de manipulation est, *a priori*, très séduisante, et ce notamment dans le contexte du "sound design". En effet, une appréhension "naturelle" de la qualité acoustique d'un lieu se fait souvent en référence à ses propriétés architecturales. Par exemple, on parlera de l'acoustique d'une cathédrale, pour décrire un espace grand et aux surfaces réfléchissantes. De manière plus générale, des réverbération longues sont associées à de grands volumes, et ce sans se soucier des propriétés acoustiques réelles des parois.

C'est pourquoi son usage, lors de la phase d'élaboration d'un monde virtuel, ne permet pas une appréhension intuitive et fine de l'évolution de la qualité acoustique d'un lieu. En effet les propriétés des matériaux, ainsi que leur disposition relative, jouent un rôle prépondérant dans l'effet de salle obtenu. Or ces propriétés sont rarement connues par un utilisateur non acousticien et des modifications sont généralement peu prévisibles.

Si l'on examine le couplage d'une telle interface avec un moteur de rendu, par l'intermédiaire d'une représentation bas-niveau, on se rend compte que le choix de celle-ci peut être prépondérant pour la contrainte de temps réel. On verra en effet au chapitre 4, que s'il semble à l'heure actuelle improbable d'arriver en temps réel à l'obtention d'une réponse impulsionnelle complète, certains paramètres d'une représentation morphologique sont eux calculables dans ce contexte (notamment en ce qui concerne la réverbération tardive). Un modèle "mixte", décrivant finement le début d'une RIF, et morphologiquement la fin peut alors être utilisé, et ce sans perte d'information du point de vue perceptif.

Interface perceptive

Elle s'intéresse à une manipulation de la perception de l'effet de salle. Une telle interface a notamment été développée au laboratoire d'acoustique des salles de l'IRCAM. On se contentera ici de présenter les résultats de ces recherches (une description plus précise de la démarche ayant abouti à ces résultats sera présentée au chapitre suivant). Un canal acoustique est décrit par 9 paramètres indépendants regroupés dans le tableau suivant.

nom (F)	Nom (US)	Effet perçu
présence de la source	source presence	niveau perçu de la source sensation de proximité
présence de la salle	room presence	niveau perçu de l'effet de salle
réverbérance	late reverberance	sensation de rémanence, de traînage du son
réverbérance courante	running reverberance	idem, mais pendant le discours courant
enveloppement	envelopment	rappelle la largeur apparente de la source
brillance	source brilliance	perception de la source en hautes fréquences
chaleur	source warmth	perception de la source en basses fréquences
vivacité	liveness	perception de la réverbérance en hautes fréquences
lourdeur	heaviness	perception de la réverbérance en basses fréquences

A cela, viennent s'ajouter deux facteurs concernant la localisation du son. Les liens avec un ensemble de critères objectifs seront explicités dans un chapitre suivant. Il faut cependant préciser que les recherches ayant permis d'aboutir à ces résultats ont été basées sur des mesures de réponses impulsionnelles utilisant des transducteurs omnidirectionnels. Il convient de rajouter des paramètres permettant de décrire la directivité de ces transducteurs, autant en ce qui concerne le comportement en propagation "directe", que "diffuse".

Un avantage certain de cette interface apparaît dans les situations de "design" acoustique. Si elle nécessite un certain temps d'apprentissage, elle offre par la suite une intuition accrue, doublée d'une hiérarchisation des paramètres de contrôles (typiquement les

contrôles **présence de la source**, **présence de la salle**, **réverbérance** permettent d'effectuer un réglage "basique" de l'effet de salle, les autres contrôles donnant accès à un réglage plus subtil pour la perception). Néanmoins, et de manière opposée à l'utilisation d'une interface géométrique, l'utilisation d'une telle interface peut s'avérer déroutante dans la mesure où elle permet de s'affranchir complètement des lois de la physique. A titre d'exemple, la configuration d'un effet de distance se traduit par une variation du contrôle **présence de la source**, mais il reste à relier si besoin les variations de ce contrôle à une loi physique. Il peut toutefois être intéressant, lors de la création d'un espace virtuel sonore, d'initialiser une telle interface à l'aide des paramètres de réglages d'une interface de type géométrique ou statistique, puis d'utiliser par la suite une interface perceptive afin de modifier dynamiquement l'effet de salle¹⁶.

Enfin on notera que cette interface présente des liens très forts avec une représentation morphologique bas-niveau, ceux-ci pouvant servir de base à des calculs d'indices objectifs et donc perceptifs. Le couplage entre ces deux entités s'effectue donc très simplement, et à très faible coût.

Interface géométrique statistique

Elle peut être considérée comme une version simplifiée d'une interface géométrique, qui prend en compte les résultats d'une modélisation statistique de l'effet de salle (et particulièrement la réverbération tardive). De part sa construction, elle élimine *a priori* le contrôle des réflexions précoces et le cluster. Néanmoins il est possible de rajouter un régime de réflexions de manière arbitraire (niveau, répartition temporelle, direction) , pour autant que le contrôle des caractéristiques de ce régime soit accessible à l'utilisateur.

S'il permet l'appréhension d'un phénomène d'effet de salle simplifié, ce type d'interface présente néanmoins l'inconvénient d'utiliser des paramètres de contrôle non indépendants, sur le plan perceptif, pour le réglage de certaines caractéristiques de l'effet de salle engendré. On se reportera aux formules issues de la théorie de Sabine afin de vérifier cette dépendance en ce qui concerne le champ réverbéré (par exemple, le volume agit simultanément sur la durée et le niveau de réverbération, notions indépendantes sur le plan perceptif (*paragraphe 2.2.1*)). De plus les paramètres utilisés étant liés et de nature différente (volume de nature géométrique, temps de réverbération d'ordre perceptif...), il se repose le problème de la confusion engendrée par leur utilisation.

De manière semblable à une interface perceptive, les liens avec une représentation morphologique sont très forts. On verra au chapitre 6 que ce type de modèle est couramment

16. Il se pose néanmoins lors de la phase "utilisation", le problème du réglage de la représentation bas-niveau. En effet, soit on conserve, de manière transparente à l'utilisateur, les données issues des deux interfaces de réglages, soit on décide d'unifier ces données en références à une seule interface, ou à une représentation bas-niveau.

utilisé dans les bibliothèques de manipulation de son 3D proposées par les sociétés informatiques.

2.3.3 Critères d'évaluation des modalités de représentation d'une scène sonore

La définition d'un certain nombre de critères permet d'effectuer d'abord une séparation entre les modalités bas-niveau et haut-niveau de représentation de l'effet de salle. Elle permet en outre de définir une classification des interfaces de manipulation.

Opérabilité

Ce critère comprend deux notions ; celle de la durée d'apprentissage et celle du caractère intuitif du processus de manipulation d'une interface, toutes deux à envisager lors de la phase de conception de l'application autant que dans la phase d'utilisation. Cette propriété sera par exemple présente lorsque l'interface adresse directement un des sens de l'utilisateur ; ainsi une interface géométrique offrira une bonne maniabilité de par son caractère visuel, tandis qu'une interface perceptive répondra également de manière satisfaisante à ce critère grâce à son caractère auditif, et par un apprentissage très court. L'opérabilité se mesure également au fait que le modèle de représentation s'appuie sur un formalisme qui autorise l'interpolation et les opérations de morphing.

Pertinence perceptive

Le critère de pertinence perceptive inclut les notions de réalisme ou de plausibilité du rendu final par rapport à la situation décrite. Il inclut également le fait que les paramètres de commandes du système se trouvent en correspondance bijective avec des effets perceptifs indépendants entre eux.

Extraction automatique des paramètres de contrôle

Ce paramètre mesure la possibilité de réglage automatique d'une interface de manipulation à partir de la donnée d'une représentation bas-niveau. Par exemple, il s'avère impossible de retrouver la configuration physique ayant engendré un effet de salle donné, ou alors selon des contraintes très fortes (forme de salle imposée, aux surfaces recouvertes du même matériau). En revanche, il est envisageable de retrouver les paramètres de réglages d'une interface perceptive ou statistique, et ce à partir d'un enregistrement. Ce type de réglage peut s'avérer utile si l'on désire créer une interaction entre l'utilisateur et l'effet de salle engendré. Par exemple lors de réalisations d'incrustations sonores, il est ainsi possible d'inclure dans une scène enregistrée un événement en cohérence avec la scène.

Modularité

Elle indique la possibilité d'utiliser des éléments fournis par le programme d'exploitation de la scène, pour différents moteurs de rendu sonore ou différentes configurations d'un même moteur de rendu. Par exemple, une réponse impulsionnelle est spécifique d'un moteur basé sur la convolution (sauf conversion préalable vers une représentation morphologique)¹⁷.

Scalabilité, Efficacité

Elle évalue l'adaptation d'une interface à différents contextes, et notamment en fonction de la puissance de calcul disponible. A titre d'exemple, l'utilisation d'une interface de type géométrique, nécessite, pour communiquer avec le moteur de rendu, la mise en œuvre d'algorithmes issus du domaine classique de l'acoustique des salles. Le coût de calcul de tel algorithme est non seulement élevé, mais est de plus dépendant de la complexité de la salle. Si l'on désire effectuer une telle utilisation, sur un ensemble de processeurs de puissances différentes, il faudra proposer des simplifications tant algorithmiques que sur la description du modèle, ce qui peut entraîner une détérioration du réalisme de la simulation.

Il faut distinguer :

- l'efficacité du codage de la scène sonore proprement dite c'est-à-dire le nombre de paramètres à entrer pour décrire la scène (liée à la complexité de la tâche du designer).
- l'efficacité du codage des informations en sortie du programme d'exploitation de l'interface vers le moteur de rendu sonore : elle prend son importance lorsque ces données sont transmises vers le moteur de rendu à travers un réseau.

Conversion entre modalités

Ce critère évalue la possibilité de cohabitation entre plusieurs modalités, c'est-à-dire la possibilité de manipuler de manière cohérente une scène sonore à partir de différents types d'interface. Cette possibilité repose sur l'existence de relations de conversion entre les modalités de description, soit directement, soit indirectement à partir d'une représentation bas niveau. Il faut noter que ce type de relation entre deux interfaces n'est pas bijective.

17. De manière similaire une RIF est spécifique d'un type de prise de son elle même associée à un dispositif de reproduction ; on ne peut s'en resservir directement pour une autre configuration d'écoute

Comparaison des modalités de description d'une scène sonore

Le tableau suivant décrit l'évaluation de trois modalités de représentation de l'effet de salle selon ces critères.

	Signal	Architecturale	Perceptive
Réalisme, Plausibilité	REALISTE <i>(authenticité)</i>	REALISTE <i>(réalisme : dépendant des performances des algo. de simulation)</i>	PLAUSIBLE
Interactivité	PAUVRE	OUI <i>(visite virtuelle)</i>	OUI
Opérabilité	PAUVRE	DELICATE <i>(composition de scène sonore difficile)</i>	GRANDE
Scalabilité	NON	OUI	OUI
Méthode d'extraction	NON	NON <i>(méthode d'inversion encore embryonnaire. Impossible à partir d'un enregistrement)</i>	OUI <i>(envisageable à partir d'un enregistrement)</i>
Modularité	NON <i>(par exemple : contributions salle/transducteurs inséparables)</i>	OUI	OUI
Efficacité codage scène sonore	FAIBLE <i>(plusieurs Koctets pour chaque réponse)</i>	ELEVEE <i>(dépendant finesse de discrétisation spatiale des parois et de la directivité des transducteurs)</i>	TRES ELEVEE
Coût de calcul pour fournir les paramètres de rendu	***	TRES ELEVE	TRES FAIBLE
Efficacité codage paramètres rendu sonore	FAIBLE	dépend du type de moteur de rendu	TRES ELEVEE
Coût de calcul du moteur de rendu associé	ELEVEE <i>(convolution)</i>	dépend du type choisi	FAIBLE <i>(effet générique)</i>

L'examen de cette évaluation conduit aux résultats suivants. Tout d'abord la modalité

signal peut être rangée du côté des représentations bas-niveau en raison de sa faible adéquation avec les critères d'interactivité, d'opérabilité, de scalabilité, d'efficacité et de modularité. Il est également possible d'effectuer une comparaison entre une interface de manipulation géométrique et une de type perceptif. Tout d'abord on notera que ce dernier type a pour cadre celui plus limité de la plausibilité. Les autres différences entre ces deux interfaces se retrouvent principalement sur les critères d'opérabilité et sur l'existence d'une méthode d'extraction associée : elles apparaissent notamment lors des phases de création d'un univers virtuel. La différence sur le coût de calcul associé à chacune de ces interfaces ne peut être considérée comme déterminante en raison de l'accroissement des capacités de calcul en informatique.

2.4 Conclusion

2.4.1 Synoptique d'un système de manipulation de l'effet de salle

L'ensemble des considérations évoquées dans ce chapitre est regroupé sous la forme d'un schéma synoptique d'un système de manipulation de l'effet de salle.

Plusieurs concepts sont représentés sur ce schéma.

- les niveaux de traitement de l'effet de salle (en cadre arrondi) : ce sont le moteur de rendu, la représentation bas-niveau, et l'interface de manipulation,
- les liens entre ces niveaux : ils peuvent comprendre plusieurs caractéristiques : l'existence, la réflexivité, le fait d'être exécutés en temps réel,
- les domaines d'expertise (en cadre droit) nécessaires au traitement de chacun de ces liens.

Lien Moteur-Représentation bas-niveau

Ils font essentiellement appel à des méthodes de traitement de signal, afin de garantir des exécutions rapides. L'utilisation d'une représentation morphologique nécessite cependant des indications quant aux méthodes de synthèse. Ces méthodes peuvent être similaires à celles utilisées lors d'une conversion *représentation morphologique* vers une *réponse impulsionnelle*.

Le passage entre les deux représentations bas-niveau nécessite des connaissances quant à la forme du signal considéré, permettant ainsi d'adapter les méthodes d'analyse et de synthèse. Ces méthodes feront l'objet du chapitre 3.

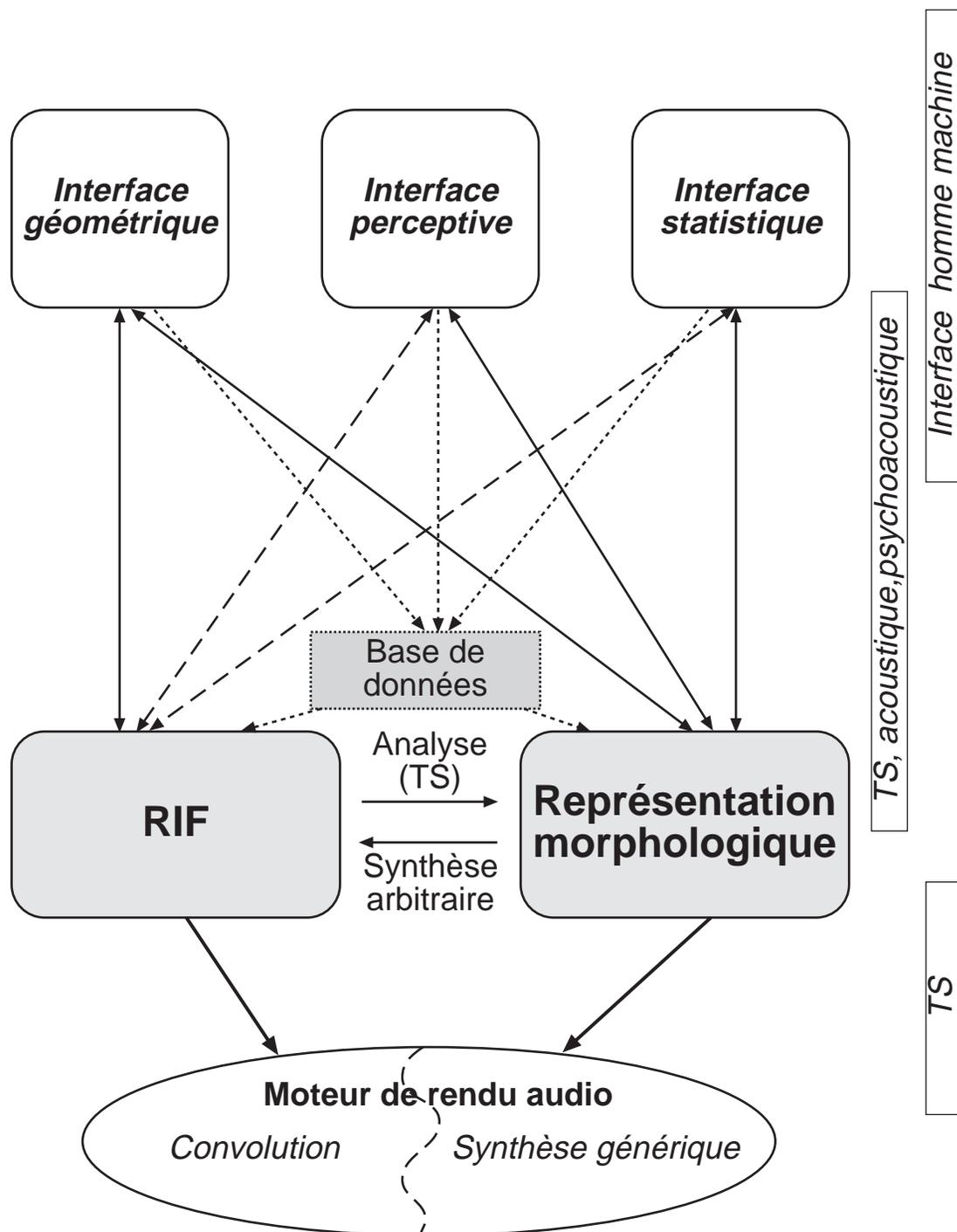


FIG. 2.10 – Synoptique d'un système de manipulation de l'effet de salle

Lien Représentation bas-niveau / Interface Haut niveau

Ces liens présentent plusieurs caractéristiques :

- l'existence: cette notion traduit la possibilité de réalisation de ces liens. Par exemple, l'existence d'un lien entre une représentation de bas-niveau et une inter-

face de manipulation géométrique ne peut *a priori* avoir lieu¹⁸.

- La réflexivité : elle traduit le fait de pouvoir parcourir les liens dans les deux directions. L'exemple précédent en est un exemple typique
- Le temps réel : il évalue la possibilité de réaliser la traduction des paramètres entre ces deux niveaux de manière instantanée. Prenons l'exemple du passage entre une interface de manipulation de type géométrique et une représentation bas-niveau. Si le passage direct vers une réponse impulsionnelle s'avère plus coûteux voire non réalisable en temps réel (il revient à utiliser les méthodes traditionnelles d'acoustique prévisionnelle), on verra au chapitre 4 que le passage vers une représentation mixte est lui réalisable sous cette condition. Si de plus, la conversion entre ces deux représentations bas-niveau est réalisable en temps réel, alors le cheminement précédent l'est également.

Les domaines d'expertise nécessaires à la réalisation de ces liens sont principalement ceux de l'acoustique prévisionnelle, pour le choix des méthodes, et de la psychoacoustique quant à la manière d'effectuer des simplifications pour une exécution temps réel, et toujours pertinente (l'oreille restant le juge final). L'examen de cette démarche constitue le cadre principal de cette thèse.

En pointillé sur le schéma, figure la notion de base de données. En effet, au lieu d'effectuer un calcul en temps réel des paramètres bas-niveau de l'effet de salle, il est possible d'utiliser le résultat de mesures pour effectuer une auralisation. Le passage entre deux points mesurés s'effectuera par interpolation des paramètres d'une représentation bas-niveau, avec le problème de la validité d'une telle procédure d'interpolation.

2.4.2 Présentation des chapitres suivants

Ce chapitre a présenté un ensemble de méthodes utilisées en acoustique des salles, en les reliant à un contexte d'utilisation. En rapport avec l'objet de cette thèse, les liens marqués "temps réel" dans le schéma précédent feront l'objet d'une étude particulière dans les chapitres suivants.

Le chapitre 3 approfondit d'abord la notion de scène sonore, en discutant les langages de description associés ; il traite également des simulations perceptives et statistiques. Le chapitre 4 présente un ensemble de bibliothèques informatiques, consacrées à la simulation temps réel d'espaces sonores, et examine leur possibilité de contrôle en rapport avec les conclusions du chapitre 3.

18. A moins d'imposer des contraintes sur la forme de la salle ou la répartition des coefficients acoustiques.

Puis, on revient à l'étude de certains liens du schéma synoptique. Les correspondances entre les représentations bas-niveau sont détaillées dans le chapitre 5. Le chapitre 6 traite de la simulation géométrique, et de son exécution en temps réel.

Finalement l'organisation d'une librairie informatique, traduction du synoptique précédent est présentée. On décrira à la fois sa structure générale, puis les parties ayant fait l'objet d'une implantation poussée.

Chapitre 3

Description d'une scène sonore virtuelle

Il a été présenté au chapitre précédent un cadre pour un dispositif de représentation et de traitement d'une scène sonore virtuelle. Cependant un tel système concerne seulement la manipulation de données présentes en mémoire; il semble donc naturel de s'intéresser à son alimentation à partir d'entités extérieures, fichier ou flux réseau. Le traitement de ces données peut être, soit une visualisation simple (lecture seule), soit une modification de leur valeurs (lecture et écriture).

Dans le cas présent, les données représentent une scène sonore virtuelle (chapitre 2.3.2). Ce chapitre est consacré à l'exposé de différentes stratégies pour l'élaboration d'un langage de description et de stockage d'une scène sonore virtuelle. Bien que naturellement très imbriquées, on essaiera cependant de conserver la séparation entre les notions de langage de description et d'interface de manipulation de son contenu. Enfin, un examen des langages existant, relatifs à l'échange de données sur l'Internet, sera effectué.

3.1 Description de l'effet de salle

3.1.1 Scène sonore virtuelle

Cette entité a été introduite au chapitre précédent; la définition donnée est "*ensemble de triplets {source-salle¹- récepteur} accompagné de règles concernant leur évolution*". La notion de règles d'évolution sert principalement à préciser le type d'interaction entre les trois composantes d'un triplet lors de sa manipulation par un utilisateur extérieur. Elle peut s'avérer particulièrement utile dans les domaines de la réalité virtuelle, dans la mesure où il n'existe aucune obligation de respect des lois physiques reliées à un canal acoustique. Par exemple, afin de recréer une situation d'apparté entre deux personnes, on peut choisir de ne pas rendre les variations de niveaux induites par les mouvements des interlocuteurs (déplacement et orientation). Ceci augmentera la sensation de face à face propre à ce genre de situation.

Un point important dans cette définition d'une scène virtuelle, est la notion d'ensemble. Plusieurs possibilités peuvent se présenter.

- Les triplets diffèrent par les récepteurs: dans le cas où le rendu du monde virtuel a lieu sur une machine unique, et pour un seul utilisateur, cette situation peut servir à préciser plusieurs angles de vue de la scène jugés importants par le concepteur. Un autre cas possible est celui où le rendu du monde virtuel est effectué par une machine, puis redistribué sous forme de flux audio spécifiques, ou de flux paramétriques, vers chaque poste de travail (par exemple pour un jeu en réseau).

1. Le terme "salle" est utilisé ici dans un sens générique, et peut également décrire des espaces ouverts

Cette spécificité peut être liée soit à l'angle de vue particulier qu'il faut restituer, soit aux conditions de restitution propres à chaque terminal².

- Les triplets diffèrent par les sources ou par les salles : typiquement cette situation décrit des scènes virtuelles riches, ou les interactions avec un utilisateur sont très élaborées. Il est dans ce cas utile de pouvoir ajouter, dans le langage, des indications à destination du moteur de rendu qui lui permettront de gérer au mieux ces situations en fonction de ses capacités de traitement. En effet, bien que des moteurs puissent avoir des capacités différentes, il est important qu'ils produisent un résultat uniforme. Pour cela la présence d'une hiérarchie entre les éléments du langage, ou la présence de stratégies de remplacement explicite, est nécessaire.
- Les triplets n'ont aucun champ commun : il s'agit d'un mélange des situations précédentes.

Enfin on notera que la notion de langage offre une utilité dès lors que plusieurs professions interviennent dans la création d'un monde virtuel. Typiquement, trois personnes sont concernées : l'auteur qui cherche à concrétiser la réalisation d'une idée, le designer qui a pour mission de transcrire et de développer l'idée selon les possibilités offertes, et le programmeur qui doit implanter le résultat. Le travail de mise au point ne peut s'effectuer qu'à travers une démarche de communication entre les différentes parties, et ce à partir d'un langage unique³.

Contraintes

Afin de réduire le champ de recherche pour la mise au point d'un tel langage, certaines propriétés et contraintes peuvent s'avérer utiles :

- Le langage de description doit être indépendant et du système de rendu audio, et, dans la mesure du possible, du type d'interface utilisée pour sa manipulation. La cohérence de l'interface et du langage peut nécessiter des traitements complémentaires.

2. Par exemple le nombre de flux audio nécessaire pourra varier selon le dispositif de restitution du terminal.

3. Si l'on prend une analogie avec les domaines graphiques, l'auteur va indiquer une couleur, le designer détermine la palette correspondante et propose des nuances, le programmeur écrit les fonctions permettant d'afficher cette couleur. Cependant, la description d'une couleur peut varier selon ces professions : ainsi l'auteur recherchera une couleur "chaude", le designer proposera un ensemble de nuances et le programmeur les codera sous forme binaire. Si un certain empirisme permet de dégager un consensus entre la notion de "chaud" et une gamme de couleur, ce problème reste plus flou dans le domaine audio, et ce d'autant plus qu'il est impossible de "désigner du doigt" une propriété auditive.

- Le langage doit posséder une hiérarchisation des données ; l'importance d'un élément est évaluée en fonction du taux d'information qu'il apporte. Le rendu de la scène peut ainsi être "scalable".
- Le langage doit pouvoir être éventuellement compris par un utilisateur expérimenté, et ce sans l'intermédiaire d'une interface. Par analogie, le langage Post-script, bien que très complexe peut néanmoins être lu sous forme de texte, et le cas échéant modifié.

Enfin, il peut être utile d'accompagner le langage d'un certain nombre de configurations de référence, permettant ainsi d'accélérer son apprentissage et d'offrir des repères communs à un ensemble de personnes.

3.2 Construction d'une scène sonore

Les contraintes proposées au paragraphe précédent laissent de nombreuses possibilités pour l'élaboration d'un langage de description d'une scène sonore. Ce paragraphe en décrit deux l'une basée sur une approche physique l'autre sur une approche perceptive. Bien que ces démarches sous entendent *a priori* l'utilisation d'une interface de manipulation adaptée, on essaiera ici de séparer ces deux notions.

3.2.1 Approche géométrique

L'approche géométrique décrit le système physique ayant engendré la scène sonore. Elle correspond à une approche visuelle, qui peut *a priori* renforcer son caractère intuitif. (2.3.2). Afin d'évaluer son adéquation avec les contraintes de construction exposées précédemment, on va imaginer un processus de construction incrémental d'une telle scène. Bien qu'il ne corresponde à aucune stratégie réelle de traitement de l'information, ce dernier permet néanmoins d'appréhender les qualités d'un tel langage.

Construction incrémentale d'une scène sonore

Les éléments du langage sont :

- les transducteurs : ils sont décrits par une position, une directivité, une orientation, des caractéristiques spectrales, une puissance pour la source,
- la salle, décrite par un ensemble de parois,
- les parois : ce sont des ensembles de points, en général coplanaires. Les propriétés acoustiques d'une paroi (absorption, diffusion) peuvent être décrites, soit par

des tableaux de coefficients pour différentes bandes de fréquences, soit par les coefficients d'un filtre.

Si l'on suppose une lecture successive de ces données à partir d'un fichier, il est possible de suivre l'évolution de la construction de la scène sonore. On place ainsi successivement :

- la source + le récepteur : cette situation représente la propagation en champ libre,
- les parois les unes après et les autres.

L'effet de salle est construit au fur et à mesure. L'information apportée par chaque paroi va dépendre de l'ordre d'ajout (*figure 3.1*) : certaines parois ne prendront de l'importance qu'au regard de la présence d'autres parois, et des effets auditifs peuvent non seulement apparaître mais aussi disparaître au cours de ce processus.

Remarques

Si l'on met cette construction en regard des critères d'évaluation précédents, il apparaît que son inconvénient majeur est l'absence de hiérarchie dans l'information apportée au moteur de rendu. Plusieurs effets en découlent :

- Dans une phase de composition d'une scène sonore, le résultat auditif ne sera révélateur de la situation finale qu'après avoir entièrement rentré les données du modèle. Une approche hiérarchique demanderait plutôt de partir d'une forme générale en précisant progressivement les détails de la géométrie. On n'est cependant pas assuré de la correspondance entre niveau de détail architectural et niveau de détail auditif.
- Si le logiciel de rendu est basé sur des algorithmes d'acoustique prévisionnelle, l'exactitude du rendu est directement dépendante de la puissance de calcul disponible, et de la complexité de la salle
- Si le logiciel utilise d'autres principes, il convient de proposer une stratégie de remplacement (*paragraphe 3.2.3*) afin d'effectuer un traitement plausible de l'information.

Exploitation

L'exploitation de ce type de représentation d'une scène sonore en vue de son auralisation fait appel à un ensemble d'algorithmes modélisant la propagation des sons dans l'espace. Le chapitre 6 y étant consacré. Nous n'aborderons donc pas ce point ici.

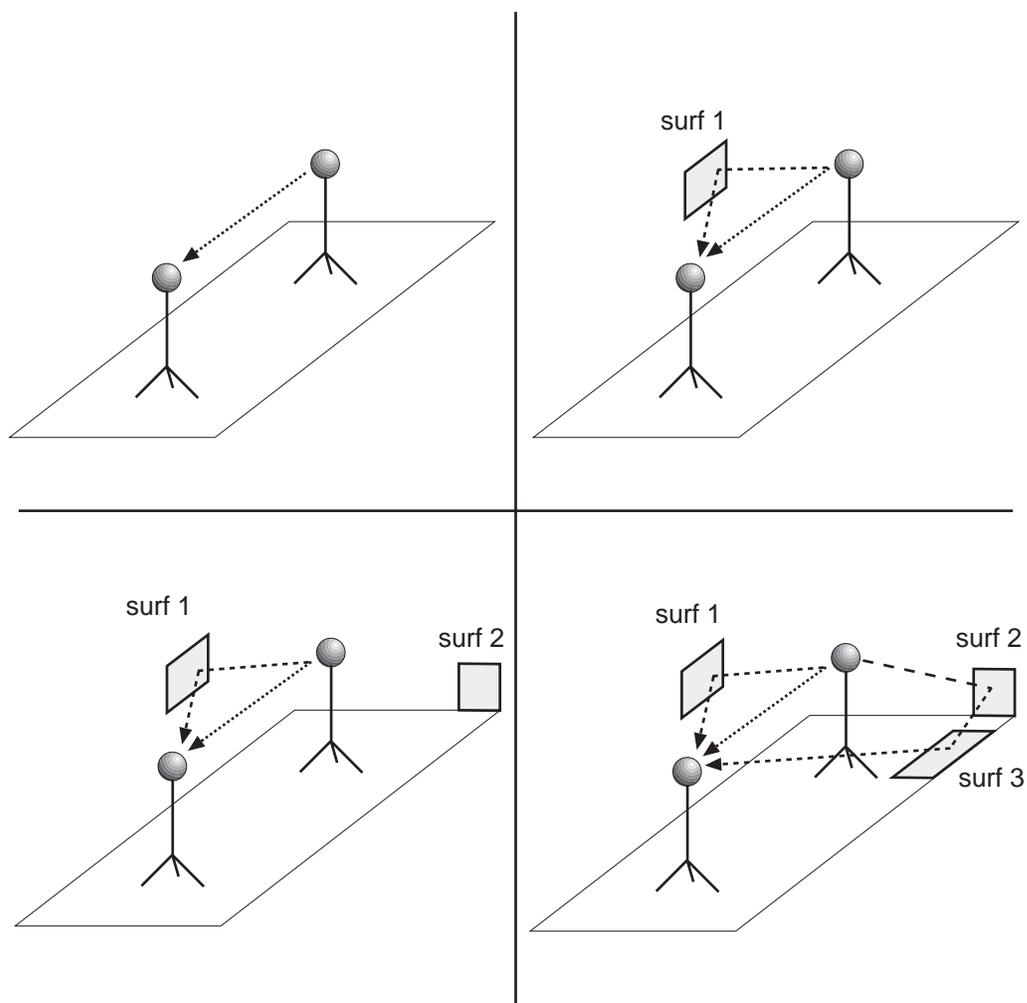


FIG. 3.1 – Ajout successif de parois : selon l'ordre d'adjonction, la contribution à l'effet de salle variera. La description ne sera complète que lorsque toutes les parois auront été placées. Par exemple, l'influence de la surface 2 ne se fera sentir qu'après le placement de la surface 3. De manière plus générale, de nouveaux phénomènes apparaissent au fur et à mesure que la salle est construite, pouvant même entraîner sur le plan perceptif la disparition de certains effets.

3.2.2 Approche perceptive

Face à la complexité des phénomènes physiques donnant lieu à l'effet de salle, une autre approche consiste à faire abstraction de ces propriétés et à se placer du point de vue de l'auditeur (on peut imaginer que celui-ci ferme les yeux). Ce travail a notamment été réalisé à l'IRCAM, dans le cadre d'études portant sur la caractérisation objective de l'effet de salle.

Mise en place

La mise en place d'une description perceptive de l'effet de salle n'a pu s'effectuer qu'à la suite d'une démarche systématique de caractérisation, démarche basée sur un ensemble de tests psychoacoustiques et une interprétation statistique des résultats de ces tests.

L'hypothèse de base est celle qu'un point dans l'espace perceptif est caractérisé par un ensemble de dimensions. A titre de comparaison, des recherches semblables sont menées afin de décrire un espace des timbres lié aux instruments de musique. Dans notre cas, les axes de cet espace représentent des sensations engendrées par un effet de salle. Ces dimensions sont supposées communes à l'ensemble des auditeurs et ne varient pas selon le type de message diffusé (parole, musique, ...) ⁴; elles forment de plus un système orthogonal. Ces dimensions sont également appelées facteurs perceptifs.

La détermination de ces facteurs, et de leur nombre, se fait en analysant les résultats de différents tests relatifs à la perception d'un effet de salle. Ces tests prennent la forme soit de tests structurés, réalisés en laboratoire, soit d'enquêtes effectuées auprès d'une série d'auditeurs en situation réelle. Les tests structurés sont construits de manière à faire varier un ensemble limité de dimensions perceptives, tandis que les enquêtes en situation réelle permettent d'évaluer une hiérarchie entre les différents aspects perceptifs ainsi que l'interaction avec d'autres paramètres tels que l'influence d'autres modalités sensorielles.

Sabine, au début du siècle, peut être considéré comme un précurseur de ce type de démarche ; il proposait un ensemble de trois critères : le niveau sonore, la distortion des sons complexes, et la confusion (réverbérance, échos). Il est intéressant de noter que ce dernier facteur est lié à un indice objectif mesurable, le temps de réverbération. Quant au fait que Sabine proposait un nombre de facteurs relativement réduit, il provient de l'hypothèse qu'il faisait de la présence d'un champ diffus dès le début de la réponse impulsionnelle. D'autres travaux suivirent à la fin des années 50, qui considèrent des aspects particuliers de la perception. Parmi ceux-ci, on peut noter l'étude sur l'effet de précedence ou encore les propriétés d'intégration de l'énergie des premières réflexions au son direct et leur effet sur la notion d'intelligibilité du discours ⁵.

Ce n'est qu'en 1962, avec les travaux de Beranek, qu'eut lieu une nouvelle approche générale de description la qualité acoustique d'une salle. Cette étude a eu lieu non à partir d'expériences en laboratoire, mais d'entretiens et d'impressions personnelles de l'auteur.

4. Les différences interindividuelles, et la notion de préférence, se traduisent par le fait que chaque auditeur affecte un poids particulier à chaque facteur, et ce selon sa sensibilité. Il ne s'agit pas ici de déterminer des critères d'appréciation de la qualité acoustique d'une salle, mais de décrire les dimensions qui gouvernent la perception de l'espace et amènent éventuellement à formuler un jugement. On peut également parler de tentative d'objectivation de la perception.

5. Cette notion, relative à la frontière entre l'énergie utile et l'énergie nuisible, a été relié à de nombreux critères dont la "Deutlichkeit", D et l'indice de clarté, $C80$.

Au total, un jeu de dix-huit facteurs non indépendants a été proposés. Ce nombre élevé tient au fait qu'aucune tentative de réduction de cet ensemble, à l'aide par exemple de méthodes statistiques, n'a été entreprise par Beranek. L'intérêt principal de l'étude de Beranek est de montrer que la donnée du temps de réverbération ne suffit pas à décrire la qualité acoustique d'une salle. Ses conclusions mettent en avant le rôle déterminant du temps d'arrivée de la première réflexion dans le jugement auditif. En revanche, dans son travail d'objectivation des attributs perceptifs (analyse de corrélations), il utilise conjointement des indices acoustiques mesurables à partir de la réponse de la salle et des indices architecturaux, dénotant une limite dans la volonté de s'affranchir de la description architecturale pour caractériser les effets auditifs.

Par la suite, différentes équipes se sont attachées à affiner la méthodologie utilisée dans les études perceptives pour s'affranchir des limites des études précédentes. On peut noter l'apparition de tests psycho-acoustiques effectués en laboratoire, à partir de stimuli enregistrés ou artificiels. Les avantages principaux de cette méthodologie sont d'offrir la possibilité de comparaisons instantanées afin de minimiser les biais liés à la mémoire auditive, à l'interprétation des oeuvres, aux inter-actions multi-modales ou cognitives (vision, réputation de la salle...). Ils ouvrent également la voie à des tests reposant sur des jugements non verbaux, limitant ainsi les problèmes d'interprétation sémantique des questionnaires. On citera ainsi les travaux du Groupe de Göttingen, ceux du Bell laboratory, ceux du groupe de Berlin, les thèse de Lehmann Wettschurek et les travaux de Baron et Marshall (importance des réflexions latérales, et de l'Early decay time). Ces recherches qui portaient généralement sur l'étude d'une composante particulière de l'effet de salle ont permis de prendre en compte l'importance de l'utilisation de méthodes d'analyse statistiques et les conditions spécifiques à l'élaboration de ces tests.

