

THÈSE DE DOCTORAT DE L'UNIVERSITÉ PARIS 6

Spécialité :
Acoustique

Sujet de la thèse :

Caractérisation et optimisation de la restitution haute-fidélité en véhicule

Présentée par

Guillaume VANDERNOOT

pour obtenir le grade de docteur de l'université Paris 6

Soutenue 5 décembre 2001
devant le jury composé de :

René CAUSSÉ	<i>Directeur de thèse</i>
Jacques JOUHANEAU	<i>Rapporteur</i>
Éric LE CHEVALIER	<i>Examineur</i>
Xavier MEYNIAL	<i>Rapporteur</i>
Jean-Dominique POLACK	<i>Examineur</i>
Olivier WARUSFEL	<i>Examineur</i>

« *L'émerveillement est le premier pas vers le respect* »

Nicolas Hulot

Remerciements

J'exprime ma profonde reconnaissance à M. René Caussé pour m'avoir fait découvrir l'acoustique musicale, et m'avoir accueilli à l'IRCAM comme stagiaire en 1993.

J'adresse mes remerciements à M. Jean-Dominique Polack pour avoir accepté de participer à mon jury de thèse, et plus particulièrement à MM. Jacques Jouhaneau et Xavier Meynial qui ont accepté d'examiner mon travail et d'en être les rapporteurs.

Je remercie le laboratoire d'acoustique des salles de l'IRCAM pour m'avoir permis de bénéficier de ses moyens et de son environnement scientifique, et tout spécialement Jean-Marc Jot, initiateur du projet.

Je remercie la Direction de la Recherche de RENAULT pour la confiance qu'elle m'a témoigné en me confiant ce projet, ainsi que l'Association Nationale pour la Recherche Technique, pour le soutien qu'elle lui a apporté à travers une convention CIFRE. J'ai ainsi pu mener, dans des conditions matérielles satisfaisantes, une activité de recherche scientifique dans un contexte industriel, au sein duquel j'ai pu m'épanouir. J'ai eu grand plaisir à collaborer avec Olivier Dumeaux, Alain Claes, Bernard André, Jean-Xavier Dubois, Mathieu Pauchet, et Alexandre Heidet.

Je souhaite témoigner mon amitié et une reconnaissance particulière à Éric Le Chevalier, pour nos discussions chaleureuses, et pour notre étroite collaboration au sein du département électronique de la Direction de la Recherche.

Mes pensées vont également vers M^{me} Catherine Oudin, qui a favorisé l'achèvement de cette thèse.

Mes remerciements vont aussi à Olivier Lartillot, qui m'a aidé dans le travail de métrologie dans le cadre d'un stage ; et j'exprime ma reconnaissance à M. Angelo Farina pour l'aide qu'il a pu nous apporter à cette occasion.

Je souhaite remercier tout spécialement M^{me} Suzanne Winsberg pour son aide précieuse en statistique.

Je ne saurais oublier Philippe Dérogis, chercheur en acoustique, pour ses généreux conseils en informatique, ainsi que l'équipe informatique de l'IRCAM qui assure le fonctionnement optimal de nos outils de travail... et qui supporte jour après jour nos remarques nerveuses et parfois désobligeantes.

J'exprime ma sympathie à Gérard Bertrand pour nos nombreuses discussions à propos de Fenêtre9x, et pour l'aide technique qu'il a pu m'apporter dans la préparation des séances de mesure.

Je remercie également les autres thésards et stagiaires de l'IRCAM, et plus particulièrement Laurent Cerveau pour sa sérénité, Federico Cruz pour son dynamisme, Alban Bassuet pour sa fougue... sans oublier Véronique Larcher pour sa spontanéité, son joli sourire, et nos échanges maraîchers.

Je me tourne enfin vers Olivier Warusfel, qui, à la suite de Jean-Marc Jot, a pris la responsabilité de guider mes travaux ; d'avoir tenté de m'éduquer, dans la bonne humeur, à une démarche scientifique rigoureuse, je lui serai toujours reconnaissant.

Et je termine par un grand merci à toutes les personnes qui, par leur impatience, m'ont poussé à terminer la rédaction de ce document.

Résumé

Cette thèse est dédiée à l'analyse et à l'optimisation de la reproduction d'un signal audio à l'intérieur des habitacles de voiture. Le nombre de transducteurs électroacoustiques à l'intérieur d'un habitacle de véhicule automobile s'est considérablement accru ces dernières années avec le développement de nouveaux produits tels que les téléphones mains libres ou les dispositifs à commande vocale. Mais la principale fonction audio reste sans nul doute la diffusion du signal provenant de l'autoradio. Comme la perception du bruit de roulement est de plus en plus réduite à l'intérieur des véhicules modernes, la qualité de la reproduction sonore est devenue un enjeu important pour les concepteurs de véhicules. Ce travail aborde le problème sur l'aspect de la métrologie, de la caractérisation objective, de l'évaluation perceptive et du traitement du signal appliqué à la correction de l'image spectrale et spatiale du signal sonore.