Les études plus récentes menées à Ircam [Lav89][Kah95] ont ceci de particulier qu'elles ont été conçues non seulement comme un outil de caractérisation, mais qu'elles ont débouché sur la mise en place d'un *opérateur de contrôle* de l'effet de salle, implanté dans le Spatialisateur. Un premier jeu de facteurs perceptifs a été issu des travaux menés au cours de la thèse de Catherine Lavandier [Lav89]. Le but de cette thèse était de vérifier si un ensemble d'indices habituellement utilisés en acoustique des salles était perceptivement pertinent. Au final un nombre de total de 14 facteurs perceptifs, pouvant être considérés comme une base de l'espace multidimensionnel de la perception a été identifié. Ces facteurs se répartissent en 4 groupes :

- facteurs relatifs à la distribution temporelle de l'énergie (énergies, temps de réverbération), au nombre de 5,

- facteurs relatifs aux effets de coloration liés à la distribution des premières réflexions, au nombre de 3,
- facteurs relatifs aux effets de coloration dûs aux valeurs du temps de réverbération et du niveau en fonction de la fréquence, au nombre de 4,
- facteurs relatifs à la spatialisation, au nombre de 2.

Par la suite ces données ont été reprises et interprétées par Jean-Pascal Jullien, au laboratoire d'acoustique des salles de l'Ircam, afin de réduire le nombre total de facteurs, tout en conservant une cohérence dans l'interprétation des tests. Le nombre total de facteurs, ainsi ramené à 10, a fait l'objet d'une validation *in situ* dans un ensemble de salles de concerts et d'opéra [Kah95] (*chapitre 2.3.2*).

Liens avec une représentation bas-niveau

On donne ici l'expression de 9 des facteurs perceptifs, retenus à l'issue des études précédentes, en fonction de la représentation morphologique de la réponse impulsionnelle de la salle. Les quantités initiales calculées à partir d'une réponse impulsionnelle sont des énergies et le temps de réverbération, R_t . On notera OD , l'énergie de la réponse sur la plage temporelle $[0, 20\text{ ms}]$, $R1$ celle sur la plage $[20\text{ ms}, 40\text{ ms}]$, $R2$ sur $[40\text{ ms}, 80\text{ ms}]$ et $R3$ sur $[80\text{ ms}, \infty]$. On notera de plus O_{XX} , l'énergie totale de $h(t)$ sur l'intervalle $[0, XX\text{ ms}]$ et R_{XX} celle sur la plage $[XX\text{ ms}, \infty]$

On indiquera par un suffixe X_{bf} , X_{mf} et X_{hf} la valeur de l'énergie de la quantité X en basses, moyennes et hautes fréquences. Les valeurs en basses fréquences sont typiquement évaluées sur les octaves 125 Hz et 250 Hz , tandis que les valeurs en hautes fréquences sont évaluées dans les octaves 4 kHz et 8 kHz .

Facteurs reliés à la source

Il s'agit des paramètres Es , niveau perçu ou "*présence*" de la source et des ses composantes fréquentielles $Desl$ ("*chaleur*") et $Desh$ ("*brillance*"). Bien que ces facteurs soient imputés à la source, leur calcul prend en compte l'intégration des premières réflexions. On calcule donc :

$$Es = 10 * \log_{10}(\max(O_{40}, R2) + 0.18\min(O_{40}, R2)) \quad (3.1)$$

Cette formule permet de prendre en compte l'intégration des réflexions précoces et l'intégration conditionnelle des réflexions tardives à la présence de la source.

Le calcul sur les variations fréquentielles liées à Es en basses et hautes fréquences s'exprime relativement à Es_{mf} par :

$$\begin{cases} DEsl &= (Es_{bf} - Es_{mf})/2 \\ DEsh &= (Es_{hf} - Es_{mf})/2 \end{cases} \quad (3.2)$$

Pour une source de référence ces deux facteurs traduisent donc plutôt l'effet de balance fréquentielle apporté par les premières réflexions et réflexions tardives sur le son direct.

Facteurs reliés à la salle

Il s'agit des facteurs concernant la perception de la salle. De manière similaire au calcul des paramètres liés à la source, ces facteurs prennent en compte les contributions de différentes sections de la réponse impulsionnelle traduisant incluant, le cas échéant, des effets de masquages associés. Ce sont l'énergie Rev niveau perçu du champ tardif ou "*présence de la salle*", et les facteurs liés à la décroissance du niveau du champ tardif, R_t et ses variations fréquentielles $Drtl$ ("*lourdeur*") et $Drth$ ("*vivacité*"). Bien que R_t soit égal au temps de réverbération, on préférera lui attribuer le nom de "*réverbérance*" dans ce cadre perceptif.

On a alors :

$$Rev = 10 * \log_{10}(Rr_{80-280} + R_{280}) \quad (3.3)$$

La séparation de $R3$ en deux sous-intervalles temporels, $[80 - 280]$ et $[280, \infty[$ permet de prendre en compte les effets de masquages dûs aux autres sections de la réponse impulsionnelle (son direct étendu). Rr_{80-280} représente en effet l'énergie restant dans cet intervalle temporel après application de ce masquage.

$$Rr_{80-280} = 10^{\log_2(2^{\log_{10}(R_{80-280})} - 2^{\log_{10}(0.18 * O_{40})})} \quad (3.4)$$

Les facteurs liés à la réverbérance sont donnés par :

$$\begin{cases} Rt &= Rt \\ Drtl &= (Rt_{bf} - Rt_{mf})/2 \\ Drth &= (Rt_{hf} - Rt_{mf})/2 \end{cases} \quad (3.5)$$

Facteurs intermédiaires

Il traduisent les effets spatiaux et temporels liés aux premières réflexions et surtout au cluster. Ce sont les facteurs Edt early decay time ou "*running reverberance*", relié à la réverbérance pendant le discours auralisé, et $Rd1$ traduisant la notion de largeur apparente de la source. On a :

$$EdT = (t[-15dB] - t[0dB]) * 4 \quad (3.6)$$

et

$$Rd1 = ANG1 - E_s \quad (3.7)$$

ou $ANG1$ est défini par

$$ANG1 = 10 * \log_{10}(R1 + 0.23 * R2) \quad (3.8)$$

Dans ce cas il est nécessaire d'effectuer une pondération spatiale lors du calcul de $R1$ et $R2$ en tenant compte de la direction d'arrivée de chacune des réflexions. Le tableau suivant indique les pondérations qui ont pu être établies à partir de l'analyse des tests perceptifs⁶.

Azimut	0	50	130	180	<i>diffus</i>
R1	0.01	0.33	0.28	–	1.1
R2	0.01	0.33	0.28	0	1.2

(3.9)

Opérateur perceptif

En pratique, une interface permettant de décrire et contrôler une scène sonore sous le formalisme perceptif doit fournir au moteur de rendu les valeurs des paramètres de la représentation morphologique de la réponse de la salle. Ces valeurs sont obtenues par un processus d'inversion des relations présentées dans le paragraphe précédent. Ces expressions ainsi que le procédé d'inversion global utilisé avec le moteur de rendu *Spatialisateur* sont indiqués en annexe B.

3.2.3 Approche statistique: stratégie de remplacement d'une interface géométrique

Tout en respectant l'approche géométrique, il est possible d'utiliser les résultats d'une approche statistique de la propagation des sons dans les salles (comme la théorie de Sabine) pour s'adapter à différentes conceptions de logiciel et/ou à des puissances de calcul limitées. Cette approche a pour but de traduire les propriétés principales d'une salle en fonction de grandeurs architecturales globales telle que le volume et la quantité d'absorption ou en fonction des déplacements des sources et récepteurs.

Par essence, l'exploitation de ce modèle s'appuie principalement sur des formulations analytiques reliant des paramètres architecturaux et indices acoustiques, et exclut *a priori* l'utilisation de modèles géométriques de la propagation acoustique. De ce fait la représentation de l'espace sur l'interface peut être de nature symbolique, la géométrie

6. Le signe "-" signifie que cette valeur est restée inaccessible, ce cas ne s'étant pas présenté dans les tests.

exacte de la salle n'étant pas exploitée. D'une certaine façon, cette approche pourrait correspondre à l'idée que l'on peut se faire de l'acoustique d'une salle en y entrant⁷.

Ce type de modèle permet principalement de contrôler un son direct et une réverbération. La création de sections intermédiaires ne peut se faire qu'en imposant une part d'arbitraire dans leur description.

La caractérisation du son direct peut obéir à une loi géométrique, donnant son atténuation en fonction de la distance source-récepteur. La prise en compte des directivités des transducteurs et des phénomènes d'absorption du milieu se traduisent par des effets spectraux :

$$E_{dir}(f)(r) = \beta(\theta_r, \phi_r, f) \Gamma(\theta_s, \phi_s, f) \frac{P_{source} e^{-m(f)r}}{4\pi r^2} \quad (3.10)$$

où $\Gamma(\theta_s, \phi_s, f)$ représente la fonction de directivité de la source, P_{source} la puissance de la source, $m(f)$ le coefficient d'absorption atmosphérique par mètre et $\beta(\theta_r, \phi_r, f)$ la sensibilité du microphone.

En utilisant les propriétés statistiques du champ réverbéré exposées au chapitre 2, un groupe de 3 données permet de caractériser cette réverbération : le temps de réverbération, le niveau de la réverbération et le temps de mélange (*chapitre 2.2.1*). Ces grandeurs peuvent être fournies directement par l'utilisateur ou déduite de la données du volume et de la quantité d'absorption en utilisant la formulation de Sabine ou ses dérivées. L'utilisation directe de ces grandeurs pose cependant le problème de leur non indépendance sur le plan perceptif (le contrôle du volume ou de la quantité d'absorption agissent simultanément sur plusieurs facteurs perceptifs).

A partir de l'approche statistique on peut également déduire un modèle de dépendance de l'énergie des premières réflexions et du champ réverbéré en fonction de la distance source/récepteur tel que décrit par Barron [BL88] .

$$E_{ref}(\tau)(r) = \frac{T_r(f)}{13.81} \frac{\Gamma_d(f)^2 \beta_d(f)^2}{V} e^{\frac{-13.81(\tau + \frac{r}{c})}{T_r(f)}} \quad (3.11)$$

$$E_{rev}(\tau)(r) = \Gamma_d(f)^2 \beta_d(f)^2 \frac{e^{\frac{-13.81(\tau + \frac{r}{c})}{T_r(f)}}}{4\pi c^3 (\tau + \frac{r}{c})^2} \quad (3.12)$$

Dans ces expressions $E_{ref}(\tau)(r)$ représente le niveau moyen d'une réflexion parvenant au temps τ à un récepteur situé à une distance r de la source, tandis que $E_{rev}(\tau)(r)$

7. On peut convenir, par exemple, de représenter la salle par une zone de l'écran dont la taille est fonction du volume supposé de la salle, et au sein de laquelle évoluent des transducteurs.

correspond au niveau du champ réverbéré après le temps τ et observé par ce même récepteur. Ces formulations font intervenir les fonctions de transfert en champ diffus de la source $\Gamma_d(f)^2$ et du récepteur $\beta_d(f)^2$.

Intégration d'éléments hétérogènes

La description peut, en outre, s'agrémenter d'éléments architecturaux ou géométriques simples, dont l'effet est supposé important. Par exemple, la description d'une frontière à l'environnement, et de parois strictement internes à cette frontière permet de représenter des situations d'obstruction et d'occlusion⁸. Elle symbolise les cas où une paroi se trouve entre une source et un récepteur et celui où une source se trouve hors de l'espace considéré. Dans ces cas, le son direct est annulé ou filtré, et la réverbération éventuellement modifiée.

La donnée de ces parois permet également de rajouter explicitement des premières réflexions, venant compléter la description de l'effet de salle. L'effet de salle est ainsi constitué de deux contributions indépendantes : celles provenant du volume et celles provenant des parois. Cette division ne correspond cependant pas à une modélisation réelle de l'effet de salle, dans la mesure où les éléments de parois internes ne sont pas considérés pour la prédiction du champ réverbéré tardif, uniquement estimé à partir de l'enveloppe. On parlera dans ce cas d'une approche physique symbolique. La modélisation de chacune de ces contributions est gérée de manière indépendante :

- celles provenant des parois internes peuvent être calculées de manière explicite (direction, amplitude, temps d'arrivée) à partir d'un modèle sources-images ou rayon limité à l'ordre 1. Les parois seront décrites par un coefficient d'absorption.
- Celles provenant de la salle sont estimées à partir du modèle statistique présenté précédemment.

Par ailleurs, on peut également faire appel à certains résultats de la psychoacoustique, notamment dans la prise en compte du rôle des premières réflexions dans la perception du niveau lié à la source ou des effets spatiaux (largeur apparente de la source, impression d'espace). On peut ainsi privilégier l'exploitation des contributions des parois situées dans un angle solide voisin de celui de la source ou dans les angles solides latéraux par rapport à l'auditeur.

Enfin afin d'orienter un concepteur dans sa démarche, la donnée d'exemple correspondant à des cas réels "types" peut s'avérer utile.

8. Ces concepts ont été définis par Creative Labs, EMU pour la version de l'environnement EAX 2.0 (*chapitre 4*).

3.2.4 Retour sur la directivité des sources

Influence de la directivité

On a vu au chapitre 3, que l'influence de la directivité des transducteurs pouvait se traduire de différentes manières sur l'effet de salle : sous la forme que l'on appellera "*directe*", elle se traduit par une modification de l'amplitude de l'onde émise (ou reçue) selon sa direction et selon la fréquence. Elle prend par conséquent la forme d'une suite de couples "atténuation-direction", et peut être exploitée pour le calcul du son direct, des contributions de source images, ou de l'énergie véhiculée par un rayon.

A cela peut être ajoutée une caractérisation, dite "omni", qui synthétise sous la forme d'une fonction de transfert, l'effet des directivités "directe" sur l'ensemble des directions. Son spectre d'amplitude est équivalent à celui de la puissance rayonnée par la source. Intuitivement, la salle effectue un mélange des contributions émises par la source dans différentes directions, et donc moyenne les effets de celle-ci. La courbe caractéristique de cette transformation est appelée "fonction de transfert en champ diffus" de la source (ou récepteur). Son estimation peut donc être calculée par moyennage spatial de toutes les valeurs de directivités fournies pour différentes directions à condition que celles-ci effectuent un balisage régulier de l'espace. Dans le cas contraire, il faut pondérer chaque direction par l'angle solide qu'elle représente.

$$\Gamma_d(f) = \sum_i \frac{\Gamma_{\theta_i}(f) d\Omega_i}{4\pi} \quad (3.13)$$

On verra au chapitre 5 comment cette fonction de transfert en champ diffus peut être mesurée directement à partir de l'analyse du champ réverbéré, excité par cette source.

La donnée de ces deux fonctions ne s'avère nécessaire que dans certains cas. Lorsque l'effet de salle est décrit selon un modèle géométrique, seule la composante "*directe*" est nécessaire. Ce sont les algorithmes utilisés qui véhiculent ces informations et "fabriqueront" naturellement une loi de mélange correspondant à un spectre "*omni*" sous jacent. Lorsque la démarche choisie est perceptive ou statistique la donnée de ces deux types de caractérisation permet de calculer d'une part les effets liés à la source (son directe et éventuellement réflexion tardive) grâce aux composantes "*directe*", d'autre part le contenu spectral du champ réverbéré grâce à la composante "*omni*".

Si les données "*directe*" sont fournies avec une discrétisation spatiale fine, la donnée du spectre "*omni*" s'avère redondante puisqu'elle peut être recalculée lors d'une étape d'initialisation. En revanche on peut supposer que dans certains cas, les données manquent pour des sources particulières (instruments, objets sonores...) et qu'on ne dispose que de quelques données ou d'un modèle approché. La donnée du spectre "*omni*" permet alors d'assurer la cohérence du champ tardif.

Types de description

Modèle géométrique paramétrique

L'examen de bibliothèques de programmation (API) consacrées au traitement du son 3D et de l'effet de salle, montre qu'un modèle couramment utilisé pour décrire les propriétés de directivité est celui d'une description géométrique, de type "*directe*" uniquement. La fonction de directivité est décrite par deux angles correspondant à des angles d'ouverture de deux cônes, ayant le transducteur pour sommet. Si on appelle α_1 et α_2 ($\alpha_1 < \alpha_2$) ces deux angles, le niveau du signal émis est maximal à l'intérieur du premier cône, et minimal (ou nul) à l'extérieur du deuxième cône. Pour la zone située entre les deux cônes, ce niveau est calculé par interpolation entre les deux valeurs aux limites. La loi d'interpolation peut varier (linéaire, logarithmique, ...).

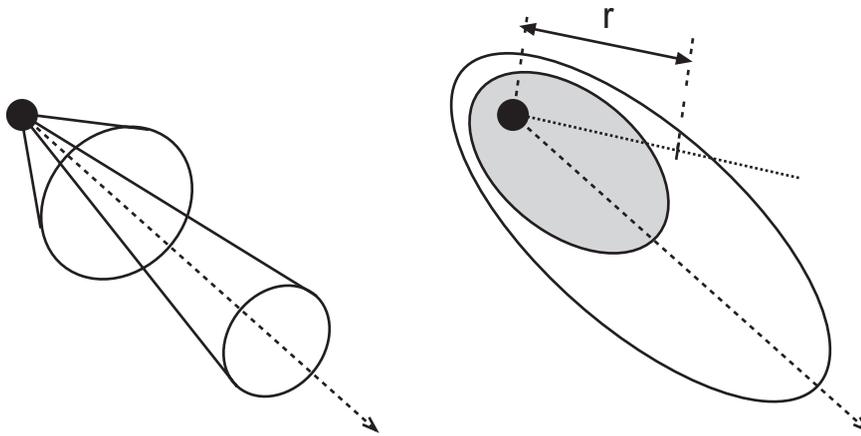


FIG. 3.2 – Description par 2 cônes, ou 2 ellipsoïdes. Lors d'une description par deux ellipsoïdes, l'atténuation est calculé en fonction de la distance r entre la source et le récepteur. Les surfaces d'isoatténuation sont également des ellipsoïdes.

Le choix de la géométrie utilisée peut varier. Ainsi le langage MPEG4, dans sa première version, utilise des ellipsoïdes, le transducteur se trouvant placé au lieu d'un des foyers des ellipsoïdes. Ce modèle est effectivement plus réaliste dans la mesure où il permet de considérer toutes les directions (et notamment les arrières).

Ces modèles négligent en général toute influence des variations fréquentielles sur la directivité, mais une extension à différentes bandes de fréquence est aisément envisageable. Ils sont de plus adaptés uniquement à une description géométrique de l'effet de salle.

Une autre approche peut utiliser des figures de directivité caractérisables mathématiquement par un petit nombre de paramètres tels que les angles d'ouverture ou l'indice de directivité. On donnera généralement une ou plusieurs valeurs d'angle permettant d'ajuster l'ouverture du transducteur, ainsi qu'une atténuation à X décibels pour cette

ouverture. Le comportement fréquentiel peut être décrit de manière analogue dans différentes bandes de fréquence. Bien qu'elles puissent être déduites des caractéristiques précédentes la valeurs de l'indice de directivité ou du spectre "*omni*" peuvent être fournies parallèlement à cette description).

On notera cependant qu'un tel modèle suppose la présence d'une direction de référence pour la directivité du transducteur, ce qui n'est pas le cas de la plupart des sources réelles, et notamment des instruments de musique. Néanmoins il peut constituer une bonne approximation des sources sonores telles que les haut-parleurs ou la voix parlée ou chantée.

Echantillonnage spatial

Une description physique précise oblige à effectuer une discrétisation de l'espace selon toutes les directions (azimuth et élévation), et à regrouper les valeurs d'amplification dans une table. La valeur de la directivité "*directe*" pour un triplet ($Az.$, $El.$, f) est calculée par interpolation sur ces 3 composantes. Le spectre "*omni*" peut être calculé à partir de sa définition, en temps que moyenne sur l'ensemble des directions.

Une autre manière de modéliser la dépendance fréquentielle des caractéristiques de directivité, est de donner pour chaque couple $Az.$, $El.$, les coefficients d'un filtre, l'ordre du filtre dépendant de la précision exigée, et pouvant éventuellement varier avec la direction considérée. Là encore, il faut noter que cette description semble mieux adaptée à une description géométrique de l'effet de salle. Il est peut-être intéressant de rajouter à cette description un filtre supplémentaire "*omni*" qui serait utilisé pour le traitement du champ tardif⁹.

On remarquera cependant que, si cette modélisation augmente considérablement la précision apportée à une description de la directivité, elle semble plutôt correspondre à des applications dans le domaine de l'acoustique, et peut être trop détaillée, voire compliquée, pour d'autres applications (par exemple, pour un logiciel de sound design).

Analogie avec les sources lumineuses

La remarque précédente suggère d'utiliser une description des sources par analogie avec des sources lumineuses, de façon à autoriser une manipulation plus intuitive dans les contextes de création.

Tout d'abord, tous les modèles décrits précédemment concernent généralement une source ponctuelle. Si l'on considère une analogie avec la description de sources lumi-

9. Si l'on travaille selon une modalité perceptive ou statistique, la validité d'un filtre "*omni*" calculé par simple moyennage des coefficients de ces filtres est nulle. Il faut calculer les fonctions de transfert de chaque filtre et recalculer un filtre à partir de la moyenne des spectres d'amplitude.

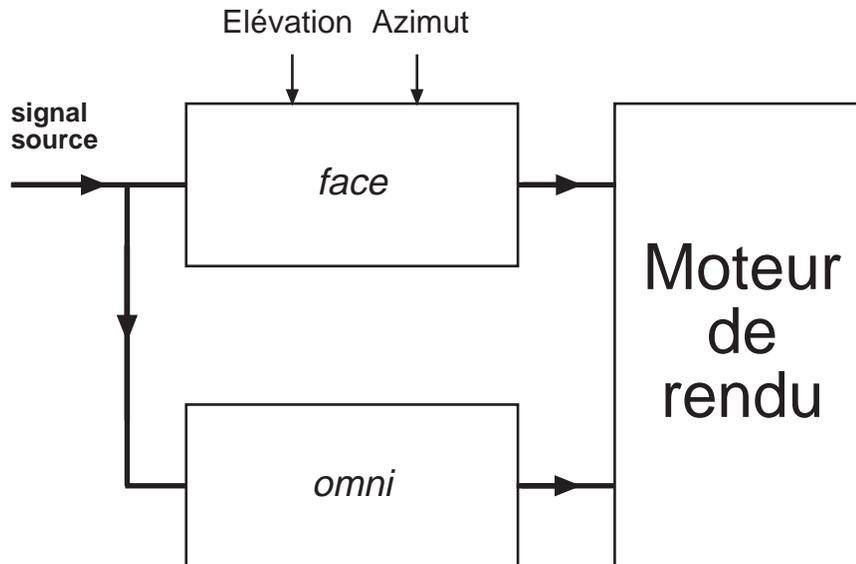


FIG. 3.3 –

neuses en programmation graphique, il est possible d'introduire un autre classement :

- les sources ambiantes, qui sont perçues de manière diffuse ; typiquement, transposées dans le domaine sonore, elles n'alimenteront que la partie tardive de l'effet de salle,
- les sources directionnelles, qui sont perçues dans une direction mais ne sont pas localisables en distance¹⁰. Dans le domaine sonore on pourrait faire rentrer dans cette catégorie le bruit lointain d'une autoroute.
- les source ponctuelles, correspondantes aux sources décrites précédemment

L'utilisation d'une telle classification requiert cependant l'usage implicite de règles pour leur rendu. Par exemple, si une source est décrite comme ambiante, le canal acoustique correspondant devra avoir des valeurs correspondantes de OD et $R1$ nulles. Le réglage des caractéristiques fréquentielles se fera à travers les caractéristiques de sa directivité "omni".

3.3 Examen de langages existant

Le développement de l'Internet et technologies de télécommunication a donné lieu à des recherches portant sur la description de monde virtuels, pouvant être transmis sur un

¹⁰. Dans le domaine de la lumière le meilleur exemple est le soleil.

réseau et interprété, de manière interactive, sur l'ordinateur personnel de l'utilisateur. Ces langages permettent de décrire plusieurs types de données (images, monde 3D, sons, scripts d'interactivité) et ce en adoptant des formats de compression divers. On s'intéressera ici à la composante audio, et particulièrement audio 3D, de deux langages existant.

3.3.1 Le langage VRML 97

VRML (Virtual Reality Modeling Language) est à la base d'un format de fichier autorisant la description de mondes virtuels interactifs en trois dimensions. Il peut contenir différents types de données : géométries 3D, données MIDI, images JPEG, scripts. Son organisation est celle d'un format de fichiers hiérarchique, orienté-objet et extensible. Les fichiers sont de type texte et donc compréhensibles par un être humain.

Organisation des données

Chaque objet élémentaire est appelé *nœud* (*node*) et contient différentes données :

- un *type* (*type name*) : il permet d'identifier la nature du nœud considéré
- des *champs* (*fields*) : ils représentent les attributs du nœud. Deux nœuds de même type diffèrent par leurs valeurs pour un même champ. Certains champs sont privés (*field*), c'est-à-dire qu'ils reçoivent une valeur initiale qui ne peut être modifiée et reste inaccessible, d'autres sont publiques (*exposedField*), et reçoivent une valeur initiale qui peut être modifiée.
- des *événements* (*events*) : ils peuvent être reçus ou envoyés par le nœud. La réception d'événements permet de modifier la valeur des champs, tandis que leur émission signale qu'un champ a subi une modification de sa valeur.

Les nœuds sont regroupés sous la forme d'un arbre avec des *parents* (*grouping nodes*) et enfants *children node*. Les nœuds présents au plus bas niveau de l'arbre sont appelés *feuilles* (*leaves*). Il est ainsi possible d'appliquer des transformations à un groupe qui seront automatiquement répercutées aux enfants.

Description du nœud Sound

Il s'agit du nœud décrivant une source sonore dans l'espace virtuel décrit par le fichier VRML. Sa structure est la suivante :

Sound {			
<code>exposedField</code>	<code>SFVec3f</code>	<code>direction</code>	100
<code>exposedField</code>	<code>SFFloat</code>	<code>intensity</code>	0.8
<code>exposedField</code>	<code>SFVec3f</code>	<code>location</code>	234
<code>exposedField</code>	<code>SFFloat</code>	<code>maxBack</code>	5
<code>exposedField</code>	<code>SFFloat</code>	<code>maxFront</code>	6
<code>exposedField</code>	<code>SFFloat</code>	<code>minBack</code>	1
<code>exposedField</code>	<code>SFFloat</code>	<code>minFront</code>	1
<code>exposedField</code>	<code>SFFloat</code>	<code>priority</code>	0
<code>exposedField</code>	<code>SFNode</code>	<code>source</code>	NULL
<code>exposedField</code>	<code>SFBool</code>	<code>spatialize</code>	TRUE
}			

La dénomination des trois premiers champs est explicite et n'appelle aucun commentaire particulier. Le champ **intensity** peut prendre des valeurs entre 0 (minimum) et 1 (maximum). Le champ **source** spécifie l'origine du son devant être traité par le système (soit un nœud de type `AudioClip` représentant un fichier audio ou un fichier MIDI, soit un nœud de type `MovieTexture`, représentant la partie audio d'un film codé au format MPEG-1).

Les champs **maxBack**, **maxFront**, **minBack**, **minFront** décrivent la directivité de la source sous la forme de deux ellipsoïdes (*chapitre 3.2.4*), la source représentant un des foyers de ces géométries. Le modèle d'interpolation conseillé par le consortium VRML est celui d'une atténuation en $1/r^2$ entre les deux ellipsoïdes (r étant la distance mesurée entre les 2 quadriques selon l'axe source récepteur, figure 3.2). Le champ **priority** permet à l'auteur du fichier d'imposer au browser VRML un ordre de priorité sur l'auralisation de différents nœuds de type `Sound`, notamment lorsque les capacités de celui-ci sont limitées.

3.3.2 La norme MPEG4

Présentation générale

MPEG (Moving Expert Picture Group) est un standard ISO consacré à la représentation de données audio et vidéo, généralement sous forme compressée. Pour la composante audio, les codages utilisés sont de types perceptifs ; le décodage doit restituer un signal qui est perçu de manière identique par l'auditeur, mais n'est pas forcément identique au signal original.

Au niveau du traitement des composantes audio, les différentes versions (*phases*) de

MPEG présentent des fonctionnalités différentes :

- MPEG-1 considère des sons mono et stéréo, à des fréquences d'échantillonnage utilisées en audio Hi-Fi (32kHz, 44.1kHz, 48, kHz).
- MPEG-2 s'intéresse à différentes extensions : codage à des fréquences d'échantillonnage plus basses, extension à des systèmes multicanaux.
- MPEG-4 s'intéresse à l'encodage de deux types de sons¹¹ : les sons *naturels*, et les sons *synthétiques*. Ces derniers comprennent des sons représentés sous d'autres formes, tels par exemple le MIDI, les systèmes de synthèse vocale à partir d'un texte (Text To Speech), l'encodage sous forme de synthèse additive. De plus des outils pour la localisation 3D de sources sont considérés, permettant ainsi la création d'environnements virtuels complets.

Flots de données

Le langage utilisé par MPEG-4 pour décrire une scène virtuelle, est, contrairement à VRML, un langage binaire : le *Binary Format for Scene Description* (BIFS). Ce langage comprend également un ensemble de nœuds permettant de décrire une scène sonore virtuelle.

3.3.3 MPEG4 Version 1 : la partie audio 3D

La première version de MPEG-4, dont les spécifications ont été arrêtées en Octobre 1998, considérait un objet audio pour le codage des sources en trois dimensions, de manière semblable à la norme VRML97. La version 2 de MPEG-4, dont les spécifications doivent être gelées 1 an après celles de la version 1, permet d'intégrer des BIFS supplémentaires. Différents objets BIFS sont proposés pour l'audio dans la Version 1:

- `AudioBuffer`
- `AudioClip`: permet l'insertion d'un clip audio dans une scène
- `AudioDelay`: permet d'ajouter un retard à la présentation d'un son
- `AudioMix`: objet de mixage
- `AudioSource`: définit une source audio dans une scène
- `AudioFX`: applique des effets de postproduction à un son
- `AudioSwitch`: permet d'allumer ou d'éteindre une source audio dans une scène

11. Le but de la norme MPEG-4 en général est celle de codages basés sur le contenu

- `ListeningPoint` : définit un point d'écoute dans une scène virtuelle
- `Sound`, `Sound2D` : définit les propriétés de rendu tridimensionnel d'un son. L'objet `Sound` est identique à celui présent dans VRML97, la différence majeure étant le type de son traité. Ceux-ci sont décodés en temps réel et peuvent regrouper des flux audio d'origines différentes. L'objet `Sound2D` est utilisé dans le cas de scène bidimensionnelle. Il s'agit d'une version simplifiée de `Sound`, mais avec une spatialisation définie uniquement dans un plan.

3.3.4 Bilan des propositions pour la Version 2

Objets proposés par le laboratoire d'Helsinki

Le laboratoire d'acoustique et de traitement du signal audio de l'université d'Helsinki a proposé des BIFS complémentaires (Advanced AudioBIFS), pour la version 2 de la norme MPEG-4 [VH98].

Modélisation de la source

Ce format, `DirectiveSound`, est relativement proche de l'objet `Sound`, mais définit la directivité de la source de manière différente. A un ensemble d'angles, relatifs à la direction de la source, sont associés des fonctions de transfert (par exemple filtre passe-bas d'ordre 1). Pour une valeur d'angle comprise entre deux de ces valeurs, une interpolation est effectuée. Dans le cas où l'ordre des filtres est nul, la directivité est réduite à un simple gain, variant avec l'angle de la source (*chapitre 3.2.4*).

La présence d'un champ `spatialize`, pouvant prendre quatre valeurs différentes, permet d'appliquer différents types de rendu sonore : de manière similaire à VRML, les deux premières valeurs possibles permettent d'activer ou non une spatialisation simple. Les deux valeurs suivantes permettent d'activer ou non le rendu de l'effet de salle : dans un cas celui-ci est effectué sans tenir compte de la localisation de la source, dans l'autre en en tenant compte.

Modélisation de la propagation

Il s'agit de champs présents dans le nœud `DirectiveSound`, permettant de contrôler les propriétés du milieu de propagation, soit le degré d'absorption de l'air, ainsi que la célérité du son dans le milieu. La présence d'un champ boolean, permet d'activer ou non le rendu de l'absorption de l'air.

Modélisation de l'environnement acoustique

MPEG-4, de même que VRML97, comprend un nœud appelé `IndexedFaceSet` permettant de définir un polygone. Ce nœud est dérivé, de façon à y introduire des proprié-

tés acoustiques, définissant ainsi un nœud, `AcousticMaterial`. La donnée d'un ensemble de nœuds de ce type, permet de calculer un ensemble de réflexions dans une pièce, en utilisant par exemple une méthode des sources-images.

Enfin un nœud, `AcousticScene` permet de décrire plusieurs environnements acoustiques différents. Chacun de ces nœuds décrit une boîte virtuelle à l'intérieur de laquelle est appliquée une réverbération tardive, et permet de regrouper différentes surfaces. Ce nœud possède, en outre, un temps de réverbération permettant ainsi de s'adapter à des environnements ne possédant pas la puissance de calcul nécessaire à un traitement géométrique complet.

Objets proposés par le CNET-France Télécom

Deux types d'objets ont été proposés. Ils sont basés sur la qualification objective et la perception de l'effet de salle ; ce sont les nœuds `PerceptualScene` et `PerceptualSound`. Leur composantes sont directement issues des recherches effectuées à l'IRCAM (*chapitre 2*). La seule information géométrique concerne la distance et l'orientation de la source par rapport au récepteur [REWJ98], [JR98].

PerceptiveSound

Ce nœud représente une source sonore dans un environnement virtuel. Il comprend des champs relatifs à la perception de la source (presence, warmth, brilliance), ainsi que la description de sa directivité. Cette dernière comprend les composantes "*directe*" et "*omni*". De plus un certain nombre de champs permettent de décrire certaines propriétés physiques utiles pour le rendu de l'onde directe (vitesse de son, absorption de l'air, direction)

PerceptiveScene

Ce nœud contient les paramètres perceptifs associés à la salle, ainsi que ceux dépendant de la position des sources mais perceptuellement attribués à la salle (*room presence, running reverberance, envelopment*). Ces trois derniers paramètres sont définis pour un ensemble de points de références, et mis à jour au fur et à mesure des déplacements de l'utilisateur.

Discussion

On retrouve à nouveau l'opposition entre une description géométrique (avantage du visuel, seuls garants du réalisme de la situation reproduite mais nécessitant des algorithmes coûteux, peu intuitifs lors d'une phase d'élaboration de monde virtuel) et une description perceptive (nécessité d'un temps d'apprentissage, concept éloigné de

certaines manipulations physiques, approche intuitive). Ces deux concepts ont été développés au début de ce chapitre.

Si l'on se réfère au premier paragraphe de ce chapitre, il semblerait que l'adjonction d'un modèle statistique conjointement au modèle géométrique, permettrait de rendre cette interface plus manipulable, et ce principalement lors d'une phase d'élaboration de monde virtuelle.

Chapitre 4

Etude de différents moteurs de rendu

Les chapitres précédents ont permis de définir les principales modalités pour des descriptions bas et haut niveau d'une scène sonore virtuelle. Afin de pouvoir construire des environnements basés sur le modèle du schéma synoptique du chapitre 1, et permettant le lien avec différents types d'interface, il est nécessaire de pouvoir accéder, directement ou non à ce contrôle bas-niveau.

Ce chapitre présente donc plusieurs moteurs de rendu, et bibliothèques consacrées au traitement du son 3D, et examine l'organisation de leurs paramètres de contrôle. Une attention plus particulière est portée à l'environnement Spatialisateur développé à l'IRCAM. Enfin, l'examen de deux bibliothèques présentes dans l'industrie, EAX de Creative Labs et Sound Sprockets d'Apple Computer, sera effectuée.

4.1 Spatialisateur de l'Ircam

4.1.1 Description de l'environnement MAX

L'environnement de programmation visuel MAX/FTS a été créé à l'IRCAM par Miller Puckette en 1990. Afin de permettre aux compositeurs de musique contemporaine de mettre en œuvre rapidement des algorithmes musicaux, cet environnement présente une interface composée de boîtes (les objets de traitement) reliées par des fils, symbolisant un système de communication par messages. Une unité fonctionnelle est représentée par un patcher, pouvant lui-même comprendre des sous-patchers.

Dédié à l'origine aux traitements des événements MIDI et existant sur le système opérationnel MacOS, MAX s'est par la suite enrichi d'un ensemble de modules consacrés au traitement du signal temps-réel, FTS. Cet environnement fonctionnait alors sur des stations de travail NeXT, équipées de cartes de traitement du signal.

L'évolution actuelle des capacités de calcul des microprocesseurs a permis de s'affranchir de la présence d'une carte externe, pour les opérations de traitement du signal. L'environnement actuel fonctionne sur deux types de systèmes opérationnels :

- les systèmes de type UNIX : il s'agit de stations de travail Silicon Graphics ou d'ordinateurs compatibles IBM PC équipé du système opérationnel Linux. Afin de présenter une portabilité accrue, le noyau de traitement du signal est écrit en C ANSI, tandis que la partie interface utilisateur utilise le langage Java. L'ensemble ainsi obtenu est dénommé jMax.
- le système MacOS : écrit entièrement en C, et faisant appel à la Toolbox de programmation proposé par Apple, cet environnement est commercialisé conjointement par les sociétés Opcode et Cycling74. La partie traitement du signal ne fonctionne sur les ordinateurs Macintosh équipés de processeurs PowerPC ; l'ensemble ainsi obtenu est dénommé Max/MSP.

L'orientation objet, ainsi que la convivialité de son interface utilisateur, font de Max un langage idéal pour le prototypage et l'implémentation rapide de nouveaux concepts. De plus l'architecture de ce système autorise l'écriture de modules externes réutilisables, en Max ou en langage C.

4.1.2 Description du Spatialisateur

Le Spatialisateur est un moteur de rendu sonore tridimensionnel et d'acoustique des salles entièrement écrit en Max. Il s'agit d'un projet développé par l'IRCAM et la société de design sonore Espaces Nouveaux, et démarré en 1991. La conception et la réalisation ont été confiées à Jean-Marc Jot, au sein de l'équipe acoustique des salles de l'IRCAM.

Son principe est celui du synthétiseur d'effet de salle. La réverbération est synthétisée à partir de matrices de retards rebouclés, les premières réflexions par unités de retard/gain. Aucun traitement par convolution n'est effectué. Une présentation plus détaillée de l'architecture et des performances du Spatialisateur est effectuée en annexe B.

Afin de s'adapter à un grand nombre de situations, l'environnement Spat se présente sous forme de modules. Les quatre modules principaux sont :

- Le module Source^{~1} : effectue le prétraitement de la source (directivités pour les caractéristiques direct et de champ diffus), et permet de rajouter l'absorption de l'air, ainsi que l'effet Doppler. Deux signaux sont issus de ce module, un signal directionnel et un signal omnidirectionnel s'adressant à la partie diffuse de la salle.
- Le module Room[~] : effectue le traitement du son direct, ainsi que l'application de l'effet de salle. La réalisation de ce module tient compte du découpage en quatre sections temporelles de la réponse impulsionnelle. Il comprend 7 sorties : une pour le son direct, deux pour les premières réflexions, et quatre pour les signaux relatifs au cluster et à la réverbération.
- Le module Pan[~] : permet d'appliquer une localisation au son direct (azimut, elevation) et au groupe des premières réflexions. Les réflexions tardive et la réverbération peuvent faire intervenir un contrôle spatial (IACC). Ces traitements sont effectués selon différentes techniques d'encodage et s'adressent à différents systèmes de reproduction : stéréophonie, techniques binaurales, Ambisonics (format B), panoramique d'intensité, techniques Surround.

1. Par convention dans l'environnement Max, les objets relatifs au traitement du signal sont suivis du sigle [~]i.e. Spat[~]. En outre l'environnement Spatialisateur propose un ensemble d'objets de contrôle ; ils sont suivis du signe _~ i.e. Spat_~

- Le module Room~ : permet l'adaptation des signaux encodés à différents systèmes de restitution. Par exemple dans le cas de l'utilisation d'un module Pan~ pour l'encodage vers le format B (Ambisonics), l'utilisation d'un module Out~ approprié permet de décoder ce format pour des systèmes composés de 4 à 8 haut-parleurs.

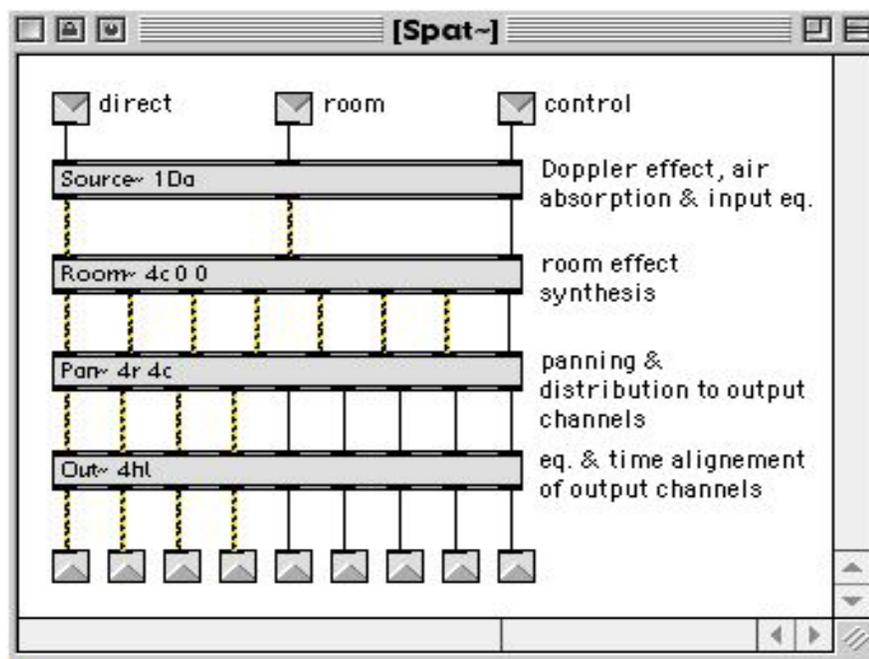


FIG. 4.1 – Modules du Spatialisateur

Ces modules peuvent être regroupés dans un seul module le module Spat~.

4.1.3 Scalabilité

L'écriture sous forme de modules assure déjà une adaptabilité à différentes situations. A cela vient s'ajouter le fait que chaque module est décliné suivant différents ordres de complexité afin de s'adapter à la puissance de calcul disponible. Ces options se présentent sous différentes formes:

- Activation ou non d'options permettant d'augmenter le réalisme du processus d'auralisation, par exemple, prise en compte de l'effet Doppler, de l'absorption de l'air ...
- Niveau de finesse des algorithmes utilisés pour la synthèse de l'effet de salle ; le module Room~ peut fonctionner selon plusieurs modes : prise en compte ou non des différentes sections temporelles du modèle de réponse impulsionnelle (direct, early,

cluster, reverb), différents niveaux de finesse dans la synthèse de la réverbération, en augmentant le nombre de canaux de rebouclage.

- Adaptation à différents systèmes d'écoute. Suivant la capacité de la machine disponible et le système de reproduction sonore différentes configurations sont disponibles. Ainsi un système à quatre haut-parleurs pourra fonctionner selon un mode de panoramiques d'intensité, Ambisonics, ou transaural amélioré (prise de relais entre deux couples de haut-parleurs placés devant et derrière l'auditeur)
- Mise en commun de certains modules : le traitement simultanée de plusieurs sources implique *a priori* l'utilisation de plusieurs Spatialisateurs. Or les résultats de l'acoustique des salles autorise la mise en commun de certaines parties notamment pour la synthèse du champ réverbéré.

4.1.4 Commandes du Spatialisateur

L'architecture du Spatialisateur est basée sur le découpage de la réponse impulsionnelle en quatre parties temporelles (représentation morphologique de la réponse impulsionnelle), ce découpage ayant été validé par la psychoacoustique.

Commande bas niveau

La commande bas-niveau est un refet fidèle de la représentation morphologique de la réponse impulsionnelle. L'ensemble de ces commandes est rassemblé dans le module `Room_`.

Ce module envoie aux modules traitement du signal du Spatialisateur des messages correspondant à chacun des contrôles correspondant. La présence de sous-modules permet en outre d'effectuer un réglage en trois bandes de fréquences de chacune des commandes énergétiques, ainsi que de décrire les premières réflexions de façon plus fine (amplitude, temps d'arrivée).

La syntaxe de ces messages est une traduction fidèle de l'interface bas-niveau². Par

² L'implantation de réglages en plusieurs de bandes de fréquence s'effectue par utilisation de filtres. Afin de respecter la commande de ceux-ci, chaque valeur est donnée en pourcentage d'une valeur fixe. Par exemple le temps de réverbération possédera une valeur *tr0*, et sa valeur en bande haute *trh* sera donné par un facteur multiplicatif

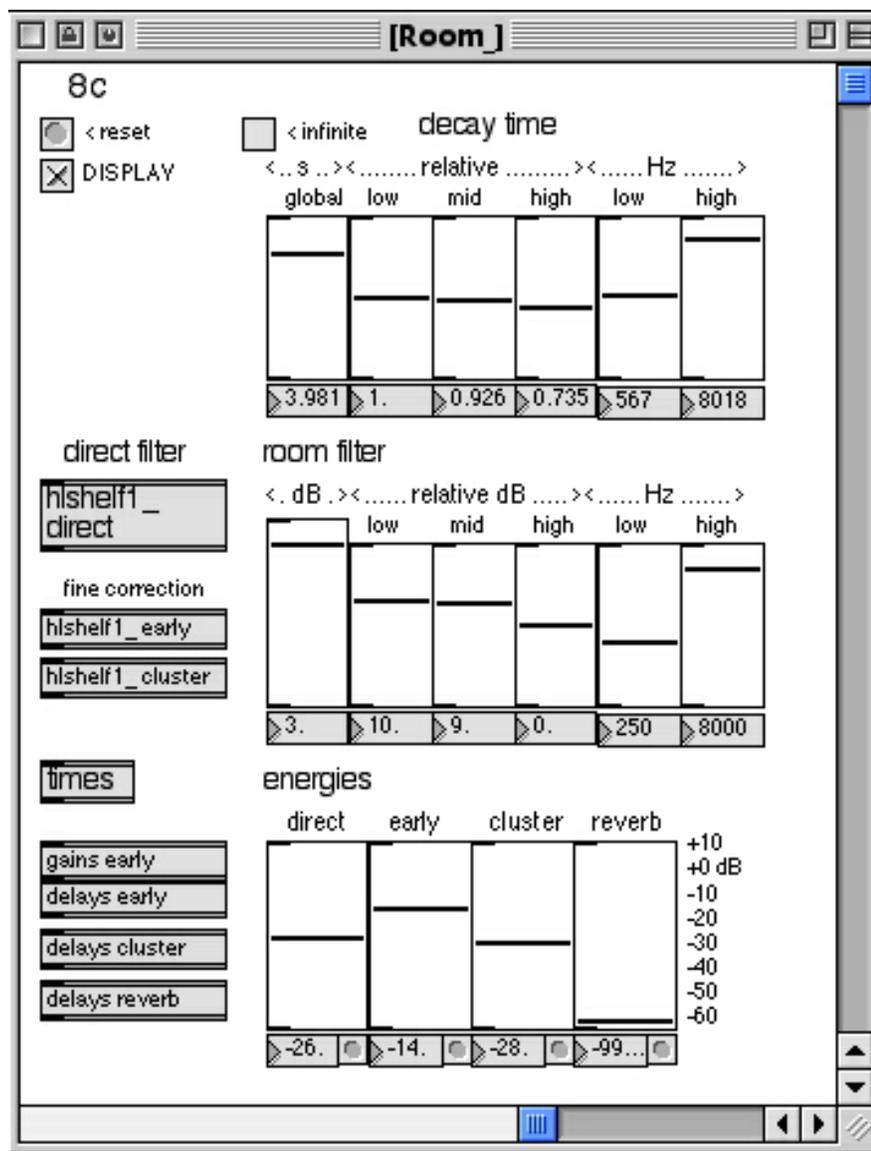


FIG. 4.2 – Module de commande bas-niveau

exemple,

Réglage	Message
Valeur du temps de réverbération	reverb tr0 2.1
Valeur de l'énergie du son direct	direct 0.005
Valeur de l'énergie de la réverbération	reverb 0.0002
Valeur de l'énergie relative du cluster en bande haute	cluster Gh 6

Interface perceptive

Afin d'offrir aux utilisateurs une interface intuitive et offrant une prise en main rapide, l'utilisation des résultats de la psychoacoustique concernant une description de l'effet de salle sous formes de facteurs perceptifs a été rajoutée au Spatialisateur. Ces facteurs sont ceux présentés au chapitre 1. Cette interface porte le nom de Spat_ OPer, et comprend, en plus, des commandes concernant le prétraitement de la source, et le positionnement dans l'espace.

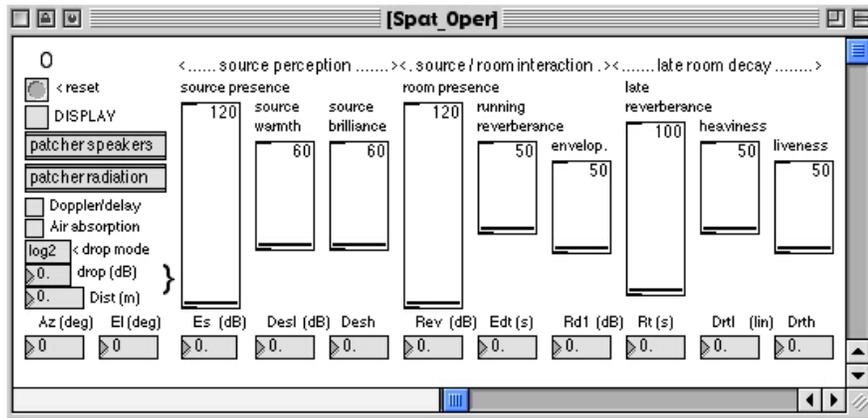


FIG. 4.3 – *Interface perceptive* ~

Adaptation d'une autre interface

La structure complète de commande bas-niveau du Spatialisateur permet *a priori* son interfaçage avec d'autres interfaces de manipulation. On verra au chapitre 5 et 6 comment une interface de type géométrique a été adaptée à cet environnement.

4.2 Environnemental Audio de Creative Labs

4.2.1 Contexte

Spécialisé dans la fabrication de cartes sons pour ordinateurs type compatibles IBM PC, équipés du système d'exploitation Windows, la société Creative Labs présente aussi aux développeurs de logiciels un ensemble de bibliothèques informatiques permettant d'exploiter au mieux les capacités de ces cartes. Ces dernières sont équipées de processeurs dédiés fabriqués par E-mu Systems. La technologie dédiée au son 3D est appelée Environmental Audio [Lab99].

4.2.2 Description de l'environnement

Cet environnement est composé de différents éléments :

- *matériel* : il s'agit de carte son équipée du processeur EMU101K. Cette carte peut fonctionner avec une fréquence d'échantillonnage allant jusqu'à 48 kHz, pour une résolution de 32 bits.
- *logiciel* : sur cette carte sont implémentées deux technologies : E-mu Environmental Modeling (EEM) permettant la synthèse d'un effet de salle, et Creative Multi-Speaker Surround (CMSS) permettant d'adapter un système de haut-parleurs (jusqu'à 8).
- *bibliothèque de développement (API)* : dénommée EAX, elle permet au développeur de logiciel d'accéder aux possibilités du matériel.
- *haut-parleurs* : Creative Labs propose un système de 5 haut-parleurs (4HP + 1 caisson de basse), permettant de recréer au mieux un environnement audio virtuel.

L'environnement considéré ici est la version 1.0 d'Environmental Audio³.

4.2.3 Interfaces de programmation

Ces interfaces sont une extension des technologies Microsoft DirectSound et DirectSound3D, dédiées au traitement des sons 3D pour le système opérationnel Windows. Cet environnement permet de créer et manipuler des sources sonores tridimensionnelles virtuelles. Il permet au développeur de logiciel de s'affranchir des problèmes liés à la multiplicité des cartes sons, en proposant une interface de programmation unique. Bien que permettant de reproduire un certain nombre d'effets (positionnement, effet Doppler), DirectSound3D ne propose pas en standard de traitements pour la réverbération. EAX permet alors de remédier à cette absence, et d'augmenter les capacités originales de DirectSound3D. EAX propose en outre un ensemble de préréglages décrivant des types de lieu.

Le programmeur utilise DirectSound par l'intermédiaire de buffers son, dans lesquels il écrit les échantillons d'un son (ou une forme d'onde). Dans le contexte de l'audio 3D, ce buffer, le *primary sound buffer* représente le récepteur (listener) et reçoit et mixe les données représentant les sources, *secondary sound buffer* ; à moins de vouloir effectuer lui-même un mixage spécifique, le programmeur ne manipule en général jamais le buffer primaire directement. Chacun de ces buffers représente un objet accessible par l'intermédiaire de deux interfaces : une interface qualifiée de standard (*standard buffer*

3. La version 2.0 présente des fonctionnalités accrues qui seront juste citées.

interface) permettant de régler des données non spécifiques au son 3D (volume, . . .), et une interface commandant des données telles la position d'une source.

L'accès à des capacités supplémentaires des cartes son, se fait par dans DirectSound par l'intermédiaire d'un mécanisme d'extension, les *property sets*, dont la manipulation se fait par l'intermédiaire d'une interface supplémentaire. EAX propose deux *property sets* :

- *sound source property set* : il possède une propriété manipulable, le **Reverb Mix**, permettant de régler le rapport des puissances entre le son direct et le son réverbéré issu d'une source. En général, EAX le règle automatiquement à partir de la distance entre la source et le récepteur.
- *listener property set* : il possède quatre propriétés manipulables :
 - **Environment Property** : elle permet de définir le type de réverbération utilisé en référence à une géométrie (salle fermée, espace ouvert). Par analogie, cette propriété règle à la fois une taille de salle et une absorption globale, agissant ainsi sur la date d'arrivée des premières réflexions, et sur le temps de réverbération. Un ensemble d'environnement type est proposé par EAX (chambre, hangar, salle de bain, canyon, ville, milieu sous-marin. . .). Les autres propriétés disponibles permettent d'affiner ces réglages.
 - **Volume Property** : elle permet de régler un niveau global de réverbération (master volume) pour le buffer primaire. Le réglage particulier pour chaque source, se fait par l'intermédiaire des buffers secondaires. Sa commande est linéaire et non logarithmique. .
 - **Decay Time Property** : elle agit sur le temps de réverbération (entre 0.1 et 20 secondes), et est présentée par une analogie avec un taux d'absorption global.
 - **Damping Property** : elle permet d'affiner la qualité de la réverbération, en offrant un réglage fréquentiel sur le temps de réverbération. Un damping de 0.5 correspond à un temps de réverbération constant : une augmentation de sa valeur, renforce le spectre aigu. Ce réglage s'effectue en deux bandes, dont la limite est 6 kHz.

Version 2.0

D'autres fonctionnalités sont présentes pour ces versions

- la présence d'obstacle par l'intermédiaire des concepts d'occlusion (source et récepteur dans deux salles différentes) et d'obstruction (obstacle entre la source et le récepteur dans une même salle).

- Contrôle fréquentiel sur la directivité des sources.
- Contrôle sur le modèle statistique sous-jacent du champ réverbéré (chapitre 3)
- Effets de distance

4.2.4 Correspondance avec une représentation bas-niveau

Le tableau suivant présente une correspondance possible entre les paramètres proposés par EAX 1.0 et une représentation morphologique bas-niveau du canal acoustique.

Représentation morphologique	EAX
OD	medium, humidity, currentLocation
R1	virtualSource, virtualSourceCount (sources-images)
R2	Aucun paramètre
R3	reverbAttenuation (niveau de signal réverbéré à inclure dans le signal de sortie)
Tr	roomSize (donnée comme une distance en mètres entre deux murs, roomReflectivity (absorption en décibels))

La version 2.0 permet d'affiner cette correspondance. Néanmoins l'accès direct à ses propriétés de bas-niveau n'est pas direct, ce qui est compréhensible vis-à-vis des programmeurs. Il semble plutôt que le modèle de contrôle de EAX aille vers celui d'une description de type géométrique statistique d'une scène sonore, tout en respectant une hiérarchie psychoacoustique dans le choix des paramètres.

4.3 Apple SoundSprocket

4.3.1 Contexte

Le développement de jeux représentant une part importante du marché informatique, la société Apple Computer a mis à la disposition des programmeurs un ensemble de quatre bibliothèques, les Game Sprockets destinées à faciliter de leur développement [App96]. L'une d'elle, SoundSprocket, est consacrée à l'audio 3D.

La version de SoundSprocket présentée ci-dessous est la version 1.0. Une nouvelle version, 1.7, est prévue pour le dernier trimestre 1999 ; elle intègre un nouveau moteur de rendu 3D audio, acquis auprès de la société Spatializer (technologie enCompass).

Les interfaces de programmation restent quasi-identiques, les différences se situant au niveau des passages de paramètres concernant les composantes 3D⁴.

4.3.2 Description de l'environnement SoundSprocket

SoundSprocket présente au programmeur un environnement audio virtuel (Virtual Audio Environment) constitué d'un récepteur unique et de une ou plusieurs sources sonores. Le nombre de sources disponibles est limité par les capacités de traitement de l'ordinateur, mais il est conseillé de se restreindre à un nombre permettant de conserver une intelligibilité pour le récepteur. Deux niveaux de manipulation sont accessibles au programmeur.

Le choix entre plusieurs dispositifs d'écoute est proposé à l'auditeur par l'intermédiaire d'un tableau de bord. Les options suivantes sont disponibles : paires de haut-parleurs, haut-parleur monophonique, casque.

Il est, de plus, précisé que l'utilisation sur une paire de haut-parleurs met en œuvre un algorithme d'annulation des trajets croisés, ce qui sous-entend l'utilisation de techniques binaurales/transaurales.

4.3.3 Interfaces de programmation

On examinera dans ce paragraphe les interfaces bas-niveau et haut-niveau.

Interfaces de bas-niveau

Les interfaces de programmation bas-niveau tiennent compte de la structure du gestionnaire audio global du système MacOS, le Sound Manager. Un composant de localisation (localization component) est installé dynamiquement et permet l'auralisation de données audio, issues d'un fichier ou stockées en mémoire, provenant du Sound Manager.

Ce composant de localisation est décrit par la structure en langage C suivante^{5 6} :

4. La version originale de la librairie utilise le formalisme de la librairie graphique QuickDraw 3D, la nouvelle permet d'utiliser celui de la librairie OpenGL.

5. Un certain nombre de champs ne concernant pas la description audio ont été supprimés afin d'améliorer la lisibilité

6. UInt32 correspond à un entier long non signé (typedef unsigned long UInt32;)

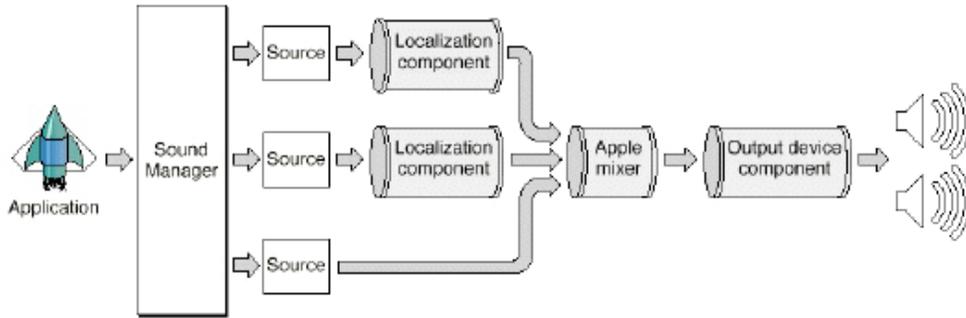


FIG. 4.4 – Architecture bas-niveau de l'environnement SoundSprocket

```

struct SSpLocalizationData {
    UInt32      medium      /*Medium for sound propagation*/
    float       humidity    /*Humidity when medium is air*/
    float       roomSize    /*Reverb model--distance bet. walls*/
    float       roomReflectivity /*Reverb model--bounce attenuation*/
    float       reverbAttenuation /*Reverb model--mix level*/
    UInt32      sourceMode  /*Type of filtering to apply*/
    float       referenceDistance /*Nominal distance for recording*/
    float       coneAngleCos /*Cos(angle/2) of attenuation cone*/
    float       coneAttenuation /*Attenuation outside the cone*/
    SSpLocationData currentLocation /*Location of the sound*/
    UInt32      virtualSourceCount /*Number of reflections*/
    SSpVirtualSourceData virtualSource[4] /*The reflections*/
}
  
```

Cette structure fait appel à deux structures décrivant l'une la localisation de la source, l'autre un ensemble de réflexions placées arbitrairement par le programmeur, auxquelles sont associés une position et un facteur d'atténuation.

```

struct SSpLocationData {
    float elevation      /*Angle of the meridian--pos is up*/
    float azimuth       /*Angle of the parallel--pos is left*/
    float distance      /*Distance between source and listener*/
    float projectionAngle /*Cos(angle) between cone and listener*/
    float sourceVelocity /*Speed of source toward the listener*/
    float listenerVelocity /*Speed of listener toward the source*/
}

```

et

```

struct SSpVirtualSourceData {
    float attenuation /*Attenuation factor*/
    SSpLocationData location /*Location of virtual source*/
}

```

Un examen plus attentif de ces structures les fait plutôt apparaître comme une combinaison d’une représentation bas-niveau de type morphologique, et d’une interface de manipulation haut niveau de type statistique. Il faut noter que la manipulation d’une telle structure suppose que le centre du repère se trouve à l’endroit du récepteur, i.e la source est décrite en coordonnées relatives par rapport aux récepteurs.

Le tableau suivant montre une correspondance possible entre l’ensemble de ces paramètres et ceux d’une représentation bas-niveau morphologique de l’effet de salle.

Représentation morphologique	SoundSprocket
OD	medium, humidity, currentLocation
R1	virtualSource, virtualSourceCount (sources-images)
R2	Aucun paramètre
R3	reverbAttenuation (niveau de signal réverbéré à inclure dans le signal de sortie)
Tr	roomSize (donnée comme une distance en mètres entre deux murs, roomReflectivity (absorption en décibels)

Les champs relatifs aux caractéristiques de la source (sourceMode, referenceDistance, coneAngleCos, coneAttenuation) correspondraient plutôt au module Source~ du Spatialisateur, et concerne la directivité de la source. Un des champs pourrait concerner la

représentation du canal acoustique. Il s'agit du champ *sourceMode*, qui permet d'attribuer un type à la source. Parmi ces types se trouve l'option `kSSpSourceMode_Ambient`, qui décrit une source entourant le récepteur. Sur une représentation morphologique bas-niveau, cela correspondrait à un canal où seul R3 aurait une valeur non nulle.

Interfaces de haut-niveau

Ces interfaces permettent de manipuler une scène sonore par l'intermédiaire d'objets sans se soucier de la configuration correspondante pour les composants de localisation. Elles permettent entre autres de créer une scène comprenant un récepteur, auxquelles sont attachées les caractéristiques de la réverbération tardive (par l'intermédiaire de la salle) et plusieurs sources.

4.3.4 Conclusion

De manière semblable à EAX, la librairie SoundSprocket propose plutôt au développeur un ensemble de paramètres basés sur une description statistique de l'effet de salle. Le couplage avec d'autres interfaces de manipulation doit donc se faire en "retrouvant" les paramètres d'une représentation bas-niveau. On notera cependant la possibilité de placer des premières réflexions, permettant de dépasser le modèle simple {son direct + réverbération}.

Chapitre 5

Représentation bas-niveau de l'effet de salle

Ainsi qu'il a été vu au chapitre premier, deux représentations dites "bas-niveau" peuvent décrire un effet de salle : la réponse impulsionnelle du canal, ou une représentation qualifiée de morphologique qui décrit la forme de cette réponse, en termes d'énergies et de critères temporels. Le chapitre 3 a de plus montré l'importance de cette représentation lors d'une manipulation de l'effet de salle à travers une interface de type perceptive ou statistique ; on verra de plus au chapitre 6 son utilité pour un rendu en temps réel de la réverbération tardive. Ce chapitre présente les méthodes utilisées pour extraire les coefficients de cette représentation morphologique à partir de la donnée d'une réponse impulsionnelle, et revient plus particulièrement sur l'analyse du champ réverbéré.

5.1 Conversion vers une représentation morphologique

Le signal de départ est une réponse impulsionnelle, obtenue par simulation ou enregistrement.

5.1.1 La courbe de décroissance intégrée

Définition

La courbe de décroissance intégrée (Energy Decay Curve) est couramment utilisée en acoustique des salles pour l'estimation de paramètres énergétiques et du temps de réverbération. Supposons que la source émette un signal permanent, bruit blanc stationnaire à niveau constant ; après interruption de l'émission de ce signal, le niveau sonore décroît dans la salle. La mesure de ce niveau au point récepteur correspond à une réalisation de la courbe de décroissance intégrée. La courbe théorique est obtenue en renouvelant l'expérience une infinité de fois et en moyennant les résultats.

Schroeder a montré qu'il était possible de retrouver cette courbe de décroissance intégrée théorique par intégration rétrograde du carré de la réponse impulsionnelle du même canal (échogramme).

$$EDC_h(t) = \int_t^{\infty} h^2(\tau) d\tau \quad (5.1)$$

La valeur de l'EDC à l'instant t représente donc l'énergie restant dans la réponse à partir du temps t .

Extraction de paramètres

Il est alors possible de calculer des paramètres morphologiques large-bande à partir de l'EDC. On a, compte tenu des limites temporelles établies au chapitre 2 (les temps sont exprimés en millisecondes):

$$\begin{cases} OD &= EDC(0) - EDC(20) \\ R1 &= EDC(20) - EDC(40) \\ R2 &= EDC(100) - EDC(40) \\ R3 &= EDC(100) \end{cases} \quad (5.2)$$

Après un certain temps, cette courbe prend la forme d'une exponentielle décroissante, ou, sur une échelle logarithmique, d'une droite. La pente de cette droite est indépendante des positions de la source et du récepteur et permet de mesurer le temps de réverbération caractéristique de la salle. Bien qu'une mesure rigoureuse implique l'utilisation d'une dynamique de 60 décibels, celle-ci est rarement présente en pratique. L'estimation s'effectue alors par régression linéaire sur un intervalle temporel, qu'il convient au préalable d'estimer. Il est limité, en amont, par l'instant où la réponse a adopté un profil exponentiel et, en aval, par l'instant d'arrivée du bruit. Ce dernier correspond à l'instant où le niveau du bruit est égal au niveau du signal utile, ce qui se traduit par la présence d'un coude sur l'échogramme (*figure 5.1*). Sans cette précaution, la présence de bruit de mesure dans la réponse analysée conduit à une surestimation du temps de réverbération.

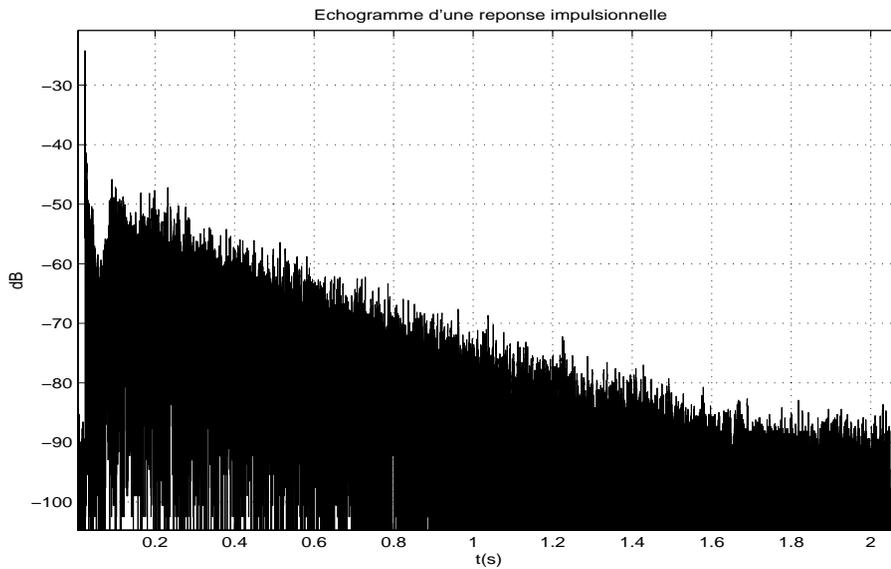


FIG. 5.1 – *Echogramme d'une réponse impulsionnelle. On notera la présence de bruit de mesure à partir de 1.6 s.*

Afin de réduire ce biais, on peut également estimer la puissance du bruit sur la fin de la réponse impulsionnelle, puis, compte tenu de l'indépendance statistique entre la réponse et le bruit, soustraire celui-ci avant intégration rétrograde (*figure 5.2*).

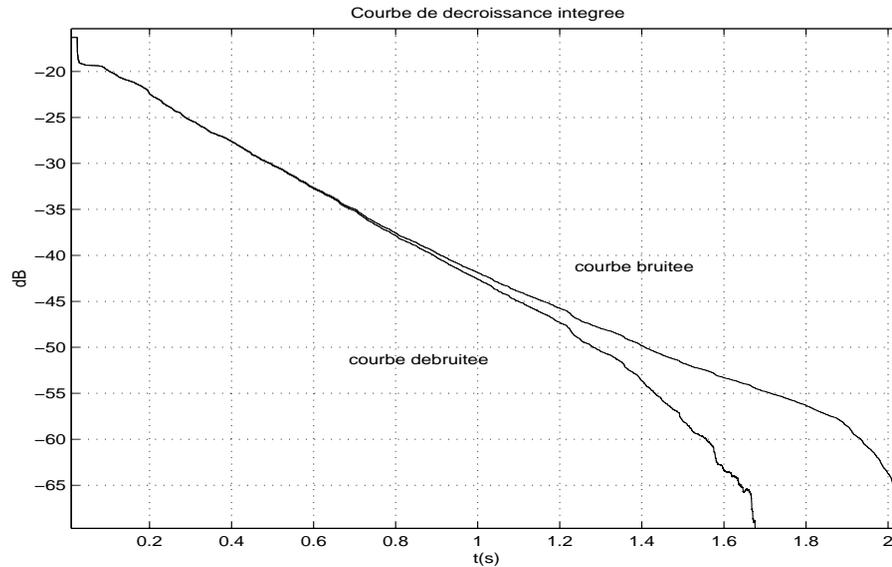


FIG. 5.2 – Courbe de décroissance intégrée large bande. La présence de bruit peut introduire une surestimation du temps de réverbération. Par soustraction du bruit avant intégration rétrograde, on peut corriger ce biais (l'estimation se fera alors sur l'intervalle temporel $[0.2 \text{ s}, 1.2 \text{ s}]$, soit une dynamique d'à peu près 30 dB).

5.1.2 Estimation de l'EDC en bandes étroites

Utilisation d'un banc de filtres

Si l'on désire mesurer les paramètres morphologiques pour différentes bandes de fréquences, un filtrage est d'abord effectué sur la réponse impulsionnelle, et ce selon la plage fréquentielle considérée (figure 5.3). Une courbe de décroissance intégrée est calculée pour chaque signal filtré (figure 5.4).

On notera cependant que le temps d'arrivée du bruit varie selon chaque canal fréquentiel. La décroissance étant habituellement plus rapide en hautes fréquences, le bruit y arrive généralement plus tôt qu'en basses fréquences. A chaque signal correspond donc un intervalle temporel spécifique et une dynamique d'estimation maximale.

Limitations

Un des inconvénients de cette méthode est, *a priori*, sa dépendance au type de filtrage utilisé. La figure 5.5 montre l'application de différents filtrages, pour une bande allant de 4KHz à 8 kHz, à une même réponse impulsionnelle, ainsi que les estimations de temps de réverbération en découlant. Bien que légèrement différents, ces écarts ne sont

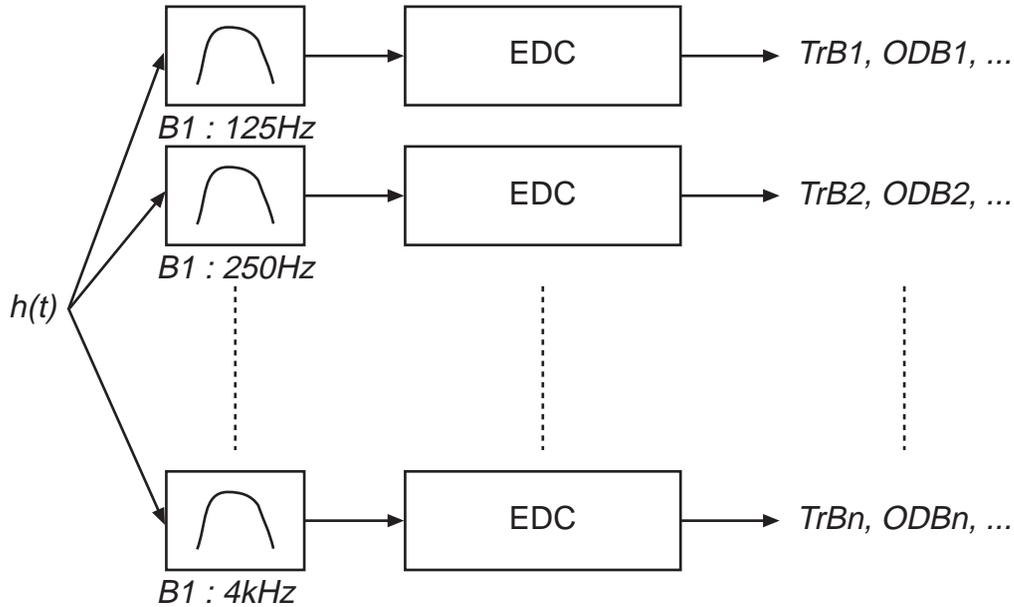


FIG. 5.3 – *Obtention de paramètres dans plusieurs bandes de fréquence, par filtrage préalable de la réponse impulsionnelle.*

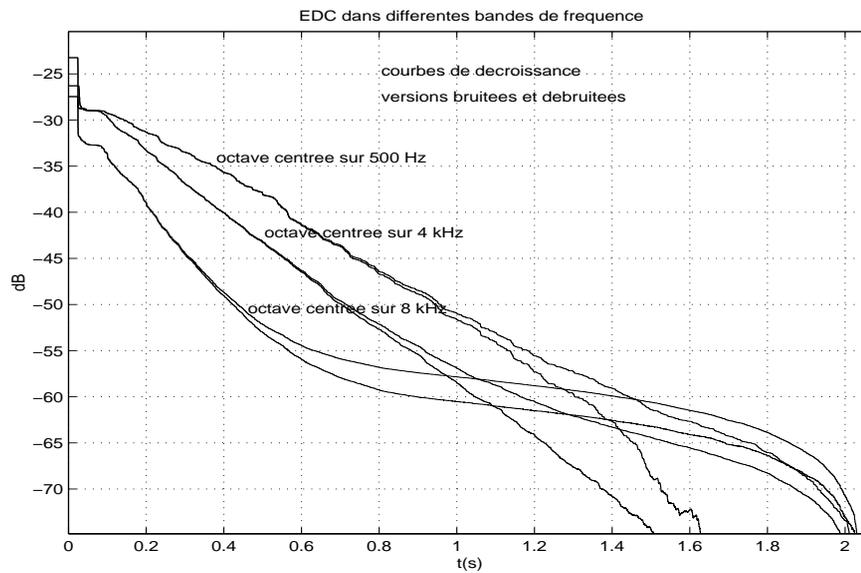


FIG. 5.4 – *Courbes de décroissance intégrées dans différentes bandes de fréquence (obtention par filtrage de Butterworth d'ordre 8), avec et sans débruitage. Le temps d'arrivée du bruit varie selon la bande fréquentielle considérée.*

pas perceptivement pertinents¹ (tableau 5.1.2).

1. En effet, pour des temps de réverbération supérieurs à 1 seconde, l'oreille ne repère des écarts qu'à partir d'une variation relative de 10%.

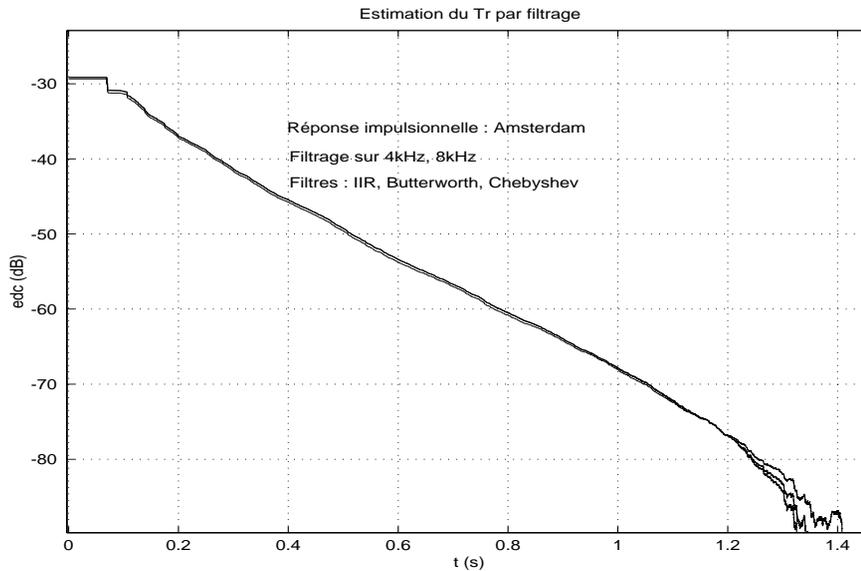


FIG. 5.5 – Courbes de décroissance intégrées pour une même bande, avec différents types de filtrage.

Les temps de réverbération obtenus pour la figure (5.5) sont :

Filtrage	Temps de réverbération
Filtre IIR	1.5159
Butterworth	1.5112
Chebyshev	1.5144

Un autre problème posé par cette méthode est celui de l'interprétation à apporter à une estimation du temps de réverbération, et notamment de savoir si la valeur estimée correspond effectivement à celle globale de la bande, ou à celle d'une région fréquentielle particulière de la bande. En effet, le temps de réverbération décroît généralement avec la fréquence et peut, lors d'une estimation sur une bande fréquentielle, être biaisé par les valeurs correspondant aux fréquences inférieures. Ce comportement est illustré par la figure 5.6 où le temps de réverbération évalué sur la plage [2 kHz, 6kHz] est "pilote" par la valeur qu'il atteint dans la partie basse de cet intervalle fréquentiel ([2 kHz, 4kHz]) masquant ainsi la valeur plus faible qu'il présente en partie haute ([4 kHz, 6 kHz]).

5.1.3 L'analyse par relief de décroissance

Une autre méthode pour extraire les valeurs de la représentation morphologique selon la fréquence consiste à effectuer une analyse temps-fréquence de la réponse, et ce en bande fine, puis de réduire l'information dans les bandes de fréquence désirées.

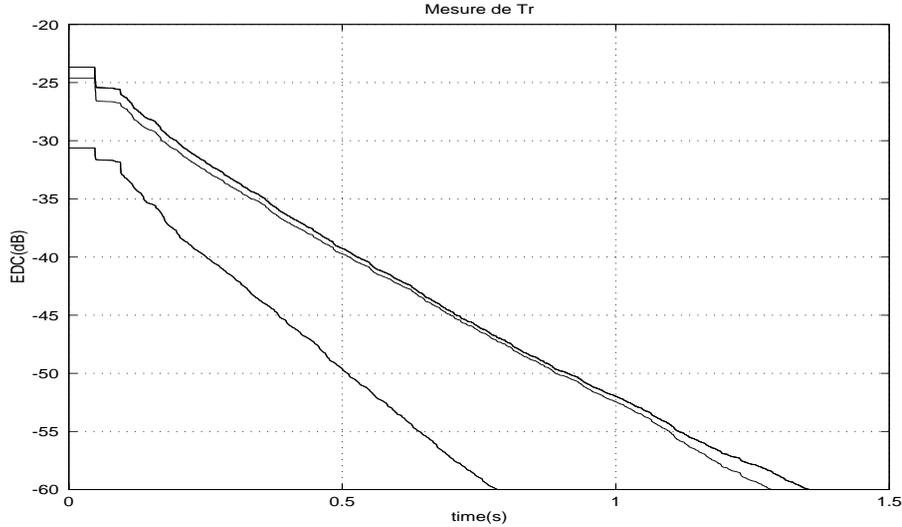


FIG. 5.6 – L'estimation du T_r est effectuée sur 3 bandes de fréquences : $[2 \text{ kHz}, 4 \text{ kHz}]$ (courbe médiane), $[4 \text{ kHz}, 6 \text{ kHz}]$ (courbe inférieure) et $[2 \text{ kHz}, 6 \text{ kHz}]$ (courbe supérieure).

Fondements mathématiques [Jot92]

Une démarche naturelle consiste à rechercher une représentation temps-fréquence de la réponse impulsionnelle, telle que toute isofréquence puisse être considérée comme une courbe de décroissance intégrée évaluée localement autour de la fréquence considérée. Ce résultat est obtenu dans le cas où l'on connaît l'évolution temporelle de l'énergie dans la réponse au voisinage de la fréquence considérée, c'est à dire une représentation temps-fréquence de l'énergie, obtenue à partir de la réponse impulsionnelle, à l'aide d'une distribution.

$$h(t) \xrightarrow{p} \rho_h(t, f) \quad (5.3)$$

La recherche d'un certain nombre de propriétés pour cette représentation permet de restreindre l'ensemble des densités candidates.

Définition 1 Si $\rho_h(t, f)$ est une représentation temps-fréquence énergétique du signal $h(t)$, le relief de décroissance associé est défini par

$$EDR_h(t, f) = \int_t^{+\infty} \rho_h(\tau, f) d\tau \quad (5.4)$$

Il est également possible de définir le relief de décroissance en accord avec la définition première de la courbe de décroissance intégrée :

Définition 2 Le relief de décroissance est défini comme la moyenne statistique sur plusieurs réalisations d'une représentation temps-fréquence de la mesure de l'énergie, après interruption du signal d'excitation, lorsque celui-ci est un bruit blanc stationnaire.

L'équivalence entre les deux définitions est réalisée quand ρ obéit à certaines propriétés, détaillées par la suite et déterminantes pour le choix idéal de ρ .

Proposition 1 *L'intégration du relief $EDR_h(t, f)$ selon les fréquences doit permettre de retrouver la courbe de décroissance intégrée $EDC_{h(t)}$. Cette propriété est satisfaite si ρ conserve les lois marginales dans le domaine temporel, soit*

$$\int_{-\infty}^{\infty} \rho_h(t, f) df = h^2(t) \implies \int_{-\infty}^{+\infty} EDR_h(t, f) df = EDC_h(t) \quad (5.5)$$

Proposition 2 *De façon analogue dans le domaine fréquentiel, l'intégration selon le temps doit permettre de retrouver le carré du module de la transformée de Fourier de $h(t)$. Cette propriété est assurée si ρ conserve les lois marginales dans le domaine fréquentiel.*

$$\int_{t=-\infty}^{\infty} \rho_h(t, f) dt = |H(f)|^2 \iff EDR_h(-\infty, f) df = |H(f)|^2 \quad (5.6)$$

Proposition 3 *ρ doit préserver la causalité, c'est-à-dire*

$$h(t) = 0 \text{ pour } t < 0 \implies \rho_h(t, f) = 0 \text{ pour } t < 0 \quad (5.7)$$

Ceci entraîne

$$\int_{t=0}^{\infty} \rho_h(t, f) dt = |H(f)|^2 \iff EDR_h(0, f) = |H(f)|^2 \quad (5.8)$$

Proposition 4 *Une généralisation de la proposition précédente conduit à vouloir trouver une interprétation de chaque isotemporelle du relief de décroissance. Si l'on considère une version tronquée de la réponse $h'(t)$, soit $h'(t) = h(t)\mathcal{U}(t - \tau)$, où \mathcal{U} est l'échelon de Heaviside, alors la représentation associée $\rho_{h'}(t, f)$ est nulle pour tout instant inférieur à τ , si ρ est compatible avec les translations temporelles. La vérification de cette propriété ainsi que des deux précédentes entraîne alors*

$$EDR_{h'}(\tau, f) = \int_{t=\tau}^{+\infty} \rho_{h'}(t, f) dt = |H'(f)|^2 = \left| \int_{t=\tau}^{+\infty} h(t) e^{-j2\pi ft} dt \right|^2 \quad (5.9)$$

Proposition 5 *Pour que l'interprétation de la proposition précédente soit valide pour $h(t)$ et non seulement $h'(t)$, il faut imposer que $\rho_{h'}(t, f)$ soit identique à $\rho(t, f)$ pour tous les instants ultérieurs à t . Il s'agit d'une contrainte d'anticausalité de la distribution ρ , c'est-à-dire que $\rho_h(\tau, f)$ ne dépend que des valeurs de $h(t)$ pour les instants ultérieurs à τ .*

L'ensemble de ces propositions permet alors d'obtenir une solution pour la distribution ρ , et par conséquent une définition unique et intuitive du relief de décroissance².

Définition 3 *Le relief de décroissance décrit la densité spectrale d'énergie restant dans le signal après l'instant t et coïncide avec le spectre courant futur.*

$$EDR_h(t, f) = \left| \int_{\tau=t}^{+\infty} h(\tau) e^{-j2\pi f\tau} d\tau \right|^2 \quad (5.10)$$

Calcul pratique

Algorithme

En pratique, le calcul du spectre courant futur nécessite un coût de calcul trop élevé. L'utilisation d'une transformée de Fourier à court terme, suivie d'une intégration temporelle rétrograde est utilisée. Afin de réduire l'influence du bruit, une procédure récursive est utilisée. En effet, le bruit étant supposé indépendant du signal, il est possible d'estimer la puissance de celui-ci et de la retrancher au canal fréquentiel considéré avant intégration rétrograde. Pour chaque canal, il faut estimer trois quantités : la puissance du bruit, le temps de réverbération et la valeur à l'origine de la droite de régression utilisée pour la mesure du temps de réverbération. On appellera cette quantité *spectre de puissance initiale*, $P_0(f)$. Sa signification sera précisée au paragraphe 5.1.4.

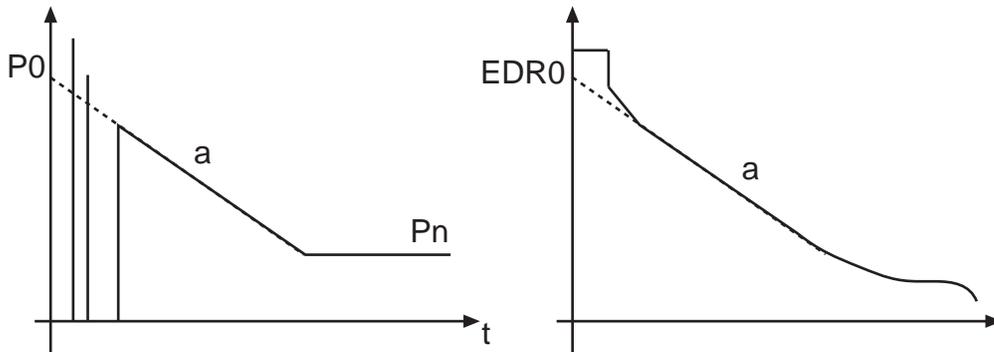


FIG. 5.7 – Représentation des 3 quantités R_t , P_0 , P_n , à partir des canaux de transformées de Fourier à court terme (gauche) et de relief de décroissance (droite). Le coefficient a est relié au temps de réverbération par $a = -60/R_t$ et EDR_0 est relié au spectre de puissance initiale par $P_0 = 6 \ln 10 \frac{EDR_0}{R_t}$

L'ensemble de la procédure est représenté sur la figure 5.8. Après calcul du spectre de Fourier à court terme de la réponse impulsionnelle, chaque canal fréquentiel est consi-

² La distribution ρ correspond alors à une distribution connue sous le nom de distribution de Page-Levin

déré en particulier. Une segmentation temporelle grossière du signal permet d'abord d'estimer des valeurs initiales de P_0 , R_t (par régression linéaire sur le début du signal) et P_n (estimation directe sur la fin du signal). Ces valeurs permettent de lancer la procédure itérative : à chaque pas un temps limite est calculé permettant de considérer une plage temporelle maximale d'estimation pour P_0 et R_t . La courbe de décroissance intégrée du signal est alors calculée après soustraction de la valeur du niveau de bruit, et de nouvelles valeurs de P_0 et R_t sont estimées. Celles-ci servent ensuite à calculer une courbe de décroissance théorique dont la contribution est retranchée de la partie bruitée du signal, afin d'améliorer l'estimation de P_n .

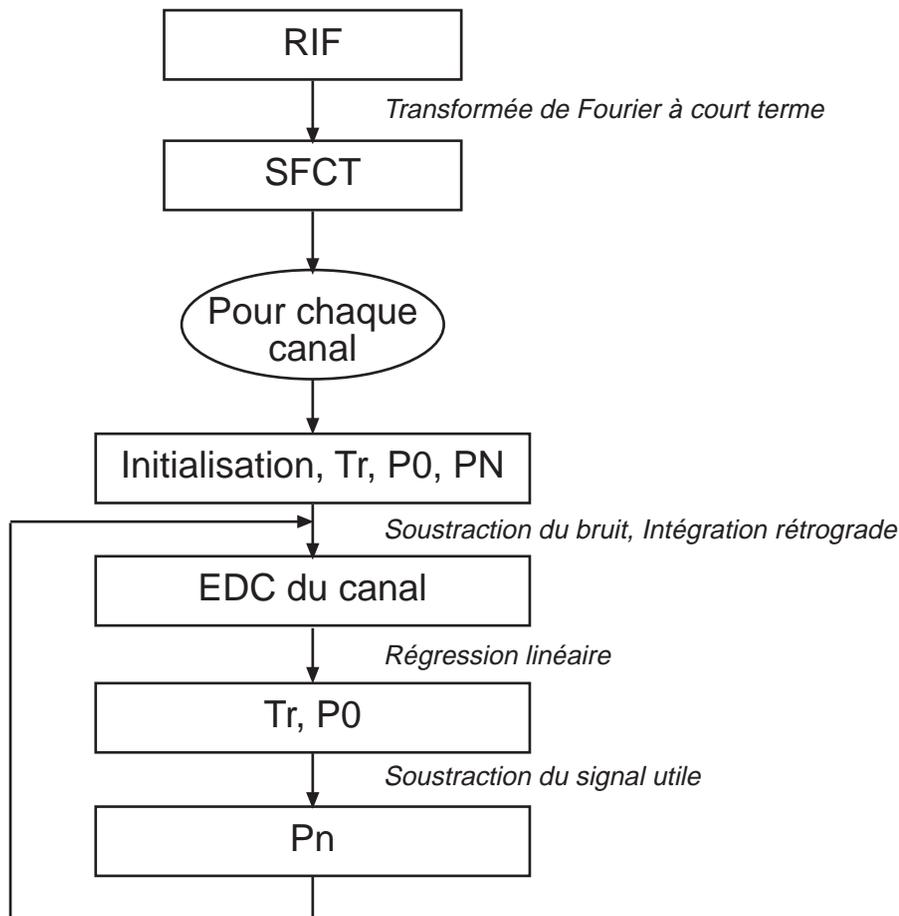


FIG. 5.8 – Procédure récursive d'estimation des paramètres T_r , P_n , P_0 . La connaissance des valeurs de ces trois paramètres pour chaque canal permet de calculer un relief de décroissance débruité.

Pour le spectre de puissance initiale, l'erreur, en dB, est donnée par : $\Delta_{\log} P_0 = 10 \log E[\widehat{P}_0] - 10 \log P_0$ et l'écart type en dB, exprimé , par $10 \log(\sigma_{\widehat{P}_0} + P_0) - 10 \log(P_0)$. Les cases hachurées correspondent à des erreurs d'estimation inférieures à 1 dB.

Erreur sur l'estimation de P_0 Ecart-type sur l'estimation de P_0

	0.25	0.50	1.00	2.00	4.00	8.00		0.25	0.50	1.00	2.00	4.00	8.00
10	2.9	2.2	////	////	1.0	1.0	10	4.2	3.0	2.4	1.8	1.4	1.1
20	2.5	////	////	////	////	1.1	20	3.9	2.1	1.3	1.3	////	////
30	////	////	////	////	////	////	39	2.2	1.6	////	////	1.1	////
40	////	////	////	////	////	////	40	1.4	1.0	////	////	////	////
50	////	////	////	////	////	////	50	1.8	1.3	////	////	////	1.0
60	////	////	////	////	////	////	60	////	////	////	////	////	////
70	////	////	////	////	////	////	70	1.1	////	////	////	////	1.1
80	////	////	////	////	////	1.1	80	////	////	////	////	////	1.0
90	////	////	////	////	////	////	90	1.1	////	////	////	////	1.1
100	////	////	////	////	////	////	100	1.0	////	////	////	////	////
Inf	////	////	////	////	////	////	Inf	////	////	////	////	////	1.0

On remarquera que les limites de la méthodes sont très liées au rapport entre la durée du signal considéré et le temps de réverbération. Ainsi si ce rapport est trop faible (dernière colonne des tableaux), le biais sur l'estimation de R_t est maximal. Ce biais est atténué lors de l'estimation de P_0 ; en effet celui-ci est calculé indirectement par utilisation de la courbe de décroissance selon $P_0 = 6 \ln 10 \frac{EDR_0}{R_t}$. La différentiation de cette formule fait apparaître un rapport $\Delta R_t / R_t$, qui pondère l'influence de l'erreur sur le temps de réverbération pour des valeurs fortes de celui-ci. L'augmentation du rapport signal à bruit agit plutôt sur l'écart type de l'estimation (sauf pour des valeurs extrêmes ou le biais peut être considérable). L'utilisation et le moyennage de plusieurs mesures s'avère alors intéressant.

5.1.4 Approfondissement sur l'analyse de la réverbération tardive

Modélisation statistique du champ diffus

L'hypothèse de Sabine pour la modélisation du champ diffus concerne la distribution spatiale de l'énergie. D'autres phénomènes apparaissent également lors de la propagation des ondes dans la salle : une augmentation de la densité d'échos $D_e(t)$ au cours du temps

et de la densité modale $D_m(f)$ en fonction de la fréquence. Ces propriétés peuvent s'exprimer par :

$$D_e(t) = 4\pi c^3 \frac{t^2}{V} \quad (5.12)$$

et

$$D_m(f) = 4\pi V \frac{f^2}{c^3} \quad (5.13)$$

où c est la vitesse du son et V le volume de la salle.

L'équation 5.13 implique qu'au delà d'une certaine fréquence, un signal va exciter simultanément plusieurs modes propres de la salle ; le signal obtenu est alors égal à la somme de leurs contributions. L'amplitude et la phase de celles-ci dépendent de la position ; en utilisant le théorème central limite, la réponse fréquentielle de la salle peut être considérée comme la somme de deux processus aléatoire gaussien, l'un réel, l'autre complexe. De manière duale, pour une position donnée de source et une excitation stationnaire, la réponse fréquentielle observée par le récepteur répond à une loi identique en fonction de la fréquence. Les propriétés de ce modèle statistique sont alors indépendantes des positions des transducteurs. Ce modèle n'est cependant valide qu'au delà d'une certaine fréquence, la fréquence de Schroeder f_s .

$$f_{Schroeder} = 2000 \sqrt{\frac{T_r}{V}} \quad (5.14)$$

Du point de vue temporel, Polack a développé un modèle, où la réponse impulsionnelle est considérée comme une réalisation d'un processus aléatoire, $h(t) = b(t)e^{-\delta t}$, ou $b(t)$ est un bruit gaussien stationnaire. Sa densité spectrale sera notée $P_0(f)$. Le coefficient δ est relié au temps de réverbération de la salle par l'équation

$$R_t = \frac{3 \ln 10}{\delta} \quad (5.15)$$

Ce modèle est également valide au delà d'un certain temps, le temps de mélange. En accord avec les propriétés d'intégration temporelle du système auditif, cela implique que dix échos se recouvrent sur 24 ms. Polack en donne une estimation par

$$t_{mélange} = \sqrt{V}, \text{ en ms, } V : \text{ volume de la salle (m}^3\text{)} \quad (5.16)$$

Intuitivement cette notion est en accord avec la théorie de Sabine. En effet, il faut qu'un nombre de réflexion soit suffisamment élevé pour atteindre une distribution spatiale uniforme. L'estimation des paramètres de ce modèle, c'est-à-dire des espérances de $h(t)^2$ et de $H(f)^2$ conduit à séparer les paramètres de temps et de fréquences. Le développement d'un modèle conjoint se fait par extension du modèle de Polack, en y introduisant une dépendance fréquentielle. Il revient alors à estimer une enveloppe énergétique de la

réverbération tardive, possédant des propriétés précises. Celle-ci peut s'exprimer selon deux quantités le spectre de puissance initial $P_0(f)$, correspondant au spectre de l'énergie avant son émission dans la salle, et le temps de réverbération $T_r(f)$, qui détermine la décroissance de cette énergie :

$$ENV_h(t, f) = P_0(f)e^{-6 \ln 10 \frac{t}{T_r(f)}} \quad (5.17)$$

En raison de sa construction cette enveloppe est reliée à l'espérance du relief de décroissance par

$$ENV_h(t, f) = \frac{6 \ln 10}{T_r(f)} [EDR_h(t, f)] \quad (5.18)$$

La recherche d'une interprétation physique du spectre de puissance initiale, permet de plus de montrer que celui-ci est relié aux fonctions de transfert en champ diffus des transducteurs utilisés pour la mesure de $h(t)$, et ce par :

$$P_0(f) = \frac{1}{V} S_d(f)^2 R_d(f)^2 \quad (5.19)$$

où $S_d(f)$ et $R_d(f)$ sont respectivement les fonctions de transfert en champ diffus de la source et du récepteur. Ceci permet de développer une méthode de mesure des fonctions de transfert en champ diffus basée sur l'utilisation du relief de décroissance (*cf* § 5.1.6).

5.1.5 Exemple d'analyse d'une réponse impulsionnelle

Les figures suivantes montrent les différentes étapes de calcul du relief de décroissance. Elles permettent de suivre l'évolution de la procédure d'analyse et de débruitage de la réponse impulsionnelle présentée à la figure 5.9.

La première étape consiste à calculer le spectre de Fourier à court terme de cette réponse (*figure 5.10*). Les variations rapides du processus stochastique lié à la réverbération dans la salle empêchent une bonne lisibilité pour estimer le temps de réverbération. On constate de plus en début de réponse, la présence d'un filtre en peigne, dû à la présence d'une forte réflexion précoce présente dans les premières fenêtres de la transformée de Fourier.

La figure 5.11 présente le résultat du calcul du relief de décroissance sans débruitage. Bien que le cumul rétrograde améliore la visibilité, la présence du bruit de mesure introduit un biais sur l'estimation des paramètres de temps de réverbération $T_r(f)$ et de spectre initial de puissance $P_0(f)$.

Comme en atteste la figure 5.12, la procédure de débruitage permet d'augmenter la dynamique utile pour le calcul de ces paramètres. La courbe foncée indique le temps limite d'arrivée du bruit pour chaque canal fréquentiel.

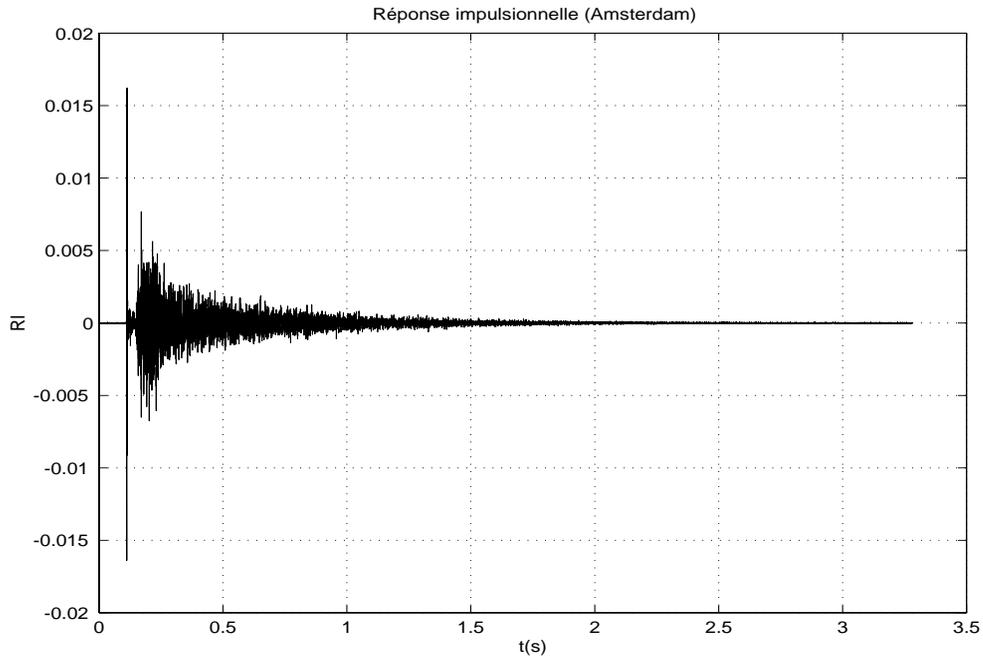


FIG. 5.9 – Réponse impulsionnelle (Amsterdam)

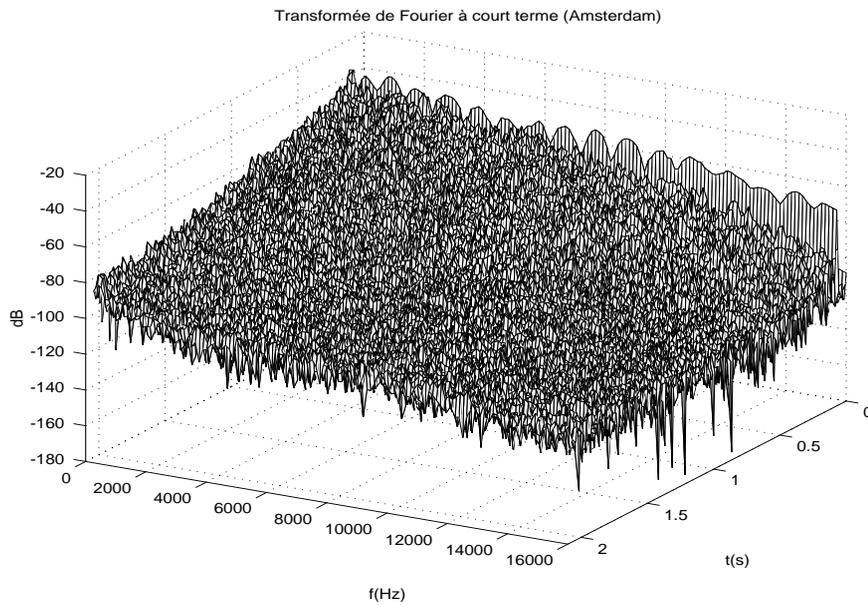


FIG. 5.10 – Analyse de la Réponse impulsionnelle par Transformée de Fourier à court terme

La dernière étape consiste à prolonger artificiellement le relief de décroissance (*figure 5.13*), en utilisant les paramètres $R_t(f)$ et $P_0(f)$. Ce relief peut alors être utilisé pour

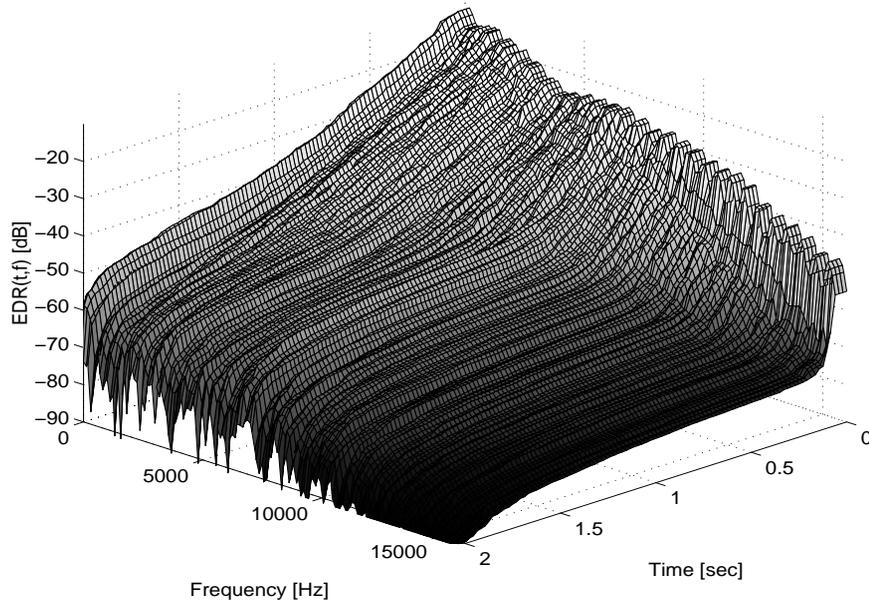


FIG. 5.11 – Calcul du relief de décroissance sans débruitage.

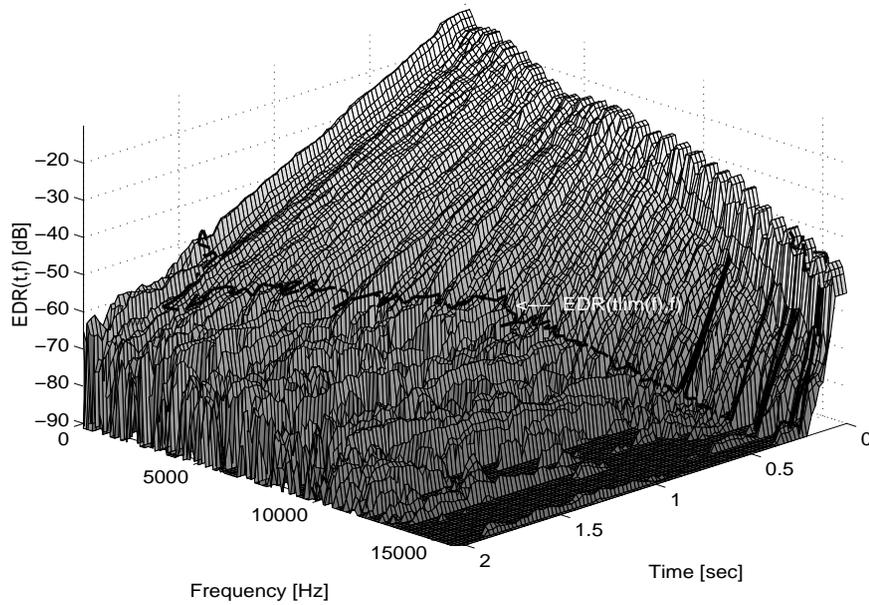


FIG. 5.12 – Calcul du relief de décroissance avec débruitage. La courbe foncée indique le temps limite d'arrivée du bruit pour chaque canal fréquentiel.

la resynthèse de la réponse débruitée.

5.1.6 Applications pratiques

Les méthodes suivantes résultent de la modélisation statistique du champ réverbéré.

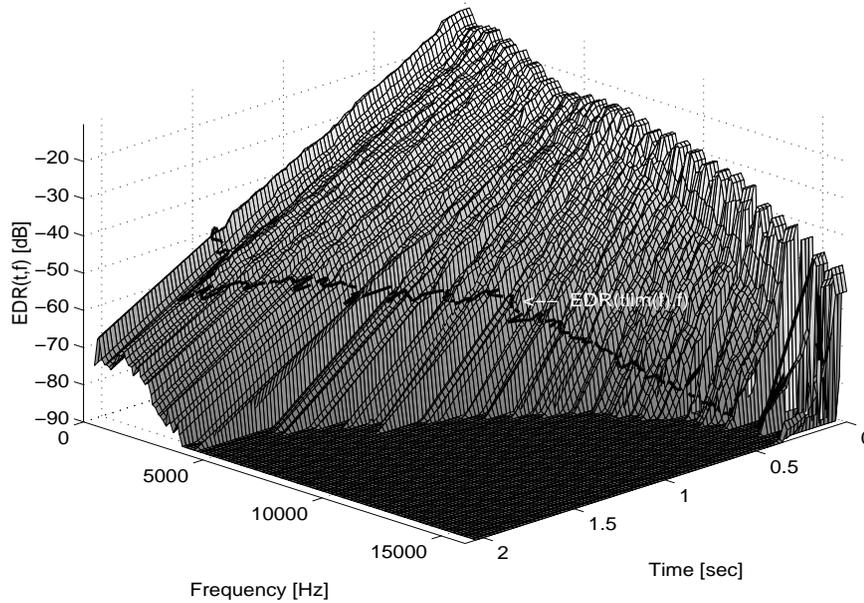


FIG. 5.13 – *Prolongement du relief de décroissance, en utilisant les paramètres $R_i(f)$ et $P_0(f)$. Ce relief peut alors être utilisé pour resynthèse de la réponse débruitée.*

Mesure de temps de réverbération

Le temps de réverbération peut être calculé à partir de réponses mesurées pour différents canaux acoustiques dans une même salle, cette quantité étant indépendante du canal considéré (placements des transducteurs, caractéristiques de directivité). Une mesure idéale est obtenue en effectuant la moyenne sur plusieurs canaux acoustiques, et ce de manière à réduire la variance sur l'estimation (*figure 5.14*).

Mesure de fonctions de transfert en champ diffus

Traditionnellement, ces mesures s'effectuent soit en chambre anéchoïque, soit en chambre réverbérante. Dans le premier cas on réalise un échantillonnage spatial autour du transducteur et la fonction de transfert en champ diffus est obtenue en effectuant la moyenne des mesures sur différentes directions (définition de la fonction de transfert en champ diffus). Dans le second cas on excite utilise un signal stationnaire pour recueillir directement la fonction de transfert (après correction du temps de réverbération).

En utilisant les propriétés du relief de décroissance il est possible d'atteindre cette caractéristique des transducteurs à partir d'un nombre limité de mesures effectuées dans une salle classique. Par moyennage de plusieurs mesures pour un même couple source récepteur, et pour plusieurs positions à l'intérieur d'une même salle (ou dans des salles différentes après application d'un facteur correctif lié au volume de la salle), il est possible d'améliorer la précision de caractérisation de ces fonctions de transfert.

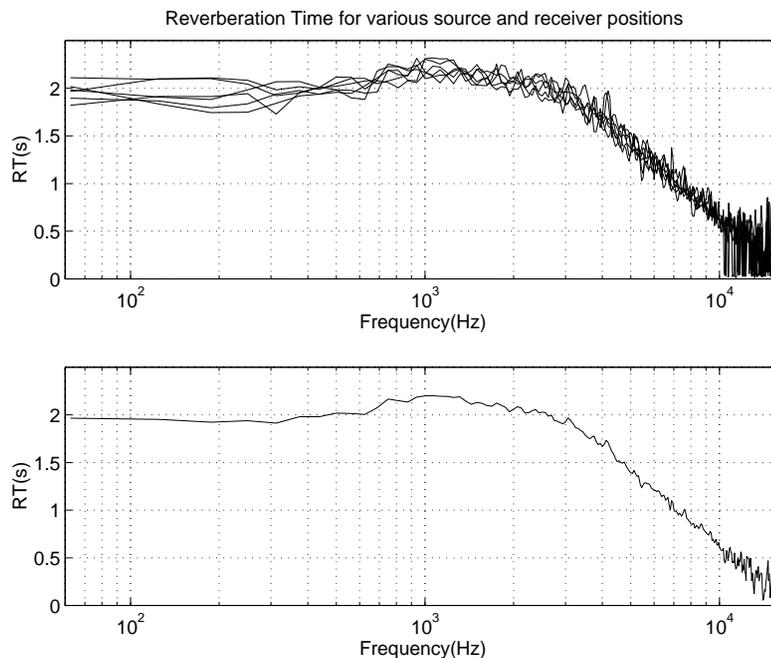


FIG. 5.14 – Mesure de temps de réverbération (Philharmonie de Berlin). Elle s'effectue en moyennant les résultats mesurés pour plusieurs canaux acoustiques (différentes positions de transducteurs, différents types de transducteurs), de manière à obtenir un lissage des résultats.

La figure 5.15 présente les résultats de telles mesures: deux transducteurs, l'un omnidirectionnel, l'autre cardioïde sont ainsi caractérisés. On retrouve un écart constant d'à peu près 5 dB entre ces deux courbes, correspondant à une différence théorique de 4.7 dB.

Débruitage de réponse impulsionnelle

L'analyse par relief de décroissance permet d'estimer l'instant d'apparition du bruit pour chaque canal de la transformée de Fourier à court-terme. Il est alors possible, connaissant les valeurs de $Tr(f)$ et de $P0(f)$ de prolonger le relief de décroissance pour chaque canal fréquentiel, et d'obtenir ainsi un relief correspondant à une réponse débruitée. La synthèse de cette réponse ne peut cependant s'effectuer à partir de la connaissance seule du relief de décroissance (absence d'informations sur la phase).

Pour cela, un bruit blanc est d'abord analysé, selon une analyse par relief de décroissance. Puis, connaissant les paramètres de la réverbération tardive de la réponse originale, l'enveloppe de cette réverbération originale est appliquée au résultat de la transformée du bruit blanc. Il est ainsi possible de recoller cette enveloppe au spectre de Fourier à court terme du signal original, puis, par transformée de Fourier inverse, et

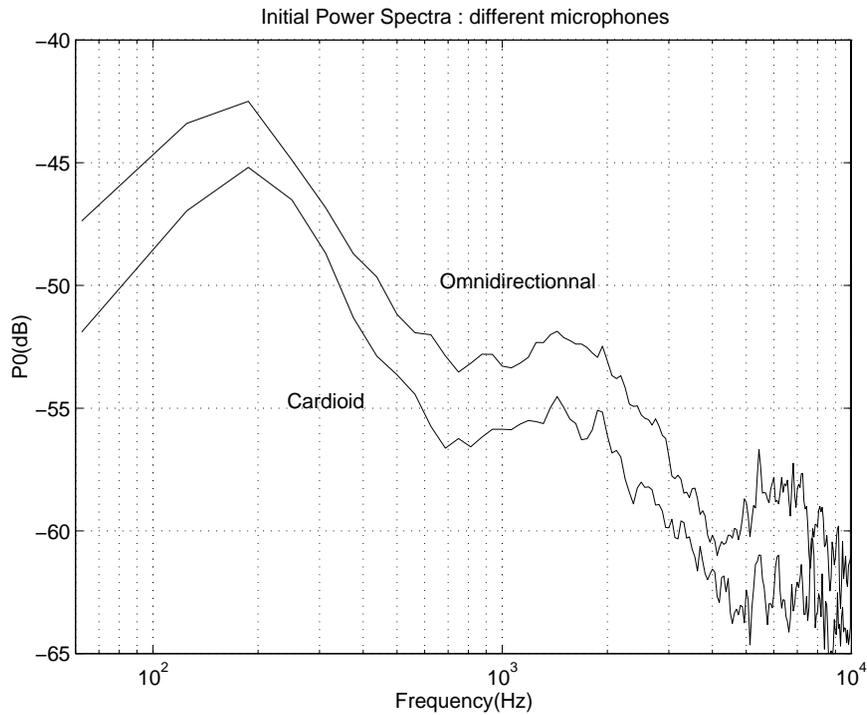


FIG. 5.15 – *Estimation de spectre de puissance initiale. Les deux courbes correspondent à des jeux de mesure effectués avec différents microphones. Le lissage des résultats s’effectue en moyennant les résultats obtenus pour plusieurs canaux acoustiques utilisant le même couple de transducteurs ; ces derniers correspondent à des positions différentes, soit dans la même salle, soit dans des salles différentes.*

overlapp-add de reconstruire une réponse débruitée³.

Ce débruitage peut être utile pour nettoyer des réponses impulsionnelles avant auralisation par convolution.

5.2 Implantation informatique

5.2.1 Librairie EDR-1.5

La méthode d’analyse par relief de décroissance a été implantée sous la forme d’une librairie Matlab : EDR-1.5. Cette librairie possède différents types de fonctions.

- Entrées/Sorties : fonctions de lecture et écriture de fichiers sons, pour différents formats de fichier (AIFF, WAV, SND, Raw Data)

3. Le fenêtrage utilisé pour cette resynthèse devra cependant obéir à certaines conditions (somme constante)

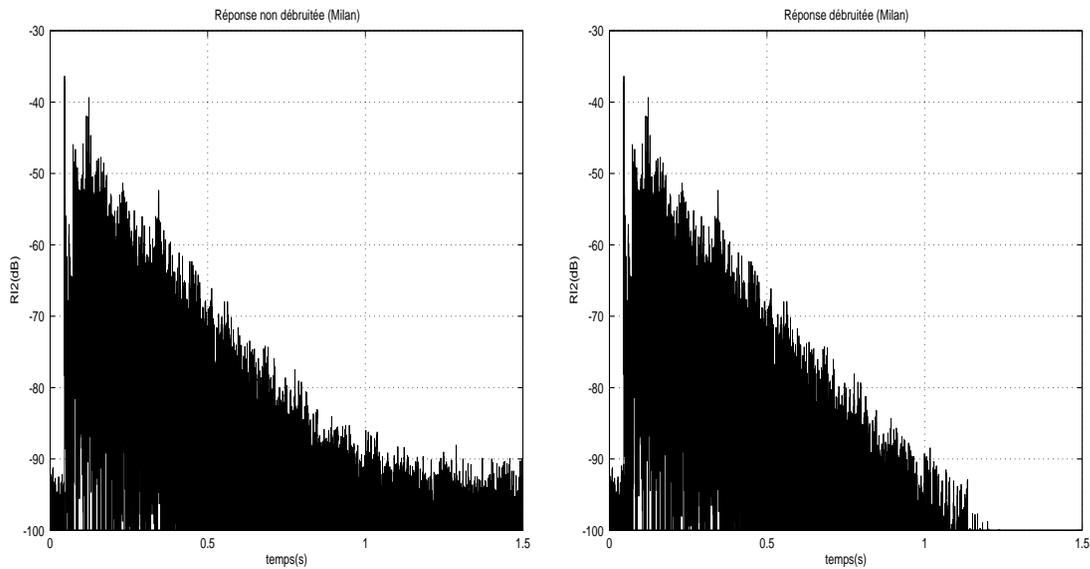


FIG. 5.16 – Débruitage et prolongement d'une réponse impulsionnelle

- Fonctions élémentaires d'analyse temps-fréquence : calcul de transformée de Fourier à court terme, calcul de courbes de décroissance intégrée et de relief de décroissance, fonctions de lissage.
- Fonctions de haut-niveau : elles permettent de réaliser des analyses complètes en une seule commande.
- Fonctions d'extraction de critères objectifs et de facteurs perceptifs.

La liste des fonctions disponibles, ainsi que des fichiers d'aide correspondant sont présentés en annexe A.

5.2.2 Applications : Réglages du Spatialisateur

Une application naturelle de l'analyse EDR a été la création d'un ensemble de réglages pour le Spatialisateur, issus de salles réelles. Afin de pouvoir présenter une interface à l'utilisateur, l'extraction de paramètres de contrôle pour une interface perceptive a été réalisée, selon les lois présentées au chapitre 3. Cette réalisation correspond au passage entre une représentation bas-niveau de type morphologique et une interface de type perceptive, si l'on se réfère au schéma final du chapitre 2. Deux séries de salles ont été considérées.

Dans le cadre d'une collaboration projet Virtuelles Raumsklangmuseum du ZKM (Zentrum für Kunst und Musik), un ensemble de lieux aux acoustiques différentes a été analysé.

Cet ensemble comprend aussi bien des salles que des espaces en plein air, et notamment :

- une salle de concert
- une église
- un salon
- une salle de bain
- un escalier (absence de son direct, le récepteur ne voit pas la source)
- un théâtre antique (présence d'un filtrage en peigne apporté par les gradins et qui apporte une coloration)
- une calanque naturelle (présence d'un écho tardif très important)

Différents cas se sont présentés : d'une part, ceux où un réglage direct après analyse EDR permettait un réglage plausible, d'autre part, ceux où une correction manuelle a été nécessaire. La validité des résultats s'est faite par comparaison des RIF initiales avec celles issues du Spatialisateur (réglé par les indices calculés), et par jugement sur la similarité auditive. Les résultats issus de l'analyse sans correction, à l'aide de la fonction *RI2Spat*, se sont montrés satisfaisant pour les salles dites classiques (salle de concert, église, salon). Le cas de la salle de bain a nécessité un réglage particulier de la densité modale pour la réverbération tardive.

La restitution des autres espaces n'a pu se faire qu'en tenant compte de l'architecture de la chaîne traitement du signal du Spatialisateur. En effet, ces salles sortant généralement du cadre psychoacoustique ayant permis l'établissement de la représentation morphologique et des facteurs perceptifs, la forme des réponses impulsionnelles n'était plus conforme avec celle décrite au chapitre 2. Le problème majeur qui est apparu concerne les limites temporelles des sections.

D'une part, celles-ci peuvent ne pas correspondre à celles établies pour la description morphologique, et une adaptation doit être effectuée pour l'analyse. De l'autre, il est des cas où le découpage temporel des réponses n'est pas en accord avec ce modèle. C'est typiquement le cas de la réponse d'une calanque où un écho apparaît à 300 ms après le son direct (*figure au chapitre 2 : 2.3*). Pour obtenir un réglage satisfaisant, il faut alors intervenir sur les paramètres de contrôle bas-niveau, afin de contraindre la forme du signal.

En effet, de part son architecture le Spatialisateur relie les différentes sections temporelles ; par exemple, les signaux issus de la synthèse des premières réflexions alimentent le cluster et la réverbération, rendant difficile le placement d'une réflexion isolée. Le contrôle par l'interface perceptive ne balaye pas l'ensemble des réponses réalisables. La

connaissance fine de ces liens autorise alors des réglages permettant de synthétiser des effets de salle hors du cadre de l'interface perceptive (*Annexe B*)⁴.

La deuxième série de salles traitées est issue d'une base de donnée sur des salles de concerts européennes. Cette base a été constituée lors d'une campagne de mesure de l'Ircam qui a eu lieu en 1991. Ont été choisies comme exemples de presets, deux salles de concert de formes différentes, un auditorium et une salle de concert. Pour chaque salle un couple semblable de canaux acoustiques a été considéré. (c'est-à-dire correspondant à deux positions identiques sur une grille de positions).

Une interface a été réalisée permettant de manipuler ces réglages. Elle permet de visualiser la forme de la salle ainsi que le canal acoustique considéré. Si l'on se réfère au schéma final du chapitre 1, ce cas correspond à celui de l'exploitation d'une base de données (sans interpolation entre les places). On peut ainsi se rendre compte des variations d'acoustique entre différentes salles pour des mêmes positions de source et récepteur ; en outre le fait de comparer des canaux identiques pour divers espaces permet de constater que la réalisation du même trajet physique donne lieu à des variations perceptives différentes selon les salles..

Enfin il faut noter que, contrairement au cas précédent, l'utilisation directe des résultats issus de l'analyse EDR s'est montrée satisfaisante. Ceci est à mettre en corrélation avec le fait que c'est cet ensemble de salles qui a servi de base à la démarche d'établissement des facteurs perceptifs.

5.3 Resynthèse arbitraire d'une réponse impulsionnelle

Les paragraphes précédents ont présenté des méthodes de conversion entre une représentation bas-niveau de type RIF et un type de représentation morphologique.

Sans entrer dans les détails d'une synthèse complète permettant d'effectuer le passage inverse, il est utile de préciser certains points. Tout d'abord le contexte d'utilisation de telles méthodes : si l'on se réfère au synoptique (*figure 2.10*) d'un système de manipulation de l'effet de salle, l'application principale semble être la resynthèse d'une réponse à partir d'une interface perceptive pour commander un moteur de rendu sonore basé sur la convolution. Deux cas peuvent alors se présenter, qui vont influencer sur le choix des méthodes. Soit ce passage doit être effectué en temps réel, auquel cas l'ordre des filtres utilisés restera relativement faible, soit il s'agit de temps différé, et c'est l'exactitude qui l'emportera⁵.

4. Par exemple, il est ainsi possible de faire commencer le cluster après le phénomène de réverbération, ce qui sort du cadre perceptif.

5. Il faut cependant remarquer qu'un moteur de rendu basé sur la synthèse générique de l'effet de

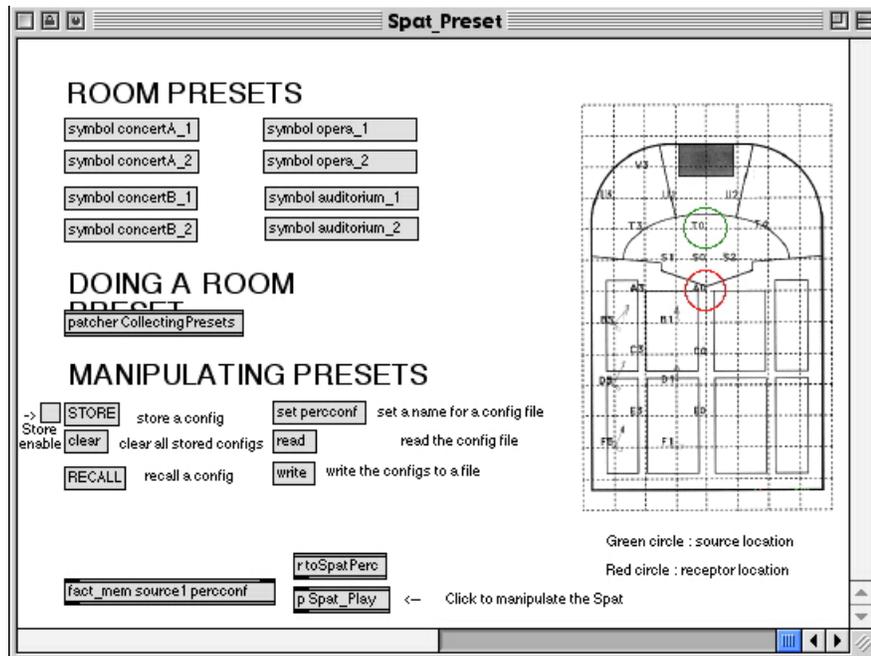


FIG. 5.17 – Ensemble de réglages du Spatialisateur

On précise que l'on s'intéresse principalement à la resynthèse d'une réponse monophonique.

5.3.1 Contributions discrètes

Cette dénomination comprend le son direct et les premières réflexions. Une méthode consiste à utiliser des filtres à réponse impulsionnelle finie, dont la fonction de transfert reproduit les caractéristiques fréquentielles de ces parties. Une contrainte à imposer est que la longueur de la réponse soit inférieure au temps de la section considéré, permettant ainsi une indépendance dans la synthèse de ces deux parties.

Un problème se pose pour la synthèse des premières réflexions. En effet soit celles-ci sont décrites précisément, soit elles sont décrites par une énergie globale. Dans ce dernier cas, il convient d'abord de donner un nombre de réflexions et trois lois pour la répartition temporelle, spatiales et énergétique de ces réflexions; le choix de ces deux quantités est purement arbitraire.

L'utilisation de filtres à réponse impulsionnelle finie semble à nouveau s'imposer afin d'éviter un recouvrement des différentes sections temporelles. Enfin, on notera que des problèmes sur la phase sont également soumis à une résolution arbitraire.

Malgré ces choix à effectuer il est important de se rappeler les effets perceptifs de ces réflexions. Cela permettra d'avoir une plus grande souplesse dans la synthèse. Par salle, constitue en soit une méthode de synthèse d'une réponse impulsionnelle

exemple, il est ainsi possible de contraindre la répartition spatiale en interdisant aux réflexions de provenir de l'arrière (dans ce cas il y aurait intégration au champ tardif).

5.3.2 Synthèse de la réverbération tardive

La première méthode utilisable est une prolongation des techniques de débruitage de la réponse impulsionnelle. Elle consiste à resynthétiser un signal de réverbération à partir de l'analyse d'un bruit auquel est appliquée une enveloppe de décroissance correspondant à la réverbération désirée. Bien que rigoureuse cette technique ne semble *a priori* pas praticable pour des utilisations en temps réels. Une technique semblable, plus rapide, est issue des travaux de Moorer, mais utilise un banc de filtres afin de découper le bruit en signaux à bandes étroites avant d'appliquer l'enveloppe exponentielle et d'effectuer la somme des signaux résultants.

Un certain nombre de tests sont à effectuer afin de vérifier la validité psychoacoustique des résultats obtenus, notamment pour le choix de la loi du temps de réverbération.

5.3.3 Le cluster

Peu d'études ont eu lieu sur l'influence de la distribution des réflexions au sein de la section "cluster". Jusqu'à présent son importance perceptive a été étudiée principalement en fonction de son énergie et de sa répartition spatiale. La démarche adoptée dans le Spatialisateur est de synthétiser une distribution aléatoire.

Chapitre 6

Simulation géométrique

L'interaction et le contrôle de l'effet de salle avec un modèle géométrique est souvent recherchée par les utilisateurs, et les programmeurs, en raison du caractère a priori intuitif de ce mode. L'application directe de méthodes d'acoustique prévisionnelle pour un calcul exhaustif de la réponse impulsionnelle dépasse cependant encore les possibilités de calcul des processeurs, pour une simulation temps-réel ; de plus, si l'on se réfère au chapitre 2, toutes ses étapes n'apportent pas une information auditivement pertinente.

L'examen des résultats de la psychoacoustique permet de simplifier cette approche, notamment pour le calcul du champ réverbéré, et offre un cadre pour une approche paramétrique à ce problème. Une approche décomposant l'effet de salle en sections temporelles est alors proposée, et différentes méthodes peuvent être utilisées. Le précalcul de certains paramètres décrivant le comportement acoustique de la salle permet en outre de diminuer le coût de calcul pour le rafraîchissement des paramètres de contrôle du moteur de rendu audio lors des déplacements des transducteurs.

6.1 Approche générale

6.1.1 Cadre de l'acoustique prévisionnelle

Le cadre classique de l'acoustique prévisionnelle met en œuvre des algorithmes de différentes sortes : on citera ainsi les méthodes type rayons ou sources-images, basées sur une approche géométrique avec hypothèse de réflexions spéculaires, les méthodes de radiosité pour la prise en compte de la diffusion des parois, et les méthodes de simulation du champs tardif (théorie de Sabine, radiosité) ; une présentation générale en a été effectué au chapitre 2.2. Ces méthodes possèdent cependant des limitations. Celles-ci peuvent être d'origines différentes :

- Elles peuvent être en rapport à la nature des phénomènes modélisés ; par exemple une méthode de type rayons ou sources-images considère en général un modèle de réflexion spéculaire, tandis qu'une modélisation de type radiosité suppose que les parois sont parfaitement diffuses. Ces deux hypothèses sont représentatives de comportements extrêmes du processus de réflexion. La réalité nécessite au moins la description de ce comportement comme combinaison pondérée de ces deux modèles [CB99] ou un modèle de radiosité généralisé [Tsi98].
- Elles peuvent être en rapport aux techniques algorithmiques utilisées pour modéliser ces phénomènes et comprennent autant les erreurs dues à la finesse physique de l'algorithme, que les erreurs inhérentes au mode de calcul choisi. Par exemple, une méthode type rayons peut commettre des erreurs d'estimations dans le signal

final calculé, dans la mesure où la réduction de l'onde à son vecteur d'onde (la direction du rayons) est imparfaite. Le test utilisé pour vérifier la contribution d'un rayon au signal final participe aussi à l'erreur finale.

En pratique, un logiciel d'acoustique prévisionnelle est souvent basé sur un modèle qualifié d'"hybride", afin de tirer parti des avantages de différentes méthodes. Ce regroupement s'effectue en général autour d'un même algorithme de base, auquel viennent s'ajouter des briques permettant de pallier certaines limitations, prendre en compte des phénomènes physiques supplémentaires, ou traiter des cas particuliers. Ainsi une méthode des rayons utilisera une méthode des sources-images pour vérifier la validité dans le calcul des premières réflexions, ou bien créera artificiellement un ensemble de sources secondaires sur les parois afin de modéliser la diffusion.

Enfin il faut noter que l'implantation informatique de ces logiciels est, le plus souvent, liée au caractère statique de l'environnement modélisé. Les données d'entrées (salle, transducteurs) sont fixées au départ et ne changent pas au cours de l'exécution du programme. Plusieurs conséquences en découlent :

- L'exécution a lieu dans un cadre particulier dédié à l'acoustique prévisionnelle : ils peuvent donc "monopoliser" la machine, autant pour la puissance de calcul que pour l'utilisation de la mémoire.
- La prise en compte du déroulement temporel s'effectue selon plusieurs échelles : on considérera ainsi la notion de temps absolu, mais aussi celle de l'ordre d'une méthode, qui adresse généralement des événements temporels non successifs : par exemple lors du calcul d'un réseau de sources-images, il est possible que des sources à un ordre donné, soit plus éloignées du récepteur que d'autres placées à l'ordre suivant. L'échelle de l'ordre n'est alors pas en relation linéaire avec celle du temps, ni même monotone.
- Le parcours du temps ne s'effectue pas de manière monotone. Par exemple, dans le cas de l'utilisation d'une méthode des rayons, le parcours de chaque entité est suivi jusqu'à un temps limite avant de passer au rayon suivant. Les logiciels peuvent également effectuer plusieurs passes pour le calcul d'une réponse impulsionnelle, repartant à chaque fois d'un temps inférieur à celui de la fin d'une passe

Un moyen de pallier ces problèmes de déroulement temporel serait d'utiliser des méthodes permettant un maximum de précalcul et une exploitation ultérieure de ces résultats au fur et à mesure du déroulement temporel lors de déplacement du récepteur, et si possible des sources. Un problème est alors celui du stockage en mémoire de ces informations [Tsi98]. A titre d'exemple, les méthodes de radiosité et des sources-images

sont de bons candidats à cette approche, permettant même une utilisation conjointe des deux algorithmes. En particulier, dans la méthode de radiosité, l'état de la salle au cours du temps peut être consigné sur les parois de la salle, seule l'observation depuis le récepteur doit être recalculé. De plus, quelle que soit la nature de l'excitation des parois, les échanges entre parois peuvent être également consignés dans une matrice. Ainsi, en cas de déplacement ou changement d'orientation des sources la prédiction de l'état de la salle requiert un recalcul du vecteur d'excitation des parois puis la multiplication de ce vecteur par la matrice d'échange.

6.1.2 Approche paramétrique

Une transposition directe de l'approche traditionnelle de l'acoustique prévisionnelle au traitement temps réel semble alors inadaptée. Par opposition au traitement global d'un logiciel d'acoustique prévisionnelle on est conduit à adopter une approche paramétrique du phénomène d'effet de salle. Par approche paramétrique, on entendra une approche qui nous permet de séparer l'effet de salle en une série de modules indépendants et contrôlables aisément. Chacun de ces modules adresse une composante particulière de la réponse impulsionnelle, et fournit, selon l'état du système la quantité d'information nécessaire pour modéliser cette composante de manière perceptivement pertinente. Cette information est indépendante du moteur de rendu utilisé pour auraliser l'effet de salle ; elle est considérée comme un paramètre de contrôle et non un signal d'entrée. A titre de comparaison, si l'on désire communiquer une information de localisation au moteur de rendu, une approche paramétrique la passera sous formes de deux angles (azimut, élévation) et non sous la forme de signaux, dont le nombre dépend de l'encodage utilisé pour la restitution spatiale.

Enfin cette approche doit être complétée par l'utilisation d'algorithme adressant chacun une composante particulière de l'effet de salle et permettant de transposer une part important des coûts de calcul vers une étape de précalcul. Ces résultats statiques seront exploités dynamiquement lors de l'évolution de la scène sonore, au besoin de manière concurrentielle.¹

6.1.3 Prise en compte des résultats de la physique

D'un point de vue physique, il en découle une séparation de l'effet de salle en plusieurs modules : un premier consacré au calcul de la contribution directe, un second permet-

1. On notera également qu'une même entité peut être stockée sous des formes différentes : par exemple, dans le cas de la discrétisation d'une salle sous forme de facettes, il peut être utile de stocker le modèle selon deux résolutions. L'une adaptée à des méthodes géométriques et réduisant au maximum le nombre de surfaces constitutives d'un même plan, l'autre adaptée à la propagation de l'énergie diffuse et augmentant la résolution

tant d'estimer les contributions précoces spéculaires et diffuses des parois et un dernier permettant d'estimer les paramètres liés au processus exponentiellement décroissant. La gestion de ces modules peut être adaptée en fonction des données d'entrée disponibles ou de la puissance de calcul disponible. A titre d'exemple, le calcul du niveau du champ réverbéré lors d'un effet de distance peut ainsi présenter plusieurs niveaux.

- selon un modèle statistique modulant le niveau du champ réverbéré,
- en tenant compte uniquement de la contribution du son direct vers la salle (exploitation indirecte d'une méthode de radiosit ,)
- en tenant compte  galement des contributions des r flexions, avant exploitation par une m thode indirecte,
- par calcul complet de tout le champ r verb r , it ration sur les sources images, rayons ou radiosit  directe.

Enfin, on notera que le calcul des contributions pr coces des parois (jusqu'au r gime de d croissance exponentiel) peut lui m me faire appel   deux modules traitant des effets physiques diff rents (un adapt  aux contributions sp culaires, l'autre aux contributions diffuses). Un mod le mixte exploitant des donn es pr calcul es est  galement possible. A titre d'exemple on peut examiner l'effet d'un d placement du r cepteur dans un sc ne virtuelle sonore. La source est consid r e comme fixe. En ce qui concerne les contributions discr tes, cette approche param trique va nous conduire   chercher une m thode permettant de r duire au maximum le temps de calcul et   consid rer la salle comme un syst me interrogeable par le r cepteur. Si l'on utilise une m thode des rayons, il faudrait conserver en m moire le trajet de tous les rayons et effectuer une interrogation sur l'ensemble de ces  l ments. Le choix d'une m thode des sources-images est plus appropri  : la source  tant fixe, il est possible de pr calculer le r seau de sources-images. Un d placement du r cepteur revient alors   effectuer une interrogation de ce r seau, sur la contribution effective de chaque sources-images.

Pour la contribution diffuse, l'emploi d'une mod lisation par processus markovien permet de pr calculer des matrices repr sentant les  changes d' nergie entre les surfaces. Elles sont exploitables selon deux modes : le mode direct permettant la propagation it rative (progression temporelle) de l' nergie dans la salle (utilisable conjointement avec le module de sources-images), et le mode inverse permettant de calculer la distribution des intensit s et le temps de r verb ration associ  lors du processus de d croissance exponentielle.

6.1.4 Prise en compte des résultats de la psychoacoustique

La prise en compte des résultats de la psychoacoustique nous a permis de voir que le système auditif réagit à un nombre limité de caractéristiques d'une réponse impulsionnelle, ou du moins qu'il est possible de dégager une hiérarchie entre les différents aspects de la perception. Ces caractéristiques correspondent aux facteurs perceptifs présentés au chapitre 3. Leur calcul est relié à la répartition temporelle et fréquentielle de l'énergie présente dans la réponse impulsionnelle.

Les principaux résultats exploitables sont basés sur cette segmentation de la réponse impulsionnelle. On peut alors examiner la quantité et la forme de l'information apportée par chacune des sections.

- L'information apportée par l'onde directe est déterminante pour la localisation de la source. La perception de son niveau et de son spectre est par contre modulée par la contribution des premières réflexions
- L'information apportée par la réverbération peut se résumer à la connaissance de deux paramètres : le temps de réverbération, et le niveau d'énergie².
- L'information apportée par les premières réflexions et le cluster va agir simultanément sur les niveaux associés à la source et à l'effet de salle et sur des effets spatiaux. Elle ne peut donc s'évaluer qu'*a posteriori*, et n'est pas prévisible telle qu'elle. Le calcul complet du signal ou de l'échogramme correspondant est alors nécessaire. Cependant il est envisageable d'exploiter cet échogramme pour extraire des paramètres à transmettre au moteur de rendu sonore

Enfin il serait intéressant de tenir compte du temps de réponse de l'oreille lors d'une modification de l'effet de salle. Par exemple, si l'on suppose une situation où le récepteur se déplace dans un espace virtuel, la question de la vitesse de rafraîchissement de l'information lui arrivant doit être examinée.

Lors d'une modification de la source, et si l'on veut respecter la réalité, on ne peut cependant rafraîchir instantanément l'ensemble des paramètres correspondant à la contribution de la salle. En effet le signal émis par cette source à un instant t ne doit être traité qu'en fonction de la réponse de la salle à l'instant t , et ce jusqu'à l'extinction de l'effet de salle associé. A titre d'exemple, si l'on suppose la disparition d'une source à un instant donné, on ne peut instantanément stopper les composantes de l'effet de salle, de façon à laisser l'effet de salle s'éteindre naturellement. Inversement, si un récepteur entre

2. On constatera cependant que ce résultat est directement lié au cadre de ces études perceptives. Des paramètres supplémentaires (densité modale, densité temporelle) sont nécessaires pour décrire d'autres types d'espaces.

dans une salle à un instant donné, il doit être capable de recevoir les contributions du passé des sources présentes dans cette salle à concurrence de la durée de réverbération.

6.1.5 Intérêts attendus de la paramétrisation

La prise en compte conjointe d'une paramétrisation perceptive et physique permet d'aboutir aux résultats suivants (résumé à la figure 6.1).

- Les paramètres nécessaires au contrôle du moteur de rendu sont indépendants de la nature de celui-ci. Leur forme répond à la fois à des critères de pertinence acoustique, et de temps de calcul.
- Les informations transmises au moteur de rendu étant sous forme de paramètres et non de réponse impulsionnelle, la transmission de ces informations est très rapide. En particulier pour des applications en réseau, on peut supposer qu'un unique serveur réalise la simulation pour un ensemble de récepteurs et envoie à chacun les paramètres correspondant à sa position spécifique dans la salle.
- Certaines options de rendu sonore peuvent être déléguées au moteur de rendu, ce qui peut être avantageux lors des situations dynamiques. Par exemple, en première approximation on peut supposer que l'effet doppler, particulièrement important pour la contribution du son direct, peut être négligé dès que le champ sonore est constitué d'une superposition de nombreux trajets dont la moyenne des variations tendra vers zéro. Pour le son direct on peut alors implanter le retard de propagation sous la forme d'un filtre à retard fractionnaire pour lequel on pourra choisir une implantation optimisée en terme de structure ou d'ordre de filtre [NNi96]
- Le système de calcul des paramètres met en œuvre des algorithmes de nature indépendante pour adresser les différents phénomènes physiques
- Le poids de chacun de ces algorithmes dans le calcul effectif des paramètres de sortie dépend du support temporel adressé, par exemple relativement à sa position par rapport à une segmentation d'ordre psychoacoustique de la réponse impulsionnelle

6.2 Calcul du son direct

6.2.1 Calcul théorique

Il correspond à la théorie de propagation en champ libre. La source sonore est modélisée comme une sphère pulsante [Ros86]. L'intensité acoustique à une distance r , s'obtient en supposant que la source émette des ondes sphériques.

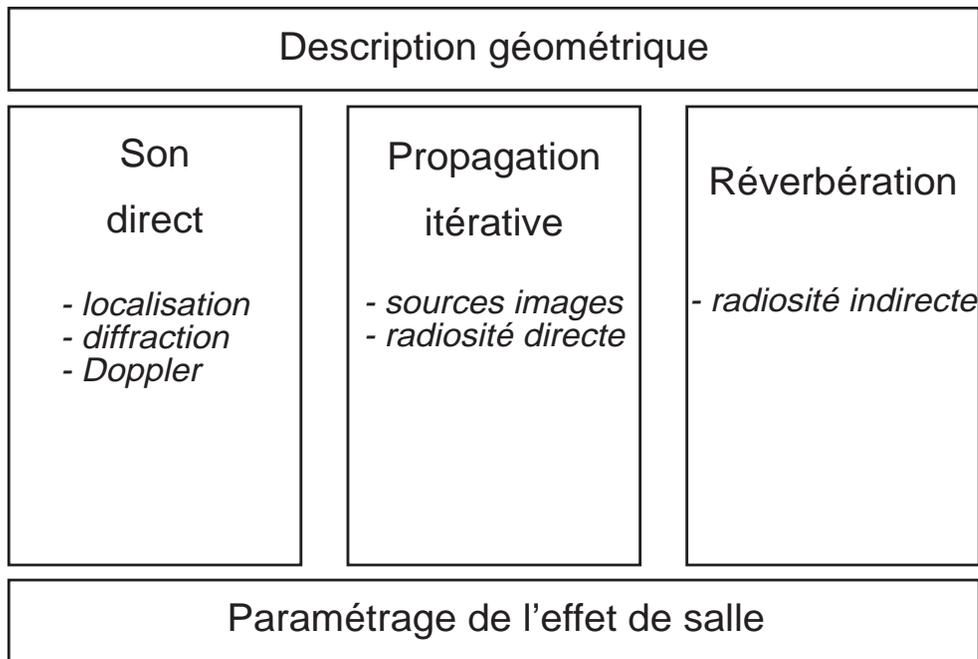


FIG. 6.1 – Approche paramétrique pour un système de traitement temps réel de l'information contenue dans un effet de salle

On a alors :

$$I_a(r) = \Gamma(\theta, \phi) \frac{I_{source}}{4\pi r^2} \quad (6.1)$$

avec $\Gamma(\theta)$ directivité selon θ, ϕ (en supposant une direction de référence). L'expression directe en décibels donne :

$$L_p(r) = L_{I_{source}} + 10 \log(\Gamma(\theta, \phi)) - 20 \log(r) - 11 \quad (6.2)$$

6.2.2 Absorption du milieu

Le milieu considéré ici est l'air. L'absorption dépend de la fréquence considérée ; son effet est celui d'un filtre passe-bas ; les caractéristiques de ce filtre dépendent notamment du taux d'humidité relative. Elle est due à trois phénomènes physiques :

- la viscosité donnant lieu à une transformation de l'énergie acoustique en chaleur
- les échanges thermiques entre particules voisines, et également du à la présence possible de rayonnement électromagnétiques.
- les phénomènes de relaxation moléculaire, impliquant des pertes.

Expérimentalement, on constate que l'absorption provoque une décroissance supplémentaire de l'intensité acoustique selon une loi exponentielle décroissant, de facteur a , appelé coefficient d'absorption. La valeur de ce coefficient dépend de la fréquence et de l'humidité relative du milieu. Pour une humidité relative h comprise entre 20% et 80%, une modélisation par une loi hyperbolique peut être utilisé [Jou97] :

$$a(h) = 5.5 \frac{50}{h} \left(\frac{f}{1000}\right)^{1.7} 10^{-4} \quad (6.3)$$

L'intégration de cette formule entre deux valeurs, permet d'obtenir un coefficient d'absorption moyen sur une bande de fréquence, permettant ainsi de calculer des valeurs de OD sur plusieurs bandes (basse, médium, haute).

6.2.3 Occlusion

Il s'agit du cas où une surface se trouve entre la source et le récepteur. Une première approche consiste à annuler la contribution de l'onde directe. Il peut également être choisie de mettre en œuvre un algorithme modélisant la diffraction par un modèle ondulatoire [Tsi98] ou approché [CB99], modélisant le phénomène d'absorption sous incidence rasante.

6.2.4 Implantation

Elle est une transposition directe des formules physiques. En général, le coût de calcul correspondant sera assez faible; ce résultat théorique ne peut cependant être utilisé tel quel dans une implémentation. En effet, il est d'abord nécessaire de vérifier qu'aucune paroi de la salle n'est présente entre la source et le récepteur, ce qui a pour conséquence d'annuler la contribution de l'onde directe ; le test de visibilité est effectué pour chaque paroi (mais peut s'arrêter dès qu'une paroi a été trouvée comme masquante³, et sa durée augmentera donc avec le nombre de parois de la salle. Ce test ne tenant pas compte des propriétés acoustiques des parois, il peut alors être intéressant de regrouper les parois jointes, et de caractéristiques différentes, afin de diminuer le nombre de parois à tester. Le temps global de calcul pour les valeurs de OD est directement lié au temps de cette opération que l'on pourrait qualifier, par analogie avec le graphisme 3D, de "détection de collision".

6.3 Calcul des contributions discrètes

On détaillera dans ce paragraphe les deux méthodes présentées au chapitre 1, c'est-à-dire la méthode des rayons, ainsi que ses dérivées et la méthode des sources-images.

3. Un traitement sera effectué dans le cas où l'on considère le phénomène de diffraction.

6.3.1 Méthodes utilisées en acoustique prévisionnelle

Méthode des rayons

Cette méthode ayant été présentée rapidement au chapitre 1, une présentation plus approfondie est effectuée ici. Un des problèmes majeurs de la méthode des rayons est dû à la modélisation des transducteurs. Au niveau de la source, se pose la problème d'effectuer une discrétisation de l'espace permettant de modéliser avec suffisamment de précision son émission. Au niveau du récepteur se pose celui du critère permettant l'imputation d'un rayon sur son échogramme et qui est relié à la taille du transducteur).

Problèmes liés au récepteur

Le récepteur est modélisé par un secteur de petites dimensions. Ce choix a pour conséquence un certain nombre de limites de cette méthode. En effet, la taille de celui-ci est déterminante sur la quantité d'énergie imputée sur l'échogramme au point récepteur. Trop petite, la procédure risque de s'arrêter trop rapidement (la distance entre 2 rayons adjacents devient supérieure aux dimensions de la source), trop grande il y a possibilité de surestimation de valeurs de l'échogramme, par imputation de l'énergie d'un nombre trop élevé de rayons. C'est pourquoi, certaines implémentations vérifient la validité d'une source par calcul de la source-image correspondante, et vérifie son unicité. Il est possible de pallier cet inconvénient en augmentant le nombre de rayons tirés, mais cela induit une augmentation des coûts de calcul. De plus se pose le problème de la répartition spatiale des rayons tirés : de façon rigoureuse, ceux-ci doivent être équirépartis sur la surface de la source⁴ ; en pratique il n'est pas possible de couvrir l'intégralité de la surface de la source.

Méthode des cônes

Elle diffère de la méthode des rayons par le test de réception, et leur forme à l'émission. Ceux-ci ne sont plus considérés comme rectilignes, mais représente un cône, siège du parcours d'une portion d'onde sphérique émise par une source ponctuelle dans un angle solide $d\Omega$. Le récepteur est considéré comme ponctuel et capte un écho à chaque fois qu'un rayon passe à une distance inférieure au rayon du cône associé. La procédure est stoppée lorsque le rayon des cônes devient supérieur aux dimensions de la salle.

Cette méthode est équivalente à une méthode des rayons ou la taille du récepteur grandirait au cours du temps. Cependant, la discrétisation de l'espace en cônes étant impossible (si l'on veut recouvrir tout l'espace, il y a recouvrement entre les cônes), certaines approximations sur le rayonnement des sources sont inhérentes à cette méthode. Ce

4. Dans le cas d'une source directive il faudra de plus effectuer une pondération de l'énergie des rayons en fonction de leur direction

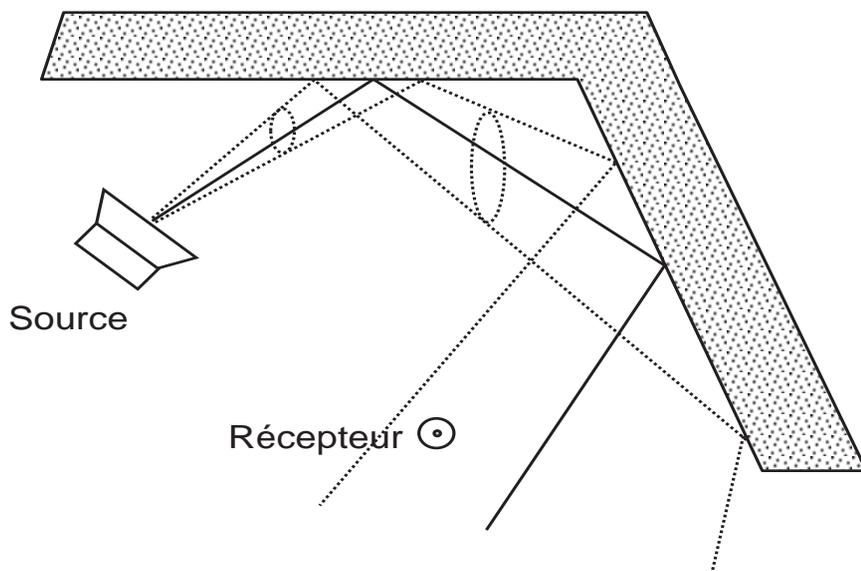


FIG. 6.2 – *Méthode des rayons, méthode des cônes*

problème de recouvrement a poussé à considérer d'autres formes pour décrire l'angle solide entourant la source ; on peut ainsi citer la méthode des "pyramides", développée par Angelo Farina [Far95a], [Far95b].

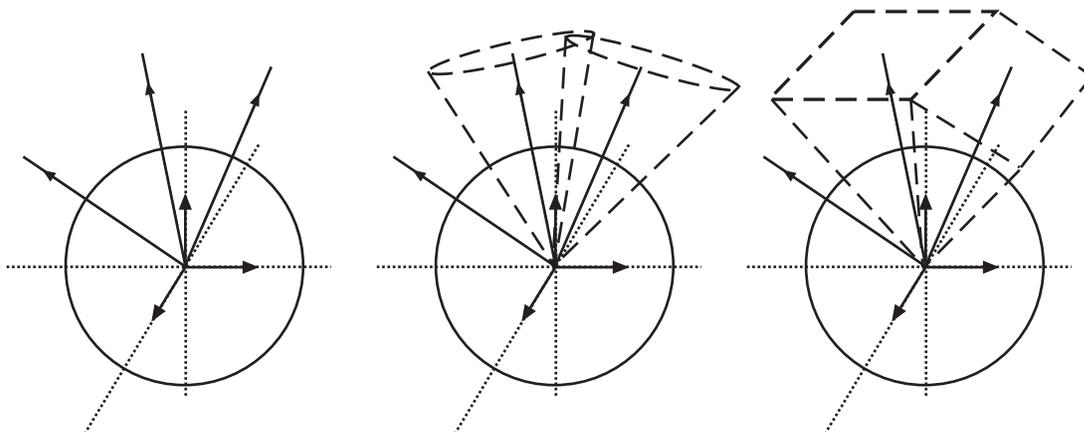


FIG. 6.3 – *Discretisation spatiale de l'énergie issue de la source : rayons, cônes, pyramides. Dans la cas de l'utilisation d'une méthode des cônes, il y a recouvrement des angles solides. Il faut alors pondérer l'énergie apportée par un cône, en fonction de sa distance à son axe.*

Méthode des sources-images

Principe

On s'intéresse ici au cas d'une salle de forme quelconque [Bor84]. Par définition une source-image d'ordre p , correspond à un rayon issu de la source et frappant le récepteur après p réflexions sur les parois de la salle. On peut alors construire, par symétries successives, l'ensemble des sources-images correspondantes à une position de la source ; les sources-images créées à un ordre, à partir d'une source-image initiale sont appelés "enfants" de la source initiale (on parlera également de "progéniture" de la source initiale). Le nombre de sources-images, dans une salle de S parois, pour un ordre donné p , est égal à

$$N_{Sp} = \frac{S}{S-2} [(S-1)^p - 1] \quad (6.4)$$

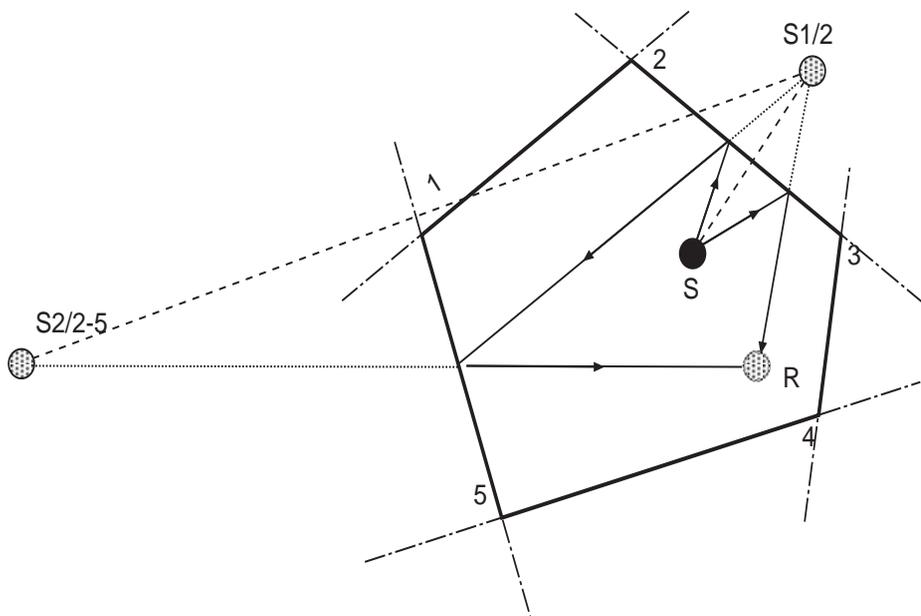


FIG. 6.4 – Construction du réseau de sources-images : $S1/2$, image d'ordre 1 de la source par rapport à la paroi 2, donne naissance à l'image d'ordre 2, $S2/2-5$, par symétrie par rapport à la paroi 5.

Examen des sources-images

L'examen du cas de la figure précédente fait cependant apparaître le problème de la contribution réelle d'une source-image, pour une position de récepteur donné. Différentes

examens sont à effectuer [Bor84].

Définition 4 Si l'on suppose que la salle est composée de parois recouvertes de miroir sur leur face interne, une sources-images est considérée comme invalide si elle se situe du côté réfléchissant de la paroi qui l'a créée. Une sources-images invalide ne possède pas de progéniture (6.5).

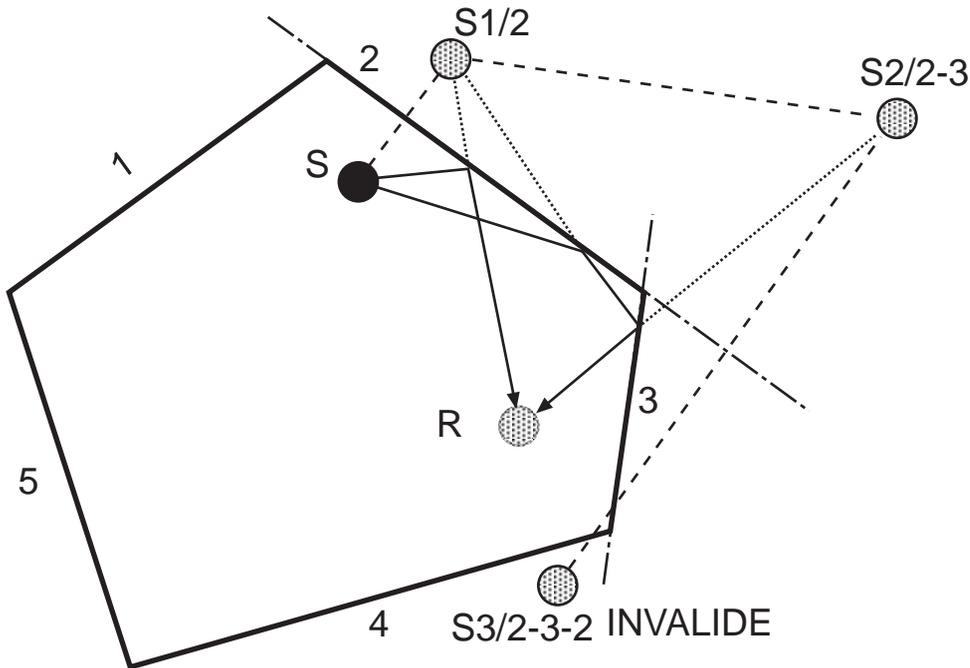


FIG. 6.5 – $S_{3/2-3-2}$, ne peut donner naissance à un trajet vers l'intérieur de la salle, dans la mesure où elle se trouve du même côté de la paroi 2 que la salle (6.6).

Définition 5 Une sources-images est considérée comme invisible, si le segment la reliant au récepteur possède son point d'intersection à l'extérieur du domaine de la paroi. Une source invisible possède une progéniture.

Cette condition ne suffit cependant pas pour des sources supérieures au premier ordre. En effet, suivant la définition d'une sources-images d'ordre p , il convient de vérifier que celle-ci correspond réellement au trajet d'un rayon ayant effectué p réflexions, et donc de vérifier la validité de tout son trajet.

Enfin, un dernier cas à considérer est celui d'une salle de forme non convexe ; il convient en plus de vérifier qu'aucune surface rentrante n'obstrue un trajet.

Le nombre réel de sources-images à considérer se réduit alors considérablement. A titre informatif, il est possible d'estimer le nombre moyen de sources visibles entre les instants

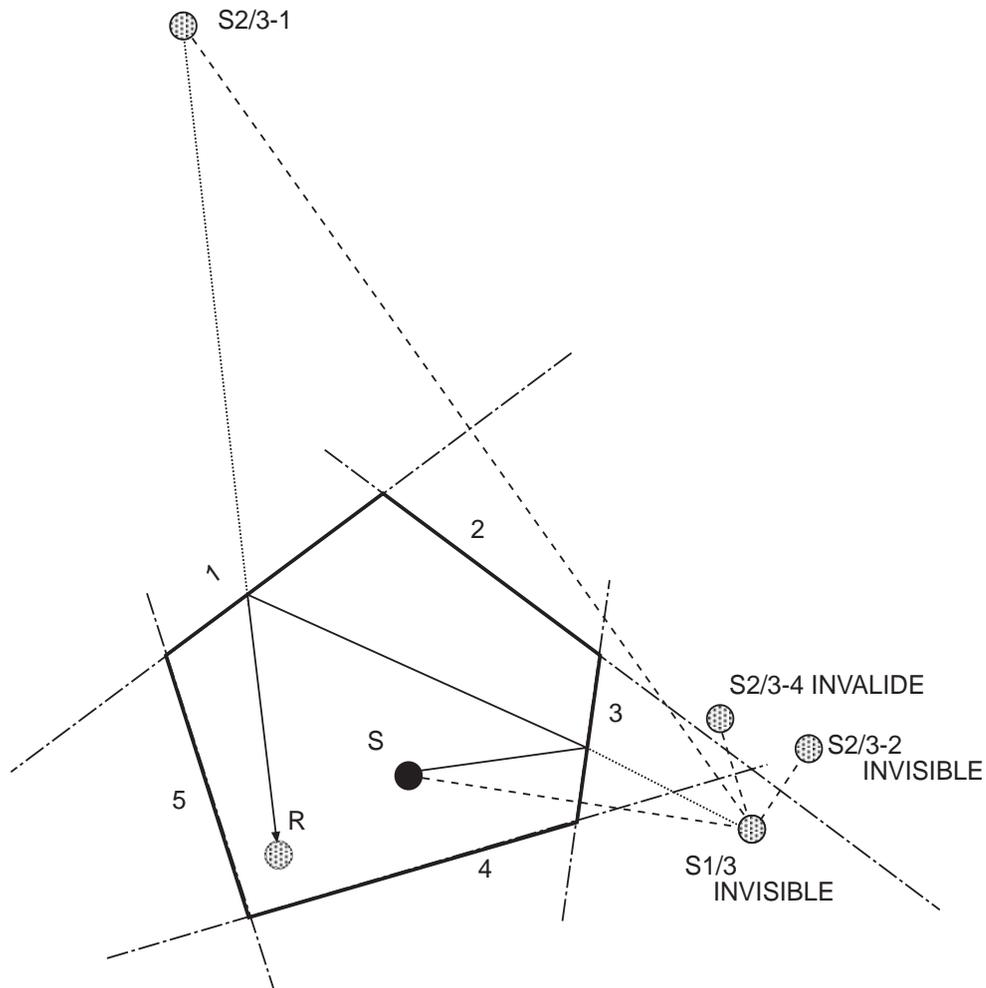


FIG. 6.6 – $S1/3$, est invisible ; parmi ses enfants, $S2/3-1$ est visible, $S2/3-2$ est invisible, et $S2/3-4$ est invalide

t et $t + \Delta t$, et ce sous l'hypothèse de champ diffus. Bien que cette hypothèse ne soit pas valide dans le cas présent, la donnée de ce nombre permet de trouver des rapports de l'ordre de 10^n avec $n > 10$ pour des réponses d'à peu près $0.5 s$, ce qui incite à penser que même sur le début de la réponse le nombre de sources-images sera considérablement diminué.

6.3.2 Formulation mathématique

Le calcul de l'intensité total au récepteur se fait par sommation au récepteur des signaux provenant des sources-images. Soit une source image S_{P_k} , d'ordre P , se situant à une distance r_{P_k} du récepteur. Sa contribution énergétique peut s'écrire

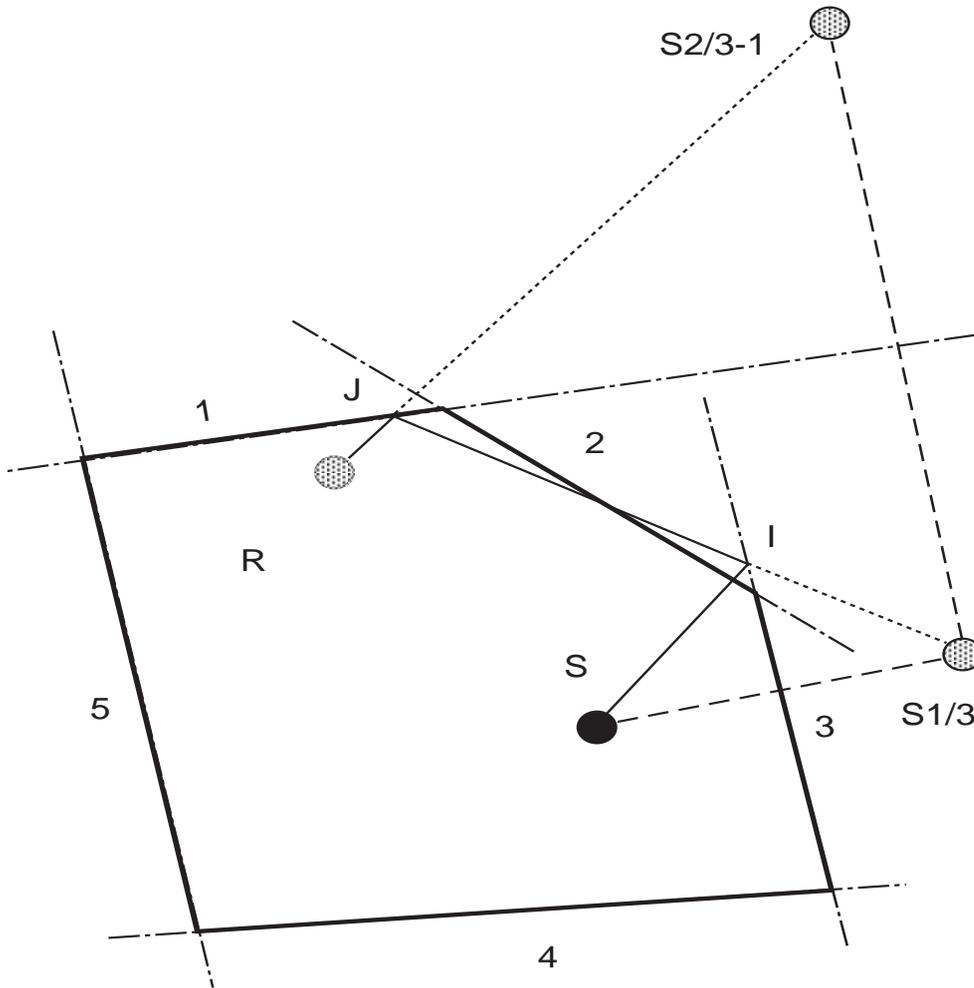


FIG. 6.7 – Bien que $S_{2/3-1}$ soit visible, la source est ne peut être prise en compte car elle ne correspond pas à un trajet réel.

$$P_{S_{P_k}} = \frac{I_{source}}{4\pi r_{P_k}^2} \prod_1^P (1 - \alpha_{S_{P_k}}) \quad (6.5)$$

où α_k représente le coefficient d'absorption de la k ième surface ayant donné naissance à S_{P_k} .

De façon analogue au cas du son direct il est possible d'introduire un facteur supplémentaire modélisant l'absorption de l'air. Si de plus l'on considère la source directive, il faudra reporter la directivité de la source par symétries successives par rapport aux parois.

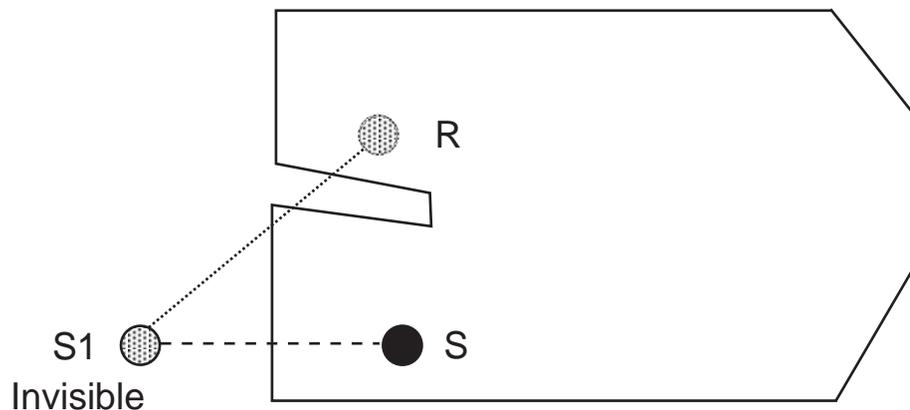


FIG. 6.8 – Salle non convexe : cas d'obstruction

6.3.3 Comparaison des méthodes

Le choix entre ces deux méthodes a déjà été évoqué lors des considérations sur une approche paramétrique de l'effet de salle. On va de plus les temps de calcul nécessaires à l'estimation de l'échogramme au point récepteur sur l'intervalle de temps considéré soit (20 – 80ms) (premières réflexions, cluster). Le cas considéré est celui d'un calcul direct, selon un temps linéaire (il arrive en effet que dans le cadre de l'acoustique prévisionnelle, on suive les rayons un par un jusqu'à leur abandon). On considère une salle de S surface élémentaires.

Coût de la méthode des rayons

Dans le cas d'une méthode des rayons, l'algorithme se déroule de la manière suivante : N rayons sont tirés à partir de la source, vont frapper les parois, et repartent vers l'intérieur de la salle. Chaque itération consiste à effectuer un nouveau tir de tous les rayons vers les surfaces. Un rayon est abandonné, dès le moment où la distance totale qu'il a parcouru est supérieure à la distance correspondant à 80 ms (soit 26,4 m).

L'initialisation consiste à effectuer une discrétisation de la source, et évaluer l'énergie portée par chaque rayon.

A chaque itération, il faut alors, pour chaque rayon, effectuer les étapes suivantes :

- Détermination de la surface touchée par le rayon, et du point d'impact : ce calcul doit être effectué en parcourant les S surfaces, moins celle d'origine du rayon. Si une surface a été trouvée dans la trajectoire du rayon, il n'est pas possible de stopper l'algorithme, dans la mesure il peut toujours exister une autre surface sur cette trajectoire, et masquant la première surface trouvée.
- Test de proximité du rayon vis à vis du récepteur, et de calcul éventuel de sa

contribution. Ce test n'est pas effectué à la première itération dans la mesure où le son direct est calculé de manière indépendante.

- Test de poursuite de l'algorithme pour le rayon considéré
- Calcul de la direction de réflexion du rayon

On introduit les grandeurs suivantes :

- T_{incid} : temps de détermination du point d'incidence par rapport à une surface
- T_{reflec} : temps de calcul de la direction du rayon réfléchi
- T_{proxi} : temps de calcul pour le test de proximité
- T_{pours} : temps de calcul du test de poursuite pour un rayon
- N_o : nombre total de rayons abandonnés après l'ordre O .

Le temps de calcul global pour un ordre d'itération est de :

$$(N - N_{o-1}) * (S - 1) * T_{incid} + (N - N_{o-1}) * T_{proxi} + (N - N_{o-1}) * T_{pours} + (N - N_o) * T_{reflec}$$

Ce temps reste constant ou diminue avec l'ordre d'itération. Il augmente linéairement avec le nombre de surfaces.

Coût de la méthode des sources-images

Dans le cas d'une méthode des sources-images, l'itération s'effectue sur l'ordre des sources-images.

Pour un ordre donné les étapes de calcul sont les suivantes :

- Placement des sources-images : ce placement s'effectue pour *a priori* sur les S parois, et pour toutes les sources-images valides à l'ordre précédent. Pour un ordre o donné on a *a priori* $(S' - 1)^o$ sources à placer.
- Test de validité de distance de toutes les sources-images par rapport au récepteur
- Tests de validité et de visibilité des sources-images par rapport au récepteur
- Calcul des contributions de toutes les sources-images valides

On introduit les grandeurs suivantes :

- T_{plac} : temps de placement d'une sources-images
- T_{pours2} : temps de calcul pour le test de poursuite de l'algorithme pour une source

- T_{valid} : temps de calcul pour le test de validité d'un sources-images
- T_{visib} : temps de calcul du test de visibilité d'une sources-images
- Np_o : nombre de sources invalidées par le test de poursuite à l'ordre O
- Nv_o : nombre de sources invalidées par le test de validité à l'ordre O
- N_o : nombres de sources valides après l'ordre O^5

Le temps de calcul pour un ordre donné O est donc de

$$N_{o-1}*(S-1)*(T_{plac} + N_{o-1}*(S-1)*T_{pours} + (N_{o-1} - Np_o)*T_{valid} + (N_{o-1} - Np_o - Nv_o)*T_{visib})$$

Ce temps de calcul évolue selon deux modes contradictoires dans le temps : le nombre de sources-images par ordre croit de manière exponentielle, tandis que les sources sont invalidées au fur et à mesure de l'évolution temporelle. Il augmente avec le nombre de surface.

Comparaison des coûts

L'estimation du temps global des deux algorithmes reste impossible sans ordre de grandeur pour toutes ces valeurs. On peut néanmoins effectuer une estimation "intuitive" des temps de calcul, et ce sur un cas d'école. Pour cela on considère une salle de concert de volume V égal à $12000m^3$, et de surface $3000m^2$. Le libre parcours moyen est donc de $16m$; on peut supposer que la source se situe alors à une distance moyenne de $16/2 = 8m$ des parois. La donnée de ces deux distances nous permet de calculer l'ordre d'itération dans une méthode des rayons : la distance moyenne parcourue par un rayon en fonction de l'ordre O est de $8 + 16 * O$. Il faut donc aller jusqu'à un ordre 2 pour arrêter l'algorithme (on ne considère un parcours maximal de $26,4m$ correspondant à $80ms$). On supposera de plus une discrétisation de $S = 100$ surfaces.

Lors de l'utilisation d'une méthode des rayons, le nombre de rayons tirés est en général de l'ordre de 10000 dans le cas d'une variante par cônes.⁶

Etape	durée	d
initialisation	$10^6 * T_{incid} + 10^4 * T_{pours} + 10^4 * T_{reflec}$	8 m
ordre 1	$10^6 * T_{incid} + 10^4 * T_{proxi} + 10^4 * T_{pours} + 10^4 * T_{reflec}$	24 m
ordre 2	$10^6 * T_{incid} + 10^4 * T_{proxi} + 10^4 * T_{pours}$	40 m
total	$3 * 10^6 * T_{incid} + 2 * 10^4 * T_{proxi} + 3 * 10^4 * T_{pours} + 2 * 10^4 * T_{reflec}$	//////

5. Cette notion de source valide sera précisée au paragraphe suivant

6. Ce nombre peut monter jusqu'à 100000 lors d'un tir de rayons simples.

Pour une méthode des sources-images, on peut estimer l'ordre de calcul par analogie avec une méthode des rayons : une sources-images d'ordre P correspond à un rayon ayant effectué P réflexions. On supposera de manière semblable que la source et le récepteur se situent à une distance moyenne de $8m$ des parois de la salle. Les sources-images d'ordre P se situent donc à une distance moyenne de $8 + 8 + (P - 1) * 16m$ du récepteur. Il faut donc aller jusqu'à un ordre 2 pour que l'algorithme soit stoppé. Il faut de plus estimer le nombre de sources éliminées par les tests de validité et de visibilité ; dans le cas le plus défavorable, aucune source n'est éliminée par ces tests⁷.

Etape	durée	d
ordre 1	$100 * T_{plac} + 100 * T_{pours2} + 100 * T_{valid} + 100 * T_{visib}$	16 m
ordre 2	$100 * 99 * T_{plac} + 100 * 99 * T_{pours2}$	32 m
total	$10000 * T_{plac} + 10000 * T_{pours2} + 100 * T_{valid} + 100 * T_{visib}$	//////

En considérant la complexité calculatoire des opérations à effectuer, on peut estimer que $T_{reflec} = T_{plac}, T_{pours} = T_{pours2}, T_{incid} = T_{visib}, T_{proxi} = T_{valid}$, ce qui laisse alors supposer que la méthode des sources-images est plus rapide que celle des rayons⁸. Il faut toutefois noter que cet avantage tient principalement au contexte de l'estimation temporelle supérieure d'estimation relativement faible, et ne correspond pas à un contexte classique d'acoustique prévisionnelle. Dans ce dernier cas l'évolution selon l'ordre est schématisé selon la figure suivante.

6.4 Calcul des contributions diffuses

L'hypothèse d'uniformité du champ réverbéré, selon la théorie de Sabine, n'est en pratique vérifiée qu'approximativement. Par exemple, au voisinage d'une paroi diffusante, il faut s'attendre à une augmentation du niveau du champ réverbéré. De même, le cas de salles composées de volumes couplés ne satisfait pas au modèle de Sabine.

Une méthode plus précise considère l'évolution du champ réverbéré comme un processus d'échange d'énergie entre les parois de la salle. Ce processus est alimenté par le signal issu de la source, qui permet d'attribuer des contributions initiales aux parois de la salle. Tout point dans la salle peut alors être considéré comme un observateur du processus. Ce modèle a notamment été décrit et appliqué dans la thèse de Christian Malcurt [Mal86]. Il se rapproche des méthodes utilisées en graphisme 3D et qualifiées de radiosité.

7. Il sera vu au paragraphe suivant que ce cas est plus que défavorable, voire irréaliste

8. Sur un cas réel, il serait plus que probable qu'il existe des sources d'ordre 2 valides, mais il est également possible que certaines sources d'ordre 1 soit invalides

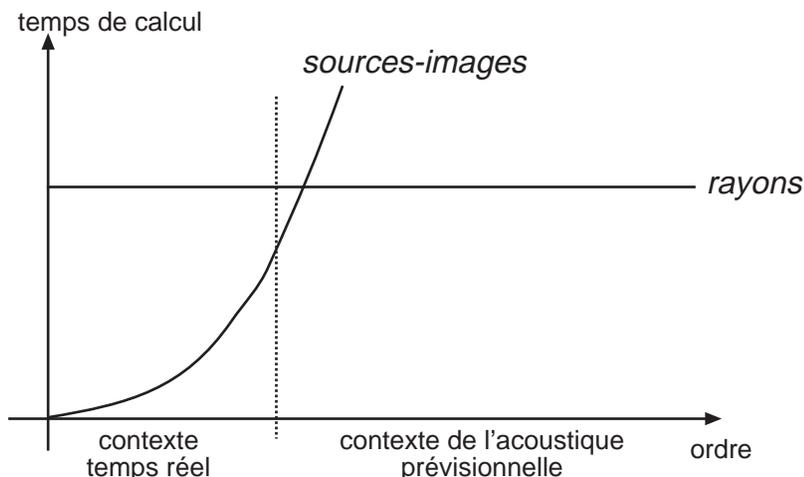


FIG. 6.9 – Dépendance du temps de calcul des méthodes géométrique selon l'ordre. Du fait que l'on cherche à estimer le résultat de l'application d'une telle méthode sur un intervalle de temps situé en début de réponse et faible, il est préférable de choisir une méthode des sources-images.

6.4.1 Fondements théoriques

Il est possible de décomposer l'étude théorique en trois parties. Dans un premier temps sera formulée l'équation de Kutruff, permettant de décrire effectivement l'énergie présente sur une surface élémentaire en fonction des énergies présentes sur les autres surfaces. Puis l'évolution temporelle de ce processus sera décrite, ainsi que son adaptation à une utilisation informatique. Enfin, les différentes données accessibles par cette méthode seront examinées.

Modèle de Kutruff

Soit un élément de surface dS , en un point x , la puissance reçue par dS au temps t peut s'écrire comme contribution des puissances présentes sur les autres surfaces de la salle, au temps de propagation près :

$$P(x, t, dS) = \int_{\Omega} \vec{I}(\vartheta) \cdot d\vec{S} \cdot d\Omega \quad (6.6)$$

avec

$$\begin{cases} \vec{I}(\vartheta) & \text{vecteur d'intensité reçu dans la direction } \vartheta \\ \vartheta & \text{angle du vecteur } \vec{I} \text{ et de la normale à la surface } dS \text{ au point } x \\ d\Omega & \text{angle solide élémentaire entourant } \vec{I}(\vartheta) \end{cases}$$

L'intensité totale reçue par dS est, par définition, égale à :

$$I(x, t) = \frac{P(x, t, dS)}{dS} \quad (6.7)$$

Suivant l'hypothèse de diffusion, si un point x' d'un élément de surface dS' reçoit une intensité $I(x', t)$, dS' émet dans toutes les directions ϑ l'intensité :

$$K(1 - \alpha(x'))I(x', t) \quad (6.8)$$

avec

$$\begin{cases} K & \text{constante} \\ \alpha(x) & \text{coefficient d'absorption de la surface } dS' \end{cases}$$

En combinant les équations ci-dessus, on peut exprimer l'intensité totale reçue par dS en fonction des intensités émises par les autres surfaces

$$I(x, t) = \int_{\Omega} K(1 - \alpha(x'))I(x', t - \frac{r}{c}) \cos(\vartheta) d\Omega \quad (6.9)$$

avec

$$\begin{cases} r & \text{distance entre les points } x \text{ et } x' \\ d\Omega & \text{angle solide de la surface } dS' \text{ vue du point } x, \text{ soit } d\Omega = \frac{dS' \cos(\vartheta)}{r^2} \end{cases}$$

La constante K est déterminée par la loi de conservation de l'énergie : si un élément de surface dS reçoit une puissance acoustique égale à dS , c'est-à-dire une intensité acoustique $I(x, t)$ égale à 1, la puissance émise par dS mesurée sur la demi-sphère S' de rayon 1 (en considérant α nul) doit être égale à dS .

En dimension 3, on obtient $K = \frac{1}{\pi}$.

L'intensité totale reçue au point x s'écrit donc

$$I(x, t) = \int_S \frac{\cos(\vartheta)}{\pi r^2} (1 - \alpha(x')) I(x', t - \frac{r}{c}) \cos(\vartheta) dS \quad (6.10)$$

Processus de Markov

Un processus est qualifié de Markovien, si son état à un instant donné ne dépend que de son état précédent. Dans le cas présent, on considère l'intensité acoustique : l'intensité acoustique émise au temps t , par une portion de surface dS intérieure à la salle, ne dépend que de l'intensité émise par tous les éléments de surface dS' , aux instant t diminués du temps de parcours.

Il est alors possible de construire une matrice de passage qui va permettre d'établir un bilan temporel et spatial de l'énergie du champ sonore présente sur la surface intérieure de la salle.

L'intensité totale reçue en un point x de la surface, au temps t peut s'écrire :

$$I(x, t) = \int_S \frac{\Omega}{\pi} (1 - \alpha(x')) I(x', t - \frac{d(x, x')}{c}) \cos(\vartheta) dS + Id(x, t) \quad (6.11)$$

avec

$$\left\{ \begin{array}{ll} I(x, t) & \text{"intensité totale" reçue au point } x \text{ au temps } t \\ \Omega & \text{angle solide de la surface } dS' \text{ du point } x' \text{ vue par le point } x \\ Id(x, t) & \text{intensité de l'onde directe reçue au point } x \text{ au temps } t \\ d(x, x') & \text{distance entre les points } x \text{ et } x' \\ \alpha(x') & \text{coefficient d'absorption de la surface au point } x' \\ \vartheta & \text{angle du rayon } (x, x') \text{ et la normale à la surface au point } x \end{array} \right.$$

Cette équation peut être discrétisée en découpant les parois de la salle en panneaux élémentaires. On obtient alors :

$$I_i(t) = \sum_j \int_{S_j} (1 - \alpha_j) I_j(t - \frac{d_{ij}}{c}) \cos(\vartheta_{ij}) \frac{d\Omega_j}{\pi} + Id_i(t) \quad (6.12)$$

avec

$$\left\{ \begin{array}{ll} I_i(t) & \text{intensité reçue par la surface } i \text{ au temps } t \\ d\Omega_j & \text{angle solide de la surface } S_j \text{ vue par la surface } dS_i \\ \vartheta_{ij} & \text{angle du rayon } S_j - S_i \text{ et de la normale à la surface } S_i \end{array} \right.$$

L'adoption d'un certain nombre d'hypothèses permet d'effectuer des simplifications :

- α_j est indépendant de la direction et constant sur toute la surface S_j
- I_j , l'intensité est constante sur toute la surface S_j
- Les surfaces sont suffisamment petites pour négliger les variations des termes géométriques de l'équation
- Les surfaces élémentaires sont alors entièrement caractérisées par leur centre.

On obtient alors :

$$I_i(t) = \sum_j \frac{\cos(\vartheta_{ji}) S_j}{d_{ij}^2} \cos(\vartheta_{ij}) \frac{(1 - \alpha_j)}{\pi} I_j(t - \frac{d_{ij}}{c}) + Id_i(t) \quad (6.13)$$

et

$$Id_i(t) = w(t - \frac{d_{si}}{c}) \frac{\Omega_{si}}{4\pi S_i} \quad (6.14)$$

avec

$$\left\{ \begin{array}{l} s_j \quad \text{surface du panneau } j \\ d_{ij} \quad \text{distance centre de la surface } i \text{ - centre de la surface } j \\ \vartheta_{ji} \quad \text{angle du vecteur centre de la surface } j \text{ - centre de la surface } i, \text{ et de la normale à la surface } j \\ d_{si} \quad \text{distance source- surface } i \\ \Omega_{si} \quad \text{angle solide de la surface } i \text{ vue de la source} \\ S_i \quad \text{aire de la surface } i \end{array} \right.$$

En un point quelconque de la salle, un bilan énergétique du champ sonore peut être établi. L'énergie diffusée par la surface i sur le récepteur est donnée par :

$$\Omega_{ri} I_i \left(t - \frac{d_{ri}}{c} \right) \quad (6.15)$$

avec

$$\left\{ \begin{array}{l} \Omega_{ri} \quad \text{angle solide de la surface } i \text{ vue par le récepteur} \\ d_{ri} \quad \text{distance du centre de la surface } i \text{ au récepteur} \end{array} \right.$$

L'énergie totale présente sur le récepteur se calcule en effectuant la somme des contributions de toutes les surfaces, soit :

$$I_{rec}(t) = \sum_i \Omega_{ri} \frac{1 - \alpha_i}{\pi} I_i \left(t - \frac{d_{ri}}{c} \right) + Id_r(t) \quad (6.16)$$

avec :

$$Id_r(t) \quad \text{intensité de l'onde directe source récepteur au temps } t, \text{ soit } Id_r(t) = \frac{w(t - \frac{d_r}{c})}{4\pi d_r^2}$$

et

$$\Omega_{ri} = \frac{\cos(\vartheta_{ri}) s_i}{d_{ri}^2} \quad (6.17)$$

ϑ_{ri} angle du vecteur récepteur - centre de la surface i avec la normale à la surface i

Formulation matricielle

Considérons la matrice A composée des coefficients

$$\left\{ \begin{array}{l} a_{ij} = \frac{\Omega_{ij}}{\pi} \cos(\vartheta_{ij}) \\ a_{ii} = 0 \end{array} \right.$$

Cette matrice peut s'interpréter comme décrivant le couplage entre chaque surface élémentaire de la salle. On notera que, la loi de conservation de l'énergie impose que la somme des coefficients d'une même colonne de la matrice A soit égale à 1, ce qui implique une normalisation avant tout traitement ultérieur.

L'équation 6.13 peut alors s'écrire

$$I_i(t) = \sum_j a_{ij}(1 - \alpha_j)I_j(t - \frac{d_{ij}}{c}) + Id_i(t) \quad (6.18)$$

6.4.2 Mode de propagation

Propagation directe

Le mode de propagation direct consiste à appliquer directement l'équation 6.18 pour calculer, pas à pas, l'évolution des intensités sur les parois. Le processus peut être poursuivi jusqu'à la limite d'audibilité du signal, ou on peut utiliser le formalisme de Kuttruff pour déterminer la distribution d'intensité et le temps de réverbération en régime tardif exponentiel (*paragraphe suivant*). Les contributions de ces parois sont ensuite collectées par le(s) récepteur(s) dans la salle.

Propagation indirecte

La résolution de l'équation matricielle 6.18 s'avère intéressante dans deux cas.

Régime permanent

Les termes dépendant du temps sont constants ; on peut donc écrire:

$$I_i = \sum_j a_{ij}(1 - \alpha_j)I_j + Id_i(t) \quad (6.19)$$

ou sous forme matricielle

$$(Id_N - A\alpha)I = I_d \quad (6.20)$$

avec, Id_N matrice identité de dimension N , α .

Dans ce cas la source alimente la salle avec un signal permanent. A celui-ci va correspondre le vecteur d'intensité excitatoire I_d . La répartition de l'énergie sur les parois va alors obéir à l'équation matricielle 6.20, et se calcule par inversion de la matrice $(Id_N - A\alpha)$. La salle effectue une amplification de l'énergie apportée par la source à chaque paroi. On peut parler de mode forcé.

Régime exponentiel décroissant

On suppose qu'après extinction des sources les intensités présentes sur les parois vont adopter un profil décroissant exponentiellement. Conformément à cette hypothèse de champ tardif, il est alors possible de négliger le terme Id_i correspondant au champ direct (6.18).

On obtient alors une équation portant sur la répartition de l'énergie sur les parois de la salle, et sur le temps de réverbération.

$$I_i = \sum_j a_{ij}(1 - \alpha_j)I_j \exp\left(\frac{d_{ij}}{c}\right) \quad (6.21)$$

On constate alors qu'il s'agit de résoudre un système de N équations à $N + 1$ inconnues. Deux choix sont possibles : supposer le temps de réverbération connu et calculer l'intensité présente sur les parois de la salle, ou supposer une distribution sur les parois et calculer le temps de réverbération associé (relié au coefficient p par $p = 13.81/Tr$). La résolution est menée de manière itérative en prenant comme condition initiale, soit une hypothèse arbitraire de distribution ou d'une excitation uniforme sur les parois, soit l'état de la distribution des intensités sur les surfaces à la fin du processus des sources-images ou du suivi propagatif de la radiosit .

R verb ration directionnelle

La th orie de Sabine suppose le champ diffus anisotrope. Perceptivement, la r verb ration n'est pas localisable. Dans certains cas cette hypoth ese est fautive, et notamment dans celui de volumes coupl s, ou lors de la pr sence d'ouvertures.

Il peut alors  tre int ressant de calculer une direction majoritaire du champ tardif. Celle-ci sera consid r e comme la barycentre des directions r cepteur-surfaces pond r es par les  nergies provenant de chaque surface, soit

$$Dir_{R3} = \frac{1}{\sum_i E_i} \sum_i E_i Dir_i \quad (6.22)$$

Si le vecteur r sultant n'est pas de norme tr s faible, on peut alors d finir une direction majoritaire de r verb ration. Le fait de trouver une norme non nulle peut cependant correspondre   des cas ou le champ r verb r  poss de des propri t s de sym trie spatiale. Il peut  tre utile dans ce cas d' tablir l'histogramme de la r partition spatiale, afin de v rifier que sa distribution est bien uniforme.

6.5 Passage entre les deux processus

Suivant la finesse de la simulation diff rents cas sont   consid rer.

6.5.1 Mod le son direct-reverb ration

Dans ce cas, l' nergie de l'effet de salle est consid r e diffuse d s le d part, la source alimentant directement les surfaces de la salle. On est alors dans le cas de propagation

indirecte. Trois étapes sont à prendre en compte pour calculer le niveau du champ réverbéré en un point de la salle. La matrice inverse est précalculée.

- imputation de la puissance de la source vers les surfaces de la salle
- application de l'amplification de la salle (radiosité indirecte).
- imputation de la puissance de chaque surface sur le récepteur.

Le temps de réverbération peut être soit précalculé, soit calculé par résolution du système 6.19 en partant de l'excitation des sources.

Ces étapes font appel aux distances transducteur-surfaces, angles solides de la portion d'espace définie par chaque surface vue des transducteurs. Un problème apparaît en pratique dû à la taille des surfaces considérées.

Un mouvement de la source nécessitera le calcul des trois étapes, tandis qu'un mouvement du récepteur ne requiert que la dernière étape.

6.5.2 Présence de premières réflexions et du cluster

Dans le cas où les premières réflexions et le cluster sont considérés, il n'est pas possible d'alimenter le processus directement avec l'énergie provenant de la source. Le processus markovien indirecte est alimenté par les contributions apportées au parois lors de la propagation des ondes en début de réponse impulsionnelle.

Deux points principaux sont à prendre en compte :

- Le mode de transition entre les deux processus
- Les types de phénomènes modélisés lors de la propagation

Transition entre les deux processus

Deux types de critères peuvent être pris en compte afin d'effectuer ce passage :

- Critères sur l'ordre : un algorithme de propagation effectue généralement une itération en fonction d'un ordre, comme par exemple un processus de sources-images. Au delà de l'ordre défini, on considère alors que toutes l'énergie est propagé selon un algorithme de radiosité indirecte.
- Critères sur le temps : dans ce cas, et si l'on reconsidère l'utilisation d'un algorithme de sources-images, l'intégration d'une source-image au processus de propagation va dépendre de sa distance par rapport au récepteur. Ce mode de transfert est plus en accord avec l'utilisation d'une représentation morphologique de la réponse impulsionnelle.

Phénomènes modélisés lors de la propagation

Il reste cependant à déterminer l'intensité présente sur chaque surface en début d'utilisation du processus de radiosité indirecte. On effectuera dans un premier temps une hypothèse sur la nature des phénomènes modélisés lors du processus de propagation, soit uniquement les phénomènes spéculaires.

Dans ce cas, lors de l'abandon du processus spéculaire, à chaque surface est associé un ensemble de sources images, qui viennent de propager leur énergie de manière spéculaire. Afin d'imputer plusieurs fois cette énergie au récepteur il convient de recalculer le réseau de sources images correspondant à l'itération suivant pour ces sources. Ce sont les éléments de ce nouveau réseau qui éclaireront chaque paroi de manière diffuse. On notera qu'à une paroi donnée peut correspondre plusieurs sources images.

En réalité, les phénomènes de propagation de l'énergie diffuse commencent dès le début de la réponse impulsionnelle. Dans ce cas il faut à chaque réflexion séparer l'énergie réfléchie de manière spéculaire et celle réfléchie de manière diffuse.

Une première correction à effectuer concerne l'énergie spéculaire véhiculée par chaque source image. Du fait que l'on suppose les parois également diffusantes, il faut appliquer un facteur correctif à l'onde véhiculée soit, pour une sources-images

$$I_{S_{Pk}} = \frac{I_{source}}{4\pi r_{Pk}^2} \prod_1^P (1 - \alpha_{S_{Pk}})(1 - \delta_{S_{Pk}}) \quad (6.23)$$

δ étant le coefficient de diffusion de la paroi considérée.

A chaque réflexion une partie de l'énergie est également réfléchie de manière diffuse. Avant que ne commence le processus de radiosité indirecte il convient alors de traiter celle-ci. A un ordre donnée correspond une valeur d'intensité diffuse associée à chaque surface et issu de la propagation spéculaire. La contribution d'ordre 1 (de même que les suivantes) peut cependant être propagé, de manière parallèle, selon un algorithme de radiosité directe. Il convient alors de considérer également un temps limite, égal au temps de passage des processus au delà duquel l'énergie propagée de manière diffuse est stockée pour chaque surface. C'est la somme de ces contributions qui servira à alimenter le processus diffus indirecte.

Une version de cette algorithme a été implanté dans le logiciel RAYFRAC/DIFFUS réalisé à l'IRCAM. On notera cependant que celui-ci est basé sur une méthode des rayons, et que le traitement de la première contribution est différent.

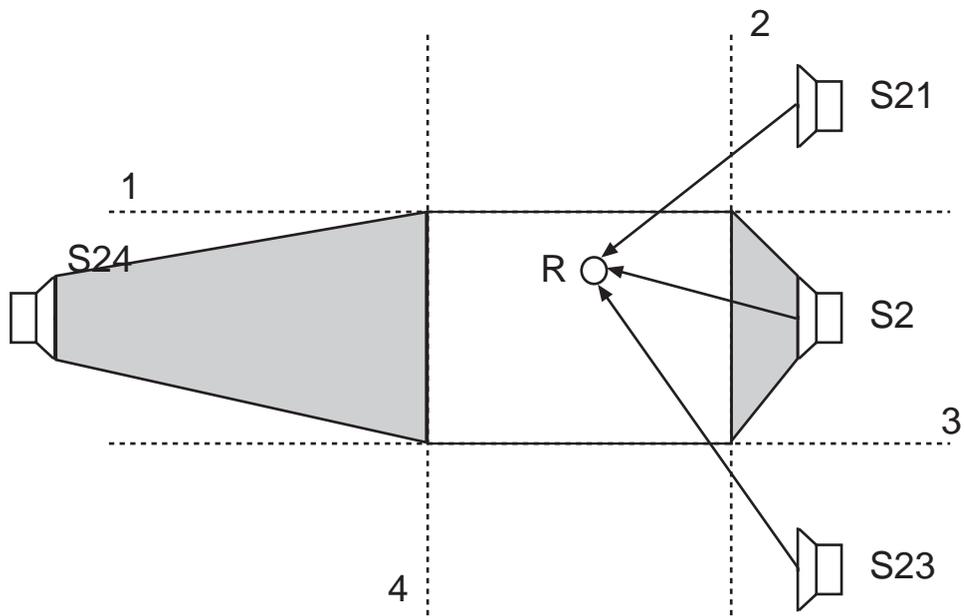


FIG. 6.10 – *Passage progressif des sources-images vers un modèle de radiosité*

Chapitre 7

Architecture d'un système de contrôle de l'effet de salle

Le schéma proposé au chapitre 1 figure 2.10, concernant la représentation et la manipulation de l'effet de salle, a donné lieu à la création d'une librairie en C permettant de manipuler un effet de salle selon différentes modalités. Cette librairie, appelée VASR (Virtual Audio Scene Rendering) est décrite dans ce chapitre.

Deux parties sont proposées. Tout d'abord une architecture générale est décrite, qui permet d'organiser les liens descendants de ce schéma (haut-niveau \rightarrow bas-niveau), et ce afin de proposer une utilisation simple par un programmeur. Puis, une attention particulière a été portée au développement concernant le traitement géométrique de l'effet de salle.

7.1 Présentation générale

7.1.1 Caractéristiques

La librairie possède les propriétés suivantes :

- écrite en C, ANSI et donc portable et recompileable sur différentes plate formes et systèmes opérationnels.
- extensible : la librairie présente une structure d'organisation des fichiers ; l'implantation particulière d'une fonction peut cependant varier. Ce point est à mettre en relation avec l'alinéum précédent : en effet, la contrainte de portabilité peut entraîner des non optimisations concernant certains traitements, qui devraient alors être réimplantés¹.
- organisation orientée objet : bien que le langage choisi ne soit pas un langage objet, VASR est organisée autour d'objets élémentaire relatifs au traitement de l'effet de salle.

Dans le cadre de cette thèse, et comme il a été vu au chapitre 5, on s'intéresse particulièrement au problème du pilotage de l'effet de salle par une interface géométrique. De plus, le moteur de rendu utilisé étant commandable par une représentation bas-niveau de type morphologique, ce sont principalement les fonctions de traitement relatives à ce trajet là qui ont été implantées. Néanmoins, VASR peut servir de cadre pour l'implantation d'autres trajets.

Certaines limitations ont cependant été effectuées : tout d'abord, les géométries de salles utilisées sont considérées fixes. De plus, seules des situations à une source et un récepteur

1. On pourra citer, entre autres, les opérations de gestion de la mémoire, reliés au système opérationnel, et surtout les opérations de type mathématique. Celles-ci auront tout intérêt à être réécrites pour certains processeurs (AMD, jeu d'extension MMX, AltiVec).

ont été considérées. Le traitement multisources nécessiterait la mise en commun de certaines ressources, suivant la configuration choisie : par exemple dans le cas où deux canaux acoustiques auraient en commun une même salle, il serait nécessaire d'établir un lien afin de n'effectuer le calcul du temps de réverbération qu'une seule fois.

7.1.2 Organisation

Différents objets sont présents :

- Les objets manipulables par l'utilisateur : ce sont principalement les transducteurs².
- Les objets relatifs au traitement acoustique géométrique : salle, paroi. . .
- Les objets de représentation des données : canal acoustique selon les deux types de représentations bas-niveau, structures pour les paramètres décrivant une interface de manipulation

A chaque objet est associé un fichier source traitant les opérations élémentaires :

- Création et destruction d'objets : la mémoire est allouée suivant le contexte de traitement
- Accession et mise à jour des champs de données
- Calcul de données internes : par exemple dans le cas d'un rendu géométrique de la réverbération, un jeu de fonctions permettra le calcul des matrices de Markov directe et indirecte.
- Interrogation indirecte sur les propriétés d'un objet : par exemple si un transducteur a sa directivité décrite par un modèle tel ceux présentés au chapitre 2, il est ainsi possible de calculer son atténuation pour une direction donnée.

Les interactions entre les objets (lors par exemple d'une manipulation par un utilisateur) sont regroupées dans un fichier particulier. Les fonctions réalisant ces traitements sont précédées du préfixe `IF_`, comme Interaction Function.

Enfin, deux fichiers regroupent un ensemble de fonctions utilisées par les routines de VASR : ce sont les utilitaires mathématiques (particulièrement géométrique) (fichier `VASR_Matut.c`), et les fonctions de tri (fichier `VASR_Sort.c`).

2. Le cas d'une géométrie de salle variable, n'est pas considéré dans cette thèse.

7.1.3 Prise en compte du contexte

Selon l'interface de manipulation choisie, le traitement effectué variera. Afin d'identifier le traitement à effectuer, chaque objet manipulable est doté d'un champ `contextType`, précisant le type de manipulation. On y reconnaît les 3 modalités principales de description de la scène sonore : les approches géométrique, physique statistique et perceptive.

```
typedef enum {
    kVASRCTXT_NONE,    //No context defined
    kVASRCTXT_GEO,    //Geometrical context defined
    kVASRCTXT_STAT,   //Statistical context defined
    kVASRCTXT_PERC,   //Perceptual context defined
    kVASRCTXT_OTHER   //Other context defined (future use)
} VASR_SimContext;
```

Un autre champ, `contextLevel`, permet d'identifier le niveau de précision dans le rendu de l'effet de salle, en faisant référence aux sections temporelles de la réponse impulsionnelle dans sa représentation morphologique (*cf*: chapitre 1).

```
typedef enum {
    kVASRCTXTLEVEL_NONE = 0x00000000, //Donorendering
    kVASRCTXTLEVEL_OD   = 0x00000001, //Dorenderingofdirectwave
    kVASRCTXTLEVEL_R1   = 0x00000002, //Dorenderingofearlyreflections
    kVASRCTXTLEVEL_R2   = 0x00000004, //Dorenderingofcluster
    kVASRCTXTLEVEL_R3   = 0x00000010 //Dorenderingofreverb
} VASR_SimLevel;
```

Suivant, par exemple la puissance de la machine, ou le type d'application ces fonctions seront combinées; ainsi, pour un rendu d'effet de salle comprenant un son direct et une réverbération, on passera en paramètre `kVASRCTXTLEVEL_OD+kVASRCTXTLEVEL_R3`.

Ces dénominations permettent au développeur d'appeler toujours les mêmes fonctions, qui prennent alors en charge le choix des méthodes utilisées. Afin d'optimiser ces traitements, les objets manipulables par l'utilisateur, c'est-à-dire les transducteurs, possèdent un champ "cache" permettant de stocker des données réutilisables par plusieurs autres fonctions. Ceci est particulièrement utile dans le cas de manipulation géométrique : il est alors possible de stocker des grandeurs telles distances, angles solides. . . .

La coexistence de plusieurs contextes de manipulation peut être réalisée à hauteur de deux interfaces. Les champs entre les deux interfaces sont *a priori* indépendants. Cette solution présente plusieurs avantages :

- En cas d'indépendance, il est ainsi possible de faire cohabiter plusieurs interfaces. Lors d'une phase d'utilisation, cela permet d'utiliser des réglages temporaires, pour créer momentanément, par exemple, un effet spécial.

- En cas de synchronisation, cela permet de voir les liens entre plusieurs modalités d'interface, ce qui peut être utile lors d'une phase d'élaboration d'espace virtuel. Cette synchronisation est basée sur les liens ascendants de la figure 2.10, et n'est pas possible dans tous les cas (*cf discussion § 2.3.3*)

7.2 La manipulation du canal acoustique

On s'intéresse ici à une utilisation par un programmeur sans rentrer dans le détail de l'implémentation pour un cadre particulier.

7.2.1 Objets en présence

Le canal acoustique au bas niveau peut être représenté de deux manières :

```

struct s_AcChannel_LL {
    VASR_LL_Type          AC_ActiveType;
    s_AcChannel_LLMorphDesc *ACMorphDesc;
    s_AcChannel_LL FIRDesc *ACFIRDesc;
}

```

Le champ `AC_ActiveType` peut prendre les valeurs `kLL_MorphDesc` (représentation morphologique) ou `kLL_FIRDesc` (réponse impulsionnelle). L'objet `s_AcChannel_LL` possède des méthodes permettant de passer d'une de ces représentations à l'autre.

La manipulation haut-niveau d'un effet de salle fait intervenir une structure permettant la connexion de plusieurs interfaces (deux au maximum) pouvant manipuler un canal acoustique. Celle-ci est décrite par :

```

struct s_AcChannel_HL {
    VASR_ManipUseFlag  whichuse;
    VASR_LL_Type       LLtype;
    s_AcChannel_Manip  *FirstManipCtxt;
    s_AcChannel_Manip  *FirstManipCtxt;
    s_Transducer       *theSourcePtr;
    s_Transducer       *theReceiverPtr;
    s_AcChannel_LL     *theAcChannelLL;
}

```

Le champ `whichuse` indique quel contexte de manipulation est actif. Chaque contexte de manipulation peut être du type statistique, géométrique ou perceptif. Le champ `theAcChannel_LL` indique la représentation bas niveau où sera écrit le résultat d'une

manipulation, après appel automatique des fonctions de conversion adéquates. Son type est précisé par le champ `LLtype`.

```

struct s_AcChannel_Manip {
    VASR_HLAcChanManipType  ManipType;
    VASR_LL_Type            LLtype;
    s_AcChannel_GeoManip    *theGeoManip;
    s_AcChannel_StatManip   *theStatManip;
    s_AcChannel_PercManip   *thePercManip;
    s_Transducer            *theSourcePtr;
    s_Transducer            *theReceiverPtr;
    s_AcChannel_LL          *theAcChannelLL;
}

```

Chaque contexte est ensuite décrit de manière plus précise et spécifique. Par exemple un contexte géométrique sera décrit par :

```

struct s_AcChannel_GeoManip {
    VASR_SimLevel          theSimLevel
    VASR_LL_Type          LLtype;
    s_RenderingOption     theRenderingOption;
    s_Room                 *theRoomPtr;
    s_RoomComputeParam    *theRoomComputeParam;
    s_Transducer          *theSourcePtr;
    s_Transducer          *theReceiverPtr;
    s_AcChannel_LL        *theAcChannelLL;
}

```

Le champ `theSimLevel` permet de préciser le niveau de finesse dans le rendu, et ce par rapport au traitement des différentes sections temporelles de l'effet de salle. `theRenderingOption` permet de préciser des options quant au type de phénomènes acoustiques pris en compte. Typiquement, il peut s'agir de l'absorption de l'air, des variations fréquentielles, de la prise en compte de la diffusion des parois ou des phénomènes comme la diffraction. Les contextes perceptifs et statistiques comprennent les données présentées au chapitre 2.

7.2.2 Utilisation pratique

Initialisation

En utilisant les fonctions d'initialisation et de création proposées par la librairie, il est possible d'instancier un canal acoustique, et ce, selon les contextes d'utilisation.

Par exemple, si l'on désire créer un programme permettant la manipulation d'un effet de salle, on instanciera un objet de type `s_AcChannel_HL`, qui comprend toutes les informations nécessaires. Si l'on désire simplement effectuer le lien avec une interface particulière, on pourra instancier directement un objet de type `s_AcChannel_XXXManip`.

Par défaut, ces objets n'allouent pas d'espace mémoire pour une source et un récepteur. Ceux-ci doivent être créés séparément, puis attachés à la structure déjà instanciée. Cette séparation permet une meilleure lisibilité dans la manière de préciser les propriétés des transducteurs. Selon le degré d'abstraction choisi pour la création d'un canal acoustique (`s_AcChannel_HL`, `s_AcChannel_XXXManip`), les transducteurs et sources des structures sous-jacentes sont directement référencés par ceux de la structure de haut-niveau.

Manipulation

Elle s'effectue par modification des champs des transducteurs. A chaque manipulation, les données cachées associées à ces transducteurs sont mises à jour automatiquement. La mise à jour du canal acoustique s'effectue ensuite par l'appel d'une fonction particulière reconnaissable au suffixe `_Update`. Cette fonction appelle les routines d'interaction appropriées. La récupération des données pour envoi vers un moteur de rendu s'effectue ensuite par interrogation du canal acoustique.

La figure 7.2.2 regroupe l'ensemble de l'architecture de la librairie VASR.

Libération des objets

Elle s'effectue simplement en appelant les fonctions de destruction des objets créés.

7.2.3 Exemple : traitement des données d'une interface géométrique

Le moteur de rendu utilisé est de type synthétiseur générique d'effet de salle. Dans un premier temps, on désire synthétiser seulement la réverbération et le son direct. Les options activées sont l'absorption de l'air et l'utilisation de 3 bandes fréquentielles. La représentation bas-niveau utilisée est de type représentation morphologique.

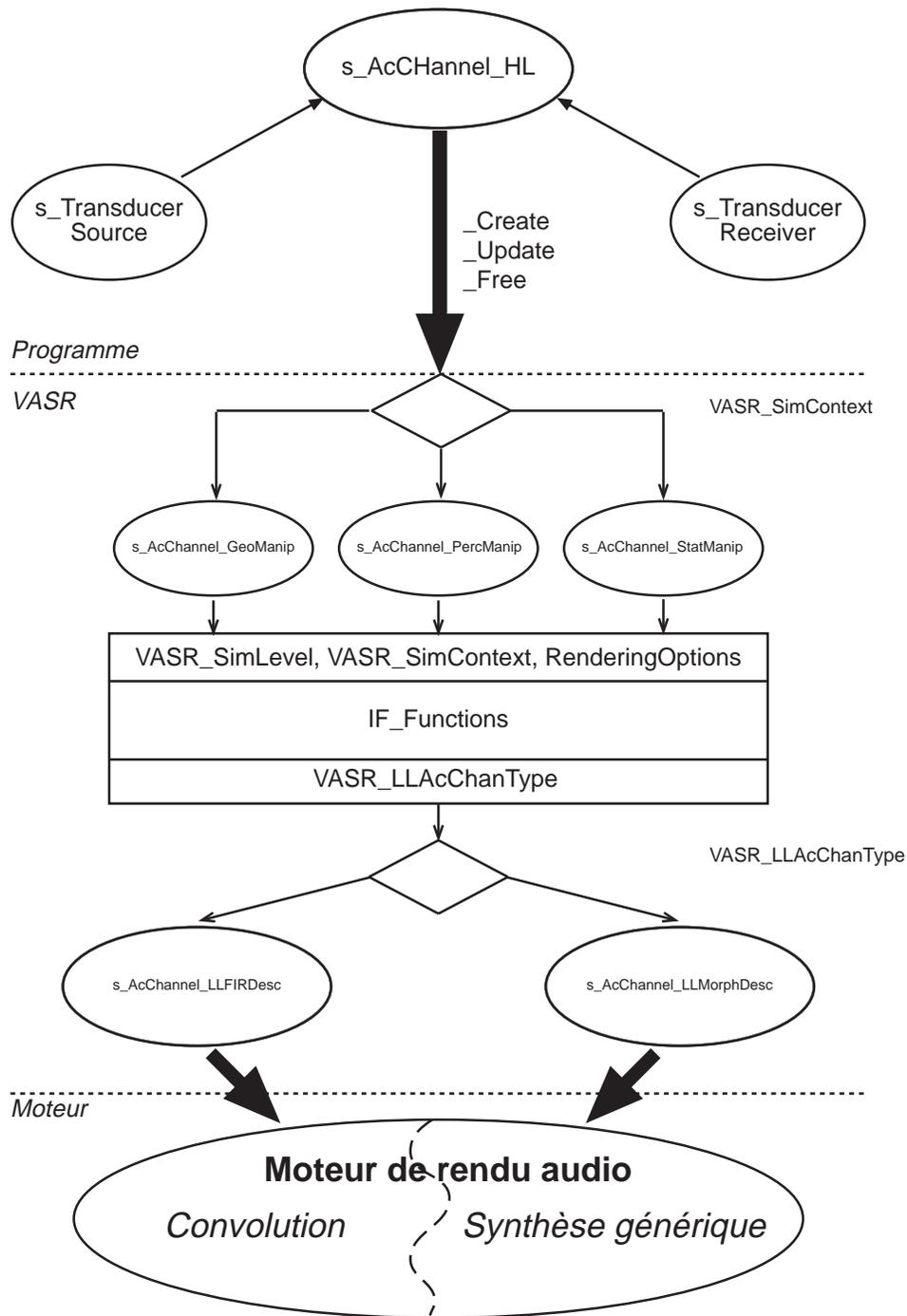


FIG. 7.1 – Structure et utilisation de la librairie VASR. L'utilisateur configure et manipule des objets de haut-niveau qui lui cache la complexité des algorithmes utilisés au bas-niveau

```
/* Création des objets, canal acoustique et transducteurs. Création du canal
spécifique à ce contexte, puis lecture de la salle dans un fichier, et mise en
place des options de rendu. La fonction de lecture de salle est désignée de
manière générique et est dépendante de la plateforme utilisée*/
```

```
AcChannel_GeoManip_InitializetoNull(&theAcChannel);
AcChannel_GeoManip_AllocFromLevel(&theAcChannel,
kVASRCTXLEVEL_OD+kVASRCTXLEVEL_R3,kMorphDescLLOption,);
AcChannel_GeoManip_RenderingOptionSet(kUseAirAbsorbionOn,
kUse3BandsOn, kLL_MorphDesc);
AcChannel_GeoManip_ReadRoom(&theAcChannel, *file);
```

```
Transducer_InitializetoNull(&theSource);
Transducer_SetType(&theSource, kTR_SOURCE);
Transducer_SetDirType(&theSource, kDIR_OMNI);
Transducer_SetPower(&theSource, 1.0f);
Transducer_InitializetoNull(&theReceiver);
Transducer_SetType(&theReceiver, kTR_RECEIVER);
Transducer_SetDirType(&theReceiver, kDIR_OMNI);
```

```
/* Rattachement des transducteurs au canal acoustique */
```

```
AcChannel_GeoManip_SetTransducerSource(&theAcChannel,&theSource);
AcChannel_GeoManip_SetTransducerReceiver(&theAcChannel,&theReceiver);
```

```
/* Manipulation de la source: on effectue un changement de position. Il
est possible d'effectuer plusieurs manipulations avant d'appeler la fonction
_Update. La fonction AcChannel_GeoLL permet de remplir une structure de
représentation morphologique bas-niveau passée en paramètre. Les champs de
cette fonction sont alors passés au moteur de rendu. */
```

```
Transducer_SetPos(&theSource, 2.0, 3.0, -1.0);
AcChannel_GeoManip_Update(&theAcChannel);
AcChannel_GetLL(&theAcChannel, &aMorphLLDataStructure);
```

```
/* Destruction des objets */
```

```
AcChannel_GeoManip_Free(&theAcChannel);
Transducer_Free(&theSource);
Transducer_Free(&theReceiver);
```

7.3 Manipulation géométrique de l'effet de salle

Dans le cadre de cette thèse, on s'est principalement intéressé à la manipulation temps réel de l'effet de salle. Le moteur de simulation utilisé étant le Spatialisateur, l'intérêt s'est principalement porté sur la conversion vers une représentation morphologique. On a ainsi approfondi les liens bas-niveaux permettant de calculer cette représentation à partir d'une RIF, ainsi que ceux avec une interface de manipulation de type perceptif. Afin de pouvoir considérer l'autre lien principal haut-niveau/bas-niveau, une interface de manipulation à contrôle géométrique a été implantée. Cette maquette a été écrite pour le Spatialisateur dans l'environnement Max/MSP, et ce sous la forme d'un ensemble d'objets externes. La communication entre objets, ainsi que les configurations d'objets se font donc sous la forme de messages MAX.

7.3.1 Description des modules de commande géométrique

Ils sont au nombre de quatre et utilisent des fonctions de la librairie VASR présentée ci-dessus. Chaque objet a été conçu pour une source et un récepteur

- *r3Dview* : il s'agit d'un visualisateur graphique en 3 dimensions, basé sur l'utilisation de la librairie QuickDraw 3D. Cette interface permet à l'utilisateur de charger une géométrie, et d'évoluer librement à l'intérieur en manipulant le clavier de l'ordinateur.
- *r3Ddirect* : il s'agit d'un objet effectuant le calcul de l'énergie OD ; il possède différentes options permettant d'activer une représentation plus ou moins fine du son direct (activation de l'absorption de l'air, rendu en plusieurs bandes de fréquence). Il faut noter que, bien que correspondant à la propagation en champ libre, cet objet accepte une géométrie en entrée. Il est ainsi possible de tester les cas où la source est masquée au récepteur
- *r3Dreflec* : ce module effectue le calcul des énergies R1 et R2, c'est à dire traite les premières réflexions et le cluster. Il est basé sur une méthode des sources-images. Afin d'éviter le calcul d'un nombre trop élevé de sources-images, seules les sources situées à une distance limite de 26,4 m du récepteur sont prises en compte (cette distance correspond à la limite temporelle de la fin du cluster soit 80 ms)³. A partir de cette distance un lien avec un modèle de propagation du champ diffus est effectué. Le calcul de R1 et R2 est ensuite fait par sommation des énergies des échos reçus⁴. A nouveau des options de rendu sont disponibles.

3. Etant données les dimensions d'une salle de concert, il faut à peu près prévoir le calcul sur 2 ordres au maximum

4. Cette méthode est également valide dans le cas de l'utilisation d'un moteur à convolution

- *r3Dreverb* : cet objet effectue le calcul de l'énergie R3 ainsi que du temps de réverbération. Il est entièrement basé sur l'utilisation d'une méthode de radiosité. Deux cas sont possibles : soit ce module est alimenté par la source, soit par une liste de résidus, correspondant à l'énergie présente sur les parois de la salle à la fin du processus de sources-images (issus du module *r3Dreflec*). A nouveau différentes options de rendu sont proposées.

7.3.2 Stockage des informations

Afin de regrouper les informations relatives à la géométrie dans un fichier unique, un format de fichier a été choisi. Il s'agit du format 3DMF (Apple Computer Inc.) et destiné au stockage de graphiques 3D. En effet, ce format est aisément extensible, et ce à travers la création d'attributs particuliers, permettant de stocker et manipuler des informations complémentaires. Le stockage des paramètres acoustiques a donné lieu à la définition d'un attribut décrit de la manière suivante :

```

struct s_AcAttribute {
    char   name      /* nom du materiau */
    float  absorb[8]; /* coefficients d'absorption par octave */
    float  diff[8];  /* coefficients de diffusion par octave */
}

```

Par ailleurs, ce format de fichier possède les propriétés suivantes : organisation hiérarchique des fichiers, stockage sous forme binaire et ASCII (permettant ainsi d'éditer manuellement le fichier), présence de nombreuses formes géométriques élémentaires dont le polyèdre.

7.3.3 Construction des modules de commande géométrique

On s'intéressera ici au fonctionnement interne des modules *r3Ddirect*, *r3Dreflec* et *r3Dreverb*.

r3Ddirect

Implantation

Elle est la transcription directe de la formule de propagation d'onde ??, ainsi que des améliorations proposées dans le chapitre ?. On suppose que les transducteurs ont été créés et initialisés, ainsi qu'une salle éventuelle (référéncé aux transducteurs). Chaque modification sur l'un des transducteurs doit être suivie des appels :

```
IF_TransducerSource_RelatPos_Receiver(&theSource,&theReceiver,    &az,  
&dist, &el);  
IF_TransducerSource_IrradiateReceiver(&theSource,                &theReceiver,  
ODResPtr,theRenderingOption);
```

Le premier appel permet de calculer les positions relatives des transducteurs, qui serviront pour la localisation, tandis que le deuxième renvoie la valeur de OD dans le pointeur `ODRes`. Dans le cas d'une utilisation à un haut niveau, l'appel à ces fonctions se fera de manière transparente au programmeur, par l'intermédiaire de la fonction `XXX_Update`.

Evaluation des performances

La simplicité de la formule théorique fait que le coût de calcul du niveau OD est minime. Néanmoins, dans le cas où une salle est chargée, un test doit être effectué afin d'examiner la possible présence d'une surface ou d'un obstacle entre la source et le récepteur. Dans le pire des cas (c'est-à-dire l'absence de surface) il faut effectuer ce test sur toutes les parois de la salle. Il peut alors être utile de simplifier la géométrie afin de regrouper les surfaces coplanaires.

r3Dreflec

Implantation

Ce module implémente une méthode des sources-images.

Evaluation des performances

r3Dreverb

Ce module est basée sur une méthode de radiosité (modélisation de l'évolution du champ tardif par l'utilisation de la matrice inverse d'crité au chapitre 6). On suppose que les transducteurs ont été créés et référencés à une salle.

Implantation

L'utilisation des fonctions relatives à cette méthode implique d'abord une phase d'initialisation permettant de calculer les temps de réverbération. Les appels fonctions relatifs à cette phase sont :

```

/*Allocation et calcul des matrices directes et inverses*/

Room_MarkovMatrixAllocateDirect(&theRoom);
Room_MarkovMatrixComputeDirect(&theRoom,      theMMComputePa-
ram,theRenderingOption);
Room_MarkovMatrixAllocateComputeInverse(&theRoom,theRenderingOption);

/*Calcul des paramètres de décroissance de la réverbération*/

Room_MarkovMatrixComputeExpFactors(&theRoom, RtPtr, theRenderin-
gOption);

```

Le champ `theMMComputeParam` permet de préciser des options sur le calcul de la matrice de Markov directe (salle ouverte, ...). A titre d'exemple, dans le cas d'un rendu pour une bande de fréquence cette phase prend environ 7 s, sur un PowerMacintosh G3 à 233 MHz, pour une salle de 242 surfaces.

La phase de manipulation sera modifiée selon que le transducteur en mouvement est la source ou le transducteur. Dans le cas d'un mouvement de la source, il faut absolument recalculer l'éclairage de la salle et réappliquer l'amplification consécutive à cette modification.

```

/*Modification due à la source*/

IF_TransducerSource_IrradiateToRoom(&theSource,&theRoom,theRenderingOption);
IF_Room_DiffuseApplyMMAmplification(&theRoom,theRenderingOption);

/*Modification due au récepteur*/

IF_Room_IrradiateToTransducerReceiver(&theRoom,&theReceiver,R3ResPtr,theRenderingOption);

```

Si l'on suppose que ce modèle n'est pas alimenté directement par la source mais par les résidus en sortie d'un processus de sources images, on remplacera le premier appel par :

```

IF_ImageTree_IrradiateToRoom(&ImageTree,&theRoom,theRenderingOption);

```

Evaluation des performances

Dans le pire des cas (manipulation de la source), le calcul le plus intensif est celui des nouvelles valeurs de l'intensité sur les parois de la salle, et ce après calcul de l'amplification apportée à la salle. Ce calcul revient à la multiplication d'une matrice de taille N^2 (nombre de surfaces de la salle) par un vecteur de taille N , soit un total de N^2 multiplication additions.

La calcul des interactions avec les récepteurs nécessite l'utilisation de produits scalaires (utilisation de cosinus), distances et angles solides. Néanmoins la présence d'un cache de donnée associé aux transducteurs permet de ne calculer ces valeurs qu'une seule fois et de les réutiliser par la suite pour ces interactions. Pour chaque transducteur, le coût d'un appel de fonctions est de l'ordre de N , soit nettement inférieur au calcul de matrice précédent.

Ces calculs ne sont cependant pas exorbitants en regard de la puissance actuelle des processeurs. Néanmoins un certain nombre de tests de visibilité entre chaque surface et les transducteurs vient à nouveau alourdir cette charge de calcul.

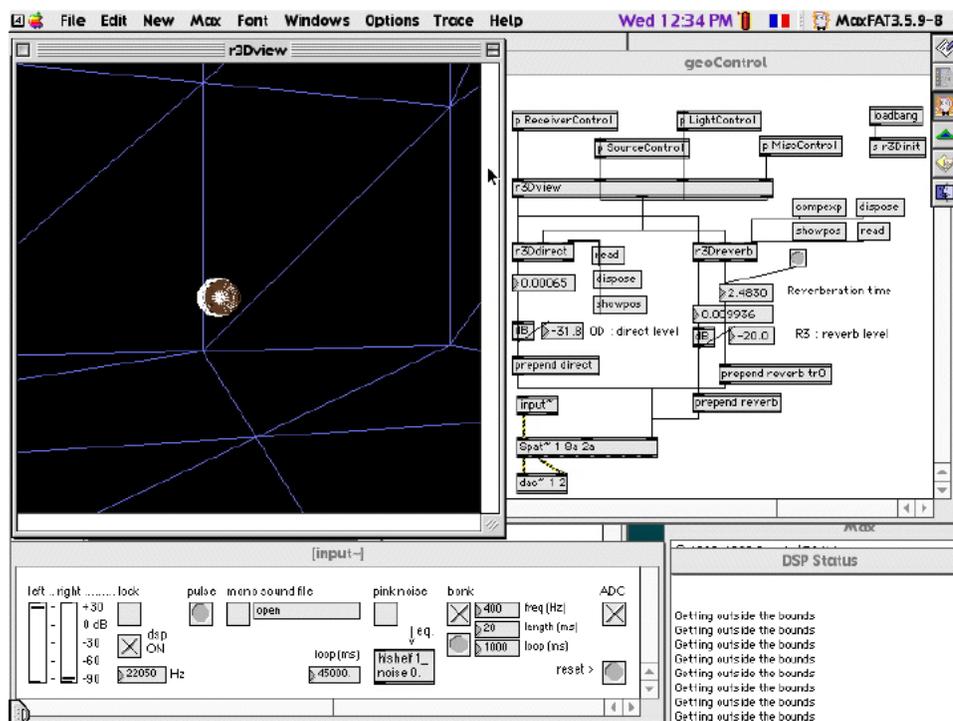


FIG. 7.2 – Utilisation des modules de commande géométrique dans l'environnement Spatialisateur

Chapitre 8

Conclusion

Le travail entrepris dans le cadre de cette thèse a permis avant tout d'éclaircir la notion de traitement temps réel dans le cadre de la manipulation d'un univers audio virtuel. Tout d'abord la notion de niveau de réalisme a été prise en compte permettant de dissocier la notion d'auralisation temps réel du cadre classique de l'acoustique architecturale. L'emploi de résultats issus de la psychoacoustique a notamment Elle s'est accompagnée de l'établissement d'une hiérarchie dans les informations contenues dans la réponse impulsionnelle qui a permis de considérer l'utilisation d'une représentation bas-niveau de celle-ci qualifiée de morphologique et regroupant l'information utile à son auralisation. L'emploi d'une telle représentation a également permis de découpler le traitement d'une scène sonore virtuelle du moteur de rendu utilisé pour en effectuer l'auralisation.

Un autre point important a été l'examen des contextes d'utilisation temps réel d'un effet de salle : on a pu ainsi séparer les situations de reproduction simple d'un effet enregistré de celles de manipulation et modification de l'effet de salle. Différents cadres de de manipulation ont ainsi été examinés (géométrique, perceptive, statistique, signal) et leur avantages et inconvénients évalués selon un ensemble de critères (interactivit'e, scalabilit]e, opérabilité, ...). Il a ensuite été possible d'examiner les liens entre de telles interfaces et la représentation morphologique proposée et d'y associer des méthodes permettant le traitement temps réel d'une modification de la scène sonore virtuelle. La séparation de la réponse impulsionnelle en différentes composantes a de de plus permis l'utilisation de méthodes adaptées, et séparées les unes des autres.

L'ensemble de ces résultat a été regroupé lors de l'élaboration de la structure d'une librairie de traitement de l'effet de salle, qui permet à un développeur de considérer différentes configurations sans se préoccuper des algorithmes mis en jeu, et sans limitation sur le moteur de rendu sonore utilisé. Celle-ci présente une interface de programmation cohérente à un développeur, et lui permet de spécifier et appréhender la qualit'e du résultat final.

D'autres sujets seraient à aborder dans la continuité de ce travail. On citera les trois points suivants :

- L'extension des recherches psychoacoustiques afin d'étendre la représentation morphologique de la réponse impulsionnelle à des espaces différents (espaces ouverts, petits volumes). Un travail portant sur la caractérisation des premières réflexions peut également être entrepris, afin de pouvoir y effectuer une sélection de l'information perceptivement utile. Ce type de procédure peut servir autant pour l'extraction des paramètres d'une représentation morphologique à partir d'une réponse impulsionnelle que pour une optimisation des algorithmes d'acoustique prévisionnelle utilisés pour le traitement de l'information issue d'une interface de manipulation géométrique.
- L'optimisation des procédures géométriques utilisées lors du traitement de données

issues d'une interface de ce type. On citera notamment les procédures de tests sur la visibilité des surfaces, utilisées pour les différents modules de rendu géométriques. La mise en place de procédures permettant de modifier dynamiquement le degré de discrétisation de la salle, et d'établir ainsi une hiérarchie entre les surfaces, peut également être envisagée afin d'optimiser les algorithmes de radiosit . On notera cependant que ces travaux concernent des fonctions pr cises et ne remettent pas en cause l'architecture de la librairie VASR.

- La coexistence conjointe de diff rentes interfaces de manipulation ne permet en effet pas dans le dispositif actuel une communication entre elles dans tous les cas. Un exemple particuli rement int ressant serait la communication entre une interface de type perceptive et de type g om trique, et ce en imposant des contraintes sur le mode de variation de celle-ci. L'utilisation de m thode, qualifi e de probl me inverse dans le domaine de l'acoustique architecturale pourrait servir de point de d part, de mani re similaire   l'utilisation de m thodes d'acoustique pr visionnelle.

Enfin le cas d'espace non compris dans le cadre de cette  tude (espaces ouverts, petites salles) serait    tudier par extension des m thodes propos es ici.

Chapitre 9

Annexes

Annexe A : Librairie EDR-1.5

Cette annexe présente les fonctions principales de la librairie EDR-1.5 d'analyse temps fréquence de l'effet de salle, en reproduisant les fichiers d'aide associés. **FONCTIONS D'ANALYSE GLOBALES**

[Tr,P0,TLim,PNoise,Tbegin,Sig,Sigtime,STFS,STFSfreq,STFStime, EDR] = EDRanalysis(mode,{submode}, data);

```
% EDRanalysis (EDR-1.5)
%
% [Tr,P0,TLim,PNoise,Tbegin,Sig,Sigtime,STFS,STFSfreq,STFStime, EDR] =
%                               EDRanalysis(mode,{submode}, data);
%                               (See explanation below for use)
%
% Computes a complete time-frequency analysis of a signal (Room Impulse
% response for example) using decay relief properties. The signal is
% supposed to be the sum of an exponential decreasing part and a
% corrupting noise. A theoretical model for the short-time Fourier
% transform of the signal is first computed and then the following
% parameters are estimated for each frequency
%   - Tr      : slope in seconde (reverberation time), versus
%               frequency
%   - P0      : Initial Power Spectrum versus frequency
%   - PNoise  : Energy of the corrupting noise
%   - TLim    : Limit time defined as the time where the level of
%               the signal is equal to the level of corrupting
%               noise
%
% The 'mode' parameters allows this function to be called in different
% ways. Values can be
%
%   's' or 'straight' : parameters or given directly
%   'f' or 'file'     : read the parameters in a file
%   'd' or 'dir'      : do the analysis for an entire directory
%                       (reserved for future versions)
%
% +'straight' mode has 2 submodes :
```

```

% * 'n' or 'name' : a filename is given with its extension, and
% then opened (for more explanations about reading sound files
% see sigRead). Only mono files are accepted. An example call
% would be :
%     EDRanalysis('s', 'n', 'myfile.aif',{win_type', nfft,
%                                     overlap,Trev, NRepeat})
%
% *'d' or 'data' : data are given directly;
% Data are directly given. These data are : the signal (Sig), the
% time scale associated with the signal (Time), and the sampling
% rate of the signal (sr). If one of these 3 data was missing, an
% error message would be output.
%     EDRanalysis('s', 'd', Sig,Time, sr,{win_type', nfft,
%                                     Trev, NRepeat})
%
% +'file' mode allows parameters to be read from a file. Only filenames can
% be stored (no data). In this mode an option to save the result of
% EDRanalysis. in a Matlab file (.mat) is permitted. In order to select
% the data to save, a filter can be precised. See the example file
% 'data.m', and more about filters in the 'README' file
%
%     EDRanalysis('f','filedata.m');
%
% +'dir' mode allows to do analysis on all files in a directory. This mode is
% not available in EDR-1.5, but will be implemented in future versions
%
% Possible input are :
%
% 'filein' : name of thefile containing the signal to be analyzed
%           Must have an extension. Obligatory field!
% win_type : name of the kind of window used for short-time Fourier transform
%           analysis. Choices are 'harris', 'triang', 'bartlett', 'kaiser',
%           'blackman', 'hamming', 'hanning'. Default is 'harris'.
% nfft      : number of FFT points; default = 512
% overlap   : overlap rate of the windows in STFS; default is 0.75
% Trev      : time (ms) where the reverberation begins (referenced to direct sound)
%           default is 100 ms
% Nrepeat   : number of iterations in the computing process (default is 12)

```

```

%
% Output are :
% Tr      : reverberation time (versus frequency)
% P0      : initial Power Spectrum (versus frequency)
% TLim    : limit time (Noise level = Signal level)
% PNoise  : noise level (versus frequency)
% Tbegin  : arrival time of the direct sound
% Sig     : signal to be analyzed
% Sigtime : time scale corresponding to Sig
% STFS    : Short time Fourier spectrum of the signal
% STFSFreq : frequency scale associated with STFS
% STFStime : temporal scale associated with STFS
% EDR     : Energy Decay Relief
%
% See also : EDRanalysis_Initialize, EDRanalysis_Iteration, edr, STFS, RI2Spat

```

[Tr,P0,TLim,PNoise,Tbegin,Sig,Sigtime,STFS,STFSfreq,STFStime,EDR,FOD,OD,R1,R2,R3,MatSpat,face,omni,crit,fd] = RI2Spat((mode,{submode}, data)

```

% RI2SPAT (EDR-1.5)
% [Tr,P0,TLim,PNoise,Tbegin,RI,RItime,STFT,STFTfreq,STFTtime, EDR,FOD,OD,R1,
%      R2,R3,MatSpat,face,omni,crit,fd] = RI2Spat((mode,{submode}, data);
%
%      (See explanation below for use)
%
% Analyses an impulse response, and saves a Max patch named for configuring a
% Spat_0Per module (in Spat-0.2 or Spat-1.x). This analysis is based on the
% Energy deccay relief analysis. The results of the analysis can be stored
% in a MAT-file (Matlab data)
%
% The 'mode' parameters allows this function to be called in different ways.
% Values can be
%      's' or 'straight' : parameters or given directly
%      'f' or 'file'     : read the parameters in a file
%      'd' or 'dir'     : do the analysis for an entire directory (reserved for
%                        future versions)
%
% +'straight' mode has 2 submodes :
%      * 'n' or 'name' : a filename is given with its extension, and then opened

```

```

%      (for more explanations about reading sound files see sigRead). Only mono
%      files are accepted. An example call would be :
%      RI2Spat('s', 'n', 'myfilein',{ 'myfileout',Times,'win_type',
%                                     nfft, overlap,Ndelays})
%
% *'d' or 'data' : data are given directly;
%      Data are directly given. These data are : the signal (RI), the time scale
%      associated with the signal (Time), and the sampling rate of the signal (sr).
%      If one of these 3 data was missing, an error message would be output.
%      RI2Spat('s', 'n',RI, RItime,sr,'myfileout',{Times,'win_type',
%                                                  nfft, overlap,Ndelays})
%
% +'file' mode allows parameters to be read from a file. Only filenames can
% be stored (no data). In this mode an option to save the result of EDRanalysis
% in a Matlab file (.mat) is permitted. In order to select the data to save, a
% filter can be precised. See the example file 'data.m', and more about
% filters in the 'README' file
%      RI2Spat('f', 'filedata.m');
%
% +'dir' mode allows to do analysis on all files in a directory. This mode is
% not available in EDR-1.5, but will be implemented in future versions
%
% Possible input are :
%
% 'myfilein' : name of the impulse response file to be analyzed. Must
%             have an extension. Obligatory field!
% 'myfileout' : name of the output patch file. This name must end with .pat
%             otherwise this suffix will be automatically added
% Times      : [t1 t2 t3] begin times of R1, R2, R3
%             (in ms, relative to direct sound). Default = [20 40 100]
% win_type   : name of the kind of window used for short-time Fourier
%             transform analysis. Choices are 'harris', 'triang',
%             'bartlett', 'kaiser','blackman','hamming', 'hanning'.
%             Default is 'harris'.
% nfft       : number of FFT points; default = 512
% overlap    : overlap rate of the windows in STFT; default is 0.75
% Ndelays    : number of feedback delays           Default = 8
%

```

```
%
% [Tr,P0,TLim,PNoise,Tbegin,RI,RItime,STFT,STFTfreq,STFTtime, EDR,FOD,OD,R1,
%     R2,R3,MatSpat,face,omni,crit,fd]
%
% See also : EDRanalysis, EDR2Ener, Ener2MatSpat, MatSpat2crit, fdens,
%     crit2patch
```

[Tr,P0,TLim,PNoise, Tbegin, RI, RItime, RIsynt,STFT, STFTfreq, STFTtime,EDR] = RIdenoising(mode,{submode}, data)

```
% [Tr,P0,TLim,PNoise, Tbegin, RI, RItime, RIsynt,STFT, STFTfreq,
%     STFTtime,EDR] = RIdenoising(mode,{submode}, data);
%
% Eliminates the noise in a Room Impulse response, using a Energy Decay Relief
% analysis.For each frequency channel, the end of the Energy decay curve is
% replaced by an artificial signal (white noise with a decreasing exponential
% value). Then the inverse Fourier transform is computed, allowing to obtain
% a denoised room impulse response
%
% The 'mode' parameters allows this function to be called in different ways.
% Values can be
%     's' or 'straight' : parameters or given directly
%     'f' or 'file'     : read the parameters in a file
%     'd' or 'dir'     : do the analysis for an entire directory (reserved
%                       for future versions)
%
% +'straight' mode has 2 submodes :
%     * 'n' or 'name' : a filename is given with its extension, and then opened
%       (for more explanations about reading sound files see sigRead). Only mono
%       files are accepted. An example call would be :
%     RIdenoising('s', 'n', 'myfilein','myfileout',{'win_type', nfft, overlap,
%                                               Trev,NRepeat})
%
%     *'d' or 'data' : data are given directly;
%       Data are directly given. These data are : the signal (Sig), the time scale
%       associated with the signal (Time), and the sampling rate of the signal (sr).
%       If one of these 3 data was missing, an error message would be output.
%     RIdenoising('s', 'd', Sig,Time, sr,'myfileout', {'win_type', nfft,
```



```
%
% The output signal can be saved in another file format but sampling rate can't
% be changed.
%
% See also : EDRanalysis, EDRanalysis, edr, stft, RI2Spat
```

ESTIMATION DES PARAMÈTRES D'UN CANAL FRÉQUENTIEL

```
[Tr, P0, PNoise, InitError] = EDRanalysis_Initialize(ETC,ETCtime,N,nfft,
roverlap)
```

```
% EDRanalysis_Initialize
%
% [Tr, P0, PNoise, InitError] =
%             EDRanalysis_Initialize(ETC,ETCtime,N,nfft, roverlap)
%
% Having a single short-time Fourier transform channel(ETC) of an room impulse
% response, this function computes values of reverberation time (Tr),
% Initial Power Spectrum (P0), and corrupting noise (PNoise) for that channel.
% This is made by comparing energy levels in different sections of the channel
% (segmentation in N sections), and then using the Energy Decay Curve properties.
% An error is computed, as the quadratic difference between the STFT Channel
% (ETC) and the theoretical curve obtained using the estimated parameters
% (exponential decreasing curve ETC=P0*exp(-60/Tr)).
%
% In order to have better estimation, only the diffuse part of the ETC should be
% passed to this routine (pre-elimination of the first reflections, ETC is
% truncated).
%
% ETC      : channel of the short-time Fourier transform of the signal, with first
%            reflections eliminated
% ETCtime  : corresponding time scale
% N        : number of division made in ETC for the initialization (segmentation)
% nfft     : number of points used in the FFT for the short-time Fourier transform
% roverlap : overlap rate of the windows during STFT
%
% Tr       : Value of Tr after the initialization
```

```
% P0      : Value of P0 (Initial Power Spectrum) after the initialization
% PNoise  : Value of the energy of the corrupting noise after the initialization
%
% See also : EDRanalysis_Iteration, EDRanalysis, edr, stft.
```

TRANSFORMATIONS TEMPS-FRÉQUENCE

[STFT, STFTtime, STFTfreq] = stft(sig, sr, win, nfft, roverlap)

```
% STFT
%
%      [STFT, STFTtime, STFTfreq] = stft(sig, sr, win, nfft, roverlap)
%
% Compute the short time Fourier transform of a signal 'sig'.
% The beginning of the signal is zero-padded with Nfft/2 zeros
% so that the beginning of the signal is not ignored. The result
% 'STFT' is a complex matrix, whose lines represents one FFT
% channel (isofrequency) and raws one FFT. Two additionnals vectors
% 'time' and 'freq' represent the scales associated with STFT. It
% uses the matlab function 'specgram'
%
% sig      : input signal
% sr       : Sampling rate of the signal 'sig'
% win      : Vector of the analysis window
% nfft     : Number of points in FFT computing;
%           must be equal to the length of win
% roverlap : Overlap rate of the windows
%
% STFS     : Matrix of the Short time fourier transform.
% time     : Temporal scale associated with the STFT (in s)
% freq     : Frequency scale associated with the STFT (in Hz)
%
% Default  : fe = 1, win : Blackmann-Harris window (order 4),
%           nfft = 1024, roverlap = 0.75;
%
% See also : STFS, SPECGRAM
```

[STFS, STFStime, STFSfreq] = stfs(sig, fe, win, nfft, roverlap)

```

% STFS
%
%   [STFS, STFStime, STFSfreq] = stfs(sig, fe, win, nfft, roverlap)
%
% Compute the short time Fourier spectrum of a signal 'sig'.
% The beginning of the signal is zero-padded with Nfft/2 zeros
% so that the beginning of the signal is not ignored. The result
% represents an energy. It is stored in a matrix, whose lines
% represents one FFT channel (isofrequency) and raws one FFT.
% Two additionnals vectors 'time' and 'freq' represent the scales
% associated with STFS. It uses the 'STFT' function
%
% sig      : input signal
% fe       : Sampling rate of the signal 'sig'
% win      : Vector of the analysis window
% nfft     : Number of points in FFT computing;
%           must be equal to the length of win
% roverlap : Overlap rate of the windows
%
% STFS     : Matrix of the Short time fourier spectrum.
% time     : Temporal scale associated with the STFS (in s)
% freq     : Frequency scale associated with the STFS (in Hz)
%
% Default  : fe = 1, win : Blackmann-Harris window (order 4),
%           nfft = 1024, roverlap = 0.75;
%
% See also : STFT, SPECGRAM

```

[sig] = istft(STFTtime, STFTfreq, STFT, nfft, origwin, roverlap)

```

% ISTFT
%
%   [sig] = istft(STFTtime, STFTfreq, STFT, nfft, origwin, roverlap);
%
% Compute the inverse Fourier transform of a short-time Fourier spectrum

```

```

%
%
% STFTtime   : Time scale for the short-time Fourier transform (seconds)
% STFTfreq   : Frequency scale for the short-time Fourier transform (Hz)
% STFT       : Short-time Fourier transform signal
% nfft       : Number of points in FFT computing of the short-time Fourier
%             transform
% originwind  : window used in the short-time Fourier transform
% roverlap    : overlap rate of the windows during
%
% sig        : output signal. Last samples (roverlap*nfft) are replaced
%             with 0.
%
% See also : STFT, STFS

```

[EDR] = edr(nfft,roverlap,STFT,STFTtime,{PNoise,P0,Tr,TLim});

```

% EDR
%
% [EDR] = edr(nfft,roverlap,STFT,STFTtime);
%   [EDR] = edr(nfft,roverlap,STFT,STFTtime,PNoise);
%   [EDR] = edr(nfft,roverlap,STFT,STFTtime,PNoise,P0,Tr,TLim);
%
% Compute the energy decay relief (EDR) for a given short time Fourier
% transform. It can be used for only one channel of STFT too. The number
% of input arguments can vary, allowing different uses.
%
%   - 4 arguments : computation of the EDR
%   - 5 arguments : before computation, a constant value (corruption
%                   noise PNoise) is subtracted to each channel of the
%                   STFT
%   - 8 arguments : same as above, but a theoretical value correponding
%                   to an exponential decay is output when the result
%                   would have been corrupted by noise
%
% nfft       : Number of points in FFT computing for the short-time
%             Fourier transform
% roverlap    : Overlap rate of the windows

```

```

% STFT      : short-time Fourier transform to be integrated
% STFTtime  : time scale corresponding to the STFT time dimension
% PNoise    : Corrupting noise, (Computed after a call to EDRanalysis).Its
%            lengths must be the same as the frequency dimension of STFT
% PO        : Initial Power Spectrum(computed after a call to EDRanalysis)
% Tr        : Reverberation time (computed after a call to EDRanalysis)
% TLim      : Noise Limit Time (computer after a call to EDRanalysis)
%
% See also : STFT, STFS, EDRanalysis

```

[Tr, PO, PNoise, Tlim, Error, EDRval, ResVal] = EDRanalysis_Iteration(ETC, ETCtime, Tr, PO, PNoise, nfft, roverlap)

```

% EDRanalysis_Iteration
%
% [Tr, PO, PNoise, Tlim, Error, EDRval, ResVal] =
%     EDRanalysis_Iteration(ETC, ETCtime, Tr, PO, PNoise, nfft, roverlap)
%
% Compute a simple iteration in a EDR analysis process. Given a STFT (ETC) channel
% and initial values for Tr, PO and PNoise, a new estimation of parameters
% Tr and PO is made by substracting the noise (PNoise) and integrating the result
% (Energy Decay Curve). Using these new values, a new value of PNoise is estimated.
% TLim is defined as the time where the energy level of corrupting noise is equal
% of the useful signal. An error is computed, as the quadratic difference between
% at the energy level the STFT Channel (ETC) and the theoretical curve obtained
% using the estimated parameters(exponential decreasing curve ETC=PO*exp(-60/Tr)).
%
% In order to have better estimation, only the diffuse part of the ETC should be
% passed to this routine (pre-elimination of the first reflections, ETC
% is truncated).
%
% ETC      : channel of the short-time Fourier transform of the signal, with first
%            reflections eliminated
% ETCtime  : corresponding time scale
% Tr       : Value of Tr before the iteration
% PO       : Value of PO (Initial Power Spectrum) before the iteration

```

```

% PNoise : Value of the energy of the corrupting noise before the iteration
% nfft   : number of points used in the FFT for the short-time Fourier transform
% roverlap: overlap rate of the windows during STFT
%
% Tr     : Value of Tr after the iteration
% PO     : Value of PO (Initial Power Spectrum) after the iteration
% PNoise : Value of the energy of the corrupting noise after the iteration
% Tlim   : Corresponding limit Time
% Error  : error between the ETC signal and a theoretical signal computed
%         with the new values of Tr, PO and PNoise
% EDRVal : value of the origin of the Energy Decay Curve for the channel
% ResVal : computing error (used by the smoothing function CleanPO)
%
% See also : EDRanalysis_Initialize, EDRanalysis, edr, stft.

```

FONCTIONS D'ENTRÉES-SORTIES

[sig,time,fe,sigsize, chan]= sigRead(varargin)

```

% SIGREAD
%
% function [sig,time,sr,sigsize,chans]=sigRead(name,{form,chanNb,prec})
%
% name      : filename with or without extension. Complete filename must
%            be provided (path included).
% form      : file format, needed if RIname is provided without extension
%            - 'snd': NeXT sound file format
%            - 'au' : Sun sound file format (equivalent to NeXT)
%            - 'max': NeXT sound issued of Max (due to a bug)
%            - 'rif': AMS format. Requires two files file.inf and
%                    file.rif
%            - 'inf': same as 'rif'
%            - 'aif': AIFF format,(or AIFC without compression)
%                    (Apple, SGI)
%            - 'raw': raw data, requires precision argument. Time
%                    and Fe won't be output. Must be there to read
%                    raw data

```

```

%           - 'wav': Wav format (Windows95, Windows 98, Windows NT)
%
% chanNb   : if the sound file is not a mono one, allows to select a
%           channel, otherwise no output is given. If an argument 0 is
%           given, a matrix of all channels will be output.
% prec     : precision of the data for the 'raw' option : 'int8', 'int16',
%           'int32', 'float32', 'float64'
%
% Parameters form, chanNb and prec are optionnal for many data types
%
% sig      : read signal
% time     : time base asoociated with Sig
% fe       : sample rate
% sigsize  : number of frames in the file
% chan     : number of channel in the file

```

[Value]=sigWrite(x,name,sr,{form,res,end})

```

% SIGWRITE
%
% function [Value]=sigWrite(x,name,sr,{form,res,end})
%
% x        : vector to be output
% name     : filename with or without extension. Complete filename must
%           be provided (path included)
% sr       : sampling rate
% form     : file format, needed if Rlname is provided without extension
%           - 'snd': NeXT sound file format
%           - 'au' : Sun sound file format (equivalent to NeXT)
%           - 'rif': AMS format. Requires two files file.inf and
%                 file.rif
%           - 'inf': same as 'rif'
%           - 'aif': AIFF format,(or AIFC without compression)
%                 (Apple, SGI)
%           - 'raw': raw data, requires precision argument. Sampling
%                 rate argument won't be used
%           - 'wav': Wav format (Windows95, Windows 98, Windows NT)
% res      : resolution : 'int8', 'int16'(default), 'int32',float32, float64

```

```
%
% Parameters 3 to 6 are optionnal. Only mono files are considered now.
%
% See also : sigRead
```

TRASCRPTION VERS LES FACTEURS PERCEPTIFS

[FOD, OD, R1, R2, R3] = EDR2Ener(EDR,STFTtime, STFTfreq, Tbegin, Tlimits)

```
% EDR2Ener (EDR-1.5)
%
% [FOD, OD, R1, R2, R3] = EDR2Ener(EDR,STFTtime, STFTfreq, Tbegin, Tlimits)
%
% This function computes the energy (versus frequency) of the differents
% parts of a room impulse response (FOD, OD, R1, R2, R3)
%
% EDR      : Energy decay relief of the room impulse response
% STFTtime : time scale associated with the EDR
% STFTfreq : frequency scale associated with the EDR
% Tbegin   : arrival time of the direct sound
% Tlimits  : begin times of the different sections of the impulse
%           room response : [t1 t2 t3] corresponding to the beginnings
%           of section R1, R2 , R3. Defaults is [20 40 100]
%
% See also : EDRanalysis, RI2Spat, Ener2MatSpat, MatSpat2crit
```

[EnerMat] = Ener2MatSpat(OD, R1, R2, R3, Tr, STFTfreq);

```
% Ener2MatSpat (EDR-1.5)
%
% [EnerMat] = Ener2MatSpat(OD, R1, R2, R3, Tr, STFTfreq);
%
% Computes a matrix wich can be used to set the Spatialisateur(TM)
% at a low-level. The matrix is a reduction in 3 frequency bandths
% of OD, R1, R2, and R3. The last row of EnerMat is a 3 bandths
% reduction of the reverberation time.
%
```

```

% OD : energy for the direct sound (versus frequency)
% R1 : energy for the first reflections (versus frequency)
% R2 : energy for the cluster(versus frequency)
% R3 : energy for the late reverb sound (versus frequency)
% Tr : reverberation time (versus frequency)
% STFTfreq : frequency scale associated with the data above
%
% EnerMat : Output matrix (5 rows, 3 columns)
%
% See also : EDRanalysis, RI2Spat, Ener2MatSpat, MatSpat2crit, Smooth8va

```

[face, omni, crit] = MatSpat2crit(EnerMat, offset);

```

% MatSpat2crit (EDR-1.5)
%
% [face, omni, crit] = MatSpat2crit(EnerMat, offset);
%
% Converts the energetical matrix to Spatialisateur criteria
%
% See also : EDRanalysis, RI2Spat, Ener2MatSpat, MatSpat2crit

```

fact] = crit2fact(crit)

```

% crit2fact (EDR-1.5)
%
% [fact] = crit2fact(crit)
%
% Computes a set of perceptual factors from a set of energetical
% criteria, in order to do a setting for the Spatialisateur
%
% See also : EDRanalysis, RI2Spat, Ener2MatSpat, MatSpat2crit,fdens

```

[Result] = crit2patch(name,face,omni,crit,Times,fd,Ndelays)

```

% crit2patch (EDR-1.5)
%

```

```

% [Result] = crit2patch(name,face,omni,crit,Times,fd,Ndelays)
%
% Creates a Max patch from a set of Spatialisateur criteria, with
% a message box which can be send to a Spatialisateur
%
% See also : EDRanalysis, RI2Spat, Ener2MatSpat, MatSpat2crit,fdens

```

[Result] = fact2patch(name,face,omni,fact,Times,fd,Ndelays, distance)

```

% fact2patch (EDR-1.5)
%
% [Result] = fact2patch(name,face,omni,fact,Times,fd,Ndelays, distance)
%
% Creates a Max patch from a set of Spatialisateur perceptual factors,
% with a message box which can be send to a Spatialisateur
%
% See also : EDRanalysis, RI2Spat, Ener2MatSpat, MatSpat2crit,fdens,
%           crit2patch, crit2fact

```

[fd,fd_min,fd_max,fd_tdens,dmax,tdens] = fdens(Tr,dmin,N,tdens)

```

% FDENS
%
%   [fd,fd_min,fd_max,fd_tdens,dmax,tdens] = fdens(Tr,dmin,N,tdens)
%
% Computes recommended value of modal density (average number of modes
% per Hz) for a feedback delay network reverberator, according to
% smallest delay length, reverberation time and number of delay units.
%
% Arguments:
% Tr : mid-frequency reverberation time (s)
% dmin : smallest delay length (ms)
% N : number of delay units (optional).
% tdens : minimum echo density after 3 loops (kHz) --default = 3.
%
% Results:
% fd_min,fd_max : min and max values for smooth time response

```

```
% fd_tdens : max value for satisfying tdens requirement
% dmax : longest delay length (ms)
% tdens : actual time density after 3 loops (kHz)
%
% See also : EDRanalysis, RI2Spat, Ener2MatSpat, MatSpat2crit
```

[ODtime, ODTimeSample] = TimeOD(RI,sr)

```
% TimeOD
%
% [ODtime, ODTimeSample] = TimeOD(RI,sr)
%
% Detect the arrival time of the direct signal in a room
% impulse response
%
% RI : Room Impulse response signal
% sr : sampling rate of the Impulse response
%
% ODtime : arrival time in seconds of the direct signal
% ODTimeSample : arrival time in samples of the direct signal
```

?? ?? ?? ?? ?? ?? ?? ?? ?? ?? ?? ?? ?? ?? ??

[Bru98] [CB99] [CB91] [Dal92] [Dal95] [EC96] [Far95a] [Far95b] [Gil88] [Hei93]
[JS73] [Kir84] [Kut92] [Kut93] [Leh93] [LL88] [MM93] [Nay92] [Nay93] [Ogu92]
[Lyn98] [Fos86] [Gar94] [Ger76] [Ger77] [VH98] [REWJ98] [JR98] [Pol92]
[Pol93] [Jot92] [Jou97] [JGS83] [Kah95] [Lav89] [Leh75] [Mal86] [Mal90]
[Pol88a] [Pie81] [Ros86] [Sch62] [Sch65] [Sch79a] [Sch79b] [SP87] [Ste92]
[Tsi98]

Bibliographie

- [App96] Apple Computer, Inc. Apple game sprocket guide. Technical report, The Apple Publishing System, 1996.
- [Bar71] M. Barron. The subjective effect of first reflections in concert halls - the need for lateral reflections. *J. Sound. Vib.*, 15(4):475–494, 1971.
- [Bar74] M. Barron. *The effects of early reflections on subjective acoustical quality in concert halls*. PhD thesis, Univ. of Southampton, 1974.
- [Beg94] D.R. Begault. *3D Sound for virtual reality and multimedia*. AP Professional, 1994.
- [Ber62] L. Beranek. *Music, Acoustics and Architectures*. John Wiley, 1962.
- [BL88] M. Barron and L.J. Lee. Energy relations in concert auditoriums. *Journal of the Acoustical Society of America*, 84:618–628, 1988.
- [BM81] M. Barron and A.H. Marshall. Spatial impression due to early lateral reflections in concert halls: the derivation of a physical measure. *J. Sound. Vib.*, 77(2):211–232, 1981.
- [Bor84] J. Borish. *Electronic simulation of auditorium acoustics*. PhD thesis, CCRMA, Department of Music, Stanford University, 1984.
- [Bru98] M. Bruneau. *Manuel d'acoustique fondamentale*. Hermes, Paris, France, 1998.
- [CB91] C.S Ciskowski and C.A. Brebbia. *Boundary Element Methods in Acoustics*. Computational Mechanics Publication co-published with Elsevier Applied Science, 1991.

- [CB99] F. Cruz-Barney. *Evaluation des performances d'un environnement informatique d'acoustique prévisionnelle*. PhD thesis, Laboratoire d'acoustique de l'Université du Maine, Le Mans, 1999.
- [Dal92] B.I. Dalenbäck. *Room acoustic prediction and auralization based on an extended image source model*. PhD thesis, Chalmers University of Technology, Goteborg, 1992.
- [Dal95] B.I. Dalenbäck. Room acoustics prediction and auralization based on a unified treatment of diffuse and specular reflections. *15th Intl. Congress on Acoustics, Trondheim, Norway*, pages 425–428, 1995.
- [DH89] P. D'Antonio and J.H. Honnert. Complex time-response measurements using time-delay spectrometry. *J. Audio Eng. Soc.*, 37(9):674–690, 1989.
- [EC96] V. Easwaran and A. Craggs. An application of acoustics finite element models to finding the reverberation times of irregular rooms. *Acta Acoustica*, 82:54–64, 1996.
- [Far95a] A. Farina. Pyramid tracing vs. ray tracing for simulation of sound propagation in large rooms. *Intl. Conf. on Computational Acoustics and its Environmental Applications*, pages 1–8, 1995.
- [Far95b] A. Farina. Verification of the accuracy of the pyramid tracing algorithm by comparison with experimental measurements of objective aic parameters. *15th Intl. Congress on Acoustics, Trondheim, Norway*, pages 445–448, 1995.
- [Fos86] S. Foster. Impulse response measurements using golay codes. *Proc. ICASSP Tokyo*, pages 929–932, 1986.
- [Gar94] W.G. Gardner. Efficient convolution without input/output delay. *AES*, 1994.
- [Ger76] M.A. Gerzon. Unitary energy preserving multichannel networks with feedback. *Electronics Letters V*, 12(11), 1976.
- [Ger77] M.A. Gerzon. Multi-system ambisonic decoder', part 1: 'basic design philosophy. *Wireless World*, 83(1499):43–47, 1977.
- [Gil88] E.N. Gilbert. Ray statistics in reverberation. *J. Acoust. Soc. Am.*, 83(5):1804–1808, 1988.

- [Hei93] R. Heinz. Binaural room simulation based image source model with addition of statistical methods to include the diffuse sound scattering of walls and to predict the reverberant trail. *Applied Acoustics*, 38:145–159, 1993.
- [JGS83] J-P. Jullien, A. Gilloire, and A. Saliou. Caractérisation d'une méthode de mesure de réponse impulsionnelle en acoustique des salles. *I.C.A.*, 6:217–220, 1983.
- [Jot92] J-M. Jot. *Etude et réalisation d'un spatialisateur de sons par modèle physique et perceptif*. PhD thesis, Ecole Nationale Supérieure des Télécommunications, 1992.
- [Jou97] J. Jouhaneau. *Acoustique des salles et sonorisation*. Collection CNAM, Lavoisier TEC & DOC, 1997.
- [JR98] J-M Jot and J-B. Rault. Extension of audio bifs: Interface and models. integrating geometrical and perceptual paradigms for the environmental spatialization of audio. *ISO/IEC/JTC1/SC29/WG11 m4223*, 1998.
- [JS73] S. Juricic and F. Santon. Images et rayons sonores dans le calcul numérique des échogrammes. *Acta Acustica*, 84(4):689–700, 1973.
- [Kah95] E. Kahle. *Validation d'un modèle objectif de la perception de la qualité acoustique dans un ensemble de salles de concerts et d'opéras*. PhD thesis, Laboratoire d'acoustique de l'Université du Maine, Le Mans, 1995.
- [Kir84] J. Kirszenstein. An image source computer model for room acoustics analysis and electroacoustics simulation. *Applied Acoustics*, 17:275–290, 1984.
- [Kut92] K.H. Kuttruff. Simulation models for auralization in room acoustics. *14th Intl. Congress on Acoustics, Beijing, China*, 1992.
- [Kut93] K.H. Kuttruff. Auralization of impulse responses modeled on the basis of ray-tracing results. *J. Audio. Eng. Soc.*, 41(11):876–880, 1993.
- [Kut95] K.H. Kuttruff. A simple iteration scheme for the computation of decay constants in enclosures with diffusely reflecting boundaries. *J. Acoust. Soc. Am.*, 98(1):288–293, 1995.

- [Lab99] Creative Labs. Environmental audio extensions: Eaax2.0. Technical report, Creative Labs - EMU, 1999.
- [Lav89] C. Lavandier. *Validation perceptive d'un modèle objectif de caractérisation de la qualité acoustique d'une salle*. PhD thesis, Labowratoire d'acoustique de l'Université du Maine, Le Mans, 1989.
- [Leh75] U. Lehmann. *Untersuchungen zur Bestimmung des Raumseindruck bei Musikdarbietungen und Grundlagen der Optimierung*. PhD thesis, TU Dresden, 1975.
- [Leh93] H. Lehnert. Systematic errors of the ray-tracing algorithm. *Applied Acoustics*, 38:161–178, 1993.
- [LL88] H. Lee and B Lee. An algorithm for the image model technique. *Applied Acoustics*, 24:85–115, 1988.
- [Lyn98] C. Lyngé. Odeon room acoustics program, version3.1. Technical report, The Department of Acoustic Technology, Technical University of Denmark, 1998.
- [Mal86] C. Malcurt. *Simulations informatiques pour prédire les critères de qualification acoustique des salles. Comparaison des valeurs mesurées et calculées dans une salle à acoustique variable*. PhD thesis, Université de Toulouse, 1986.
- [Mal90] D.G. Malham. Ambisonics - a technique for low cost, high precision, three dimensional sound diffusion. *ICMC Glasgow Proceedings*, pages 118–120, 1990.
- [MM93] D. V. Maercke and J. Martin. The prediction of echograms and impulse responses within the epidaure system. *Applied Acoustics*, 38:03–114, 1993.
- [Nay92] G. Naylor. Treatment of early and late reflections in a hybrid computer model for room acoustics. *124th ASA meeting, New Orleans, USA*, 1992.
- [Nay93] G. Naylor. Odeon - another hybrid room acoustical model. *Applied Acoustics*, 38:131–143, 1993.

- [NNi96] R. NNicol. Implantation et évaluation d'un traitement d'antenne acoustique. Technical report, Conservatoire National des Arts et Métiers, 1996.
- [No99] M. Noël. Implémentation du module de mesure et de compensation du contexte d'écoute du spatialisateur de l'ircam: Time delay spectrometry et déconvolution d'enveloppe énergétique. Technical report, 1999.
- [Ogu92] K. Oguchi. Ray tracing simulation technique regarding distribution of early reflections for studying fundamental room shape of concert halls. *14th Intl. Congress on Acoustics, Beijing, China*, 1992.
- [Pie81] A.D. Pierce. *Acoustics, an introduction to its physical principles and applications*. Acoustical Society of America, 1981.
- [Pol88a] J-D. Polack. *La transmission de l'énergie sonore dans les salles*. PhD thesis, 1988.
- [POL88b] Mark A. POLETTI. Linearly swept frequency measurements, tds and the wigner distribution. *JAES*, 36(6):457–468, June 1988.
- [Pol92] J-D. Polack. Modifying chambers to play billiard: the foundations of reverberation theory. *Acustica*, 76:257–272, 1992.
- [Pol93] J-D. Polack. Playing billiards in concert hall: the mathematical foundations of geometrical room acoustics. *Applied Acoustics*, 38:235–244, 1993.
- [REWJ98] J-B. Rault, M. Emerit, O. Warusfel, and J-M. Jot. Audio rendering of virtual room acoustics and perceptual description of the auditory scene. *ISO/IEC/JTC1/SC29/WG11 m4222*, 1998.
- [Ros86] M. Rossi. *Electroacoustique*. Presses Polytechniques romandes, 1986.
- [Sch62] M.R. Schroeder. Natural sounding artificial reverberation. *J. Audio Eng. Soc.*, 10(3):219–223, 1962.
- [Sch65] M.R. Schroeder. New method for measuring reverberation time. *J. Acous. Soc. Am.*, 37:409–412, 1965.

- [Sch79a] M.R. Schroeder. Integrated impulse method for measuring sound decay without using impulses. *J. Audio Soc. Am.*, 66(2):497–500, 1979.
- [Sch79b] M.R. Schroeder. Integrated impulse method for measuring sound decay without using impulses. *J. Audio Soc. Am.*, 66(2):497–500, 1979.
- [SP87] J. Stautner and M. Puckette. Statistical parameters of the frequency response curves of large rooms. *J. Acoust. Soc. Am.*, 35(5):299–306, 1987.
- [Ste92] U.M. Stepenson. Diffraction algorithms for ray-tracing: discussion of some approaches. *14th Congress on Acoustics, Beijing, China*, 1992.
- [Tsi98] N. Tsingos. *Simulations de champs sonores de haute qualité pour des applications graphiques de haute qualité*. PhD thesis, Université Joseph Fourier - Grenoble 1, 1998.
- [Van94] J. Vanderkooy. Aspects of mls measuring systems. *JAES*, 42(4):219–231, April 1994.
- [VH98] R. Väänänen and J. Huopaniemi. Spatial representation of sound in scene description languages. 1998.