Concernant l'aspect métrologique, une nouvelle procédure de mesure de réponses impulsionnelles acoustiques est présentée. En particulier, on étudie et démontre la possibilité de réaliser une mesure asynchrone, c'est-à-dire avec un dispositif d'enregistrement découplé du dispositif de lecture ; un protocole de mesure entièrement automatique est développé.

Le travail de caractérisation objective des conditions d'écoute est mené à partir d'une base de données enregistrées dans un grand nombre de véhicules et avec plusieurs dispositifs microphoniques. Les réponses impulsionnelles sont ensuite analysées sous l'angle de leur distribution temps-fréquence en énergie. Différents indices objectifs sont étudiés afin de rendre compte de la qualité de la restitution audio dans le domaine temporel, fréquentiel et spatial.

La conduite de tests perceptifs dédiés à la comparaison des véhicules ou l'évaluation de traitements audio requiert le développement préalable d'une plate-forme de restitution en laboratoire. Afin de reproduire aussi fidèlement que possible les conditions d'écoute en véhicule, cette plate-forme est basée sur le mode de restitution binaural. Cependant, les mécanismes rentrant en jeu dans ce principe dépendent fortement de l'auditeur ; nous recherchons la meilleure stratégie pour assurer une reproduction au casque aussi proche que possible de l'écoute individuelle en véhicule. Différentes corrections des réponses impulsionnelles binaurales effectuées sur un mannequin acoustique sont étudiées et comparées pour déterminer celle qui permet la meilleure approximation des conditions d'écoute réelles. Cette solution est validée par un test d'écoute perceptif réalisé à l'intérieur d'un véhicule.

A la suite des observations effectuées lors de l'analyse objective, nous étudions différents traitements visant à optimiser la qualité de l'écoute. Notamment, l'analyse temps-fréquence des réponses impulsionnelles, laisse supposer qu'il est nécessaire de mener une égalisation spectrale séparée pour le son direct et pour l'effet d'habitacle ; cette hypothèse est confirmée par un test d'écoute perceptif réalisé en laboratoire. Pour la correction de l'image spatiale, nous étudions la possibilité de restaurer les conditions d'écoute idéales de la stéréophonie classique, en nous basant sur le principe de création de haut-parleurs virtuels. A cette occasion, nous développons une technique d'inversion de réponse impulsionnelle mettant en œuvre un principe de régularisation. Enfin, nous proposons plusieurs solutions permettant d'augmenter la largeur apparente de la source et la sensation d'enveloppement sonore.

Tous les traitements proposés ont été implantés dans un environnement logiciel fonctionnant en temps réel et permettant de tester la pertinence de la modification envisagée. Un système modulable a été développé, dont l'étage de traitements peut être connecté soit au dispositif électroacoustique réel d'un véhicule, soit à un environnement de simulation de la restitution audio embarquée. Ce système permet une comparaison instantanée entre le dispositif électroacoustique d'origine et celui modifié.

TABLE DES MATIÈRES

Table des matières	i
Introduction	1
I Caractérisation objective	7
I.1 Introduction	7
I.2 Définitions	7
I.3 Mesure de réponses impulsionnelles	9
I.3.1 Présentation des différentes techniques	9
I.3.1.1 Principe général	9
I.3.1.2 Les séquences de longueur maximale	10
I.3.1.3 Balayage fréquentiel	17
I.3.1.4 Comparaison des méthodes	24
I.3.1.5 Utilisation dans le contexte automobile	26
I.3.2 Dispositifs découplés	26
I.3.2.1 Première approche	26
I.3.2.2 Estimation du décalage des fréquences d'échantillonnage	28
I.3.2.3 Modification de la fréquence d'échantillonnage	34
I.3.2.4 Prise en compte des retards relatifs de propagation	36
I.3.2.5 Validation	40
I.4 Description de la réponse impulsionnelle	42
I.4.1 Considérations générales	42
I.4.1.1 Premières réflexions	42
I.4.1.2 Réverbération tardive	42
I.4.2 Description temps-fréquence de la réverbération	45
I.4.2.1 Analyses temps-fréquence	45
I.4.2.2 Analyses statistiques	49
I.4.2.3 Modèle stochastique temps-fréquence de la réverbération tardive	52
I.4.3 Comportement en basses fréquences	55
I.5 Analyse temps-fréquence de réponses impulsionnelles d'habitacle	57
I.5.1 Protocole de mesure	58
I.5.2 Temps de réverbération	59
I.5.3 Densité spectrale de puissance initiale	62
I.5.4 Décroissance temporelle de l'énergie	63
I.5.5 Répartition spatiale de l'énergie	65

I.5.6	Corrélation interaurale	68
I.6	Conclusion	71
II	Restitution binaurale	73
II.1	Introduction	73
II.2	Simulation binaurale	74
II.2.1	Cadre de l'étude	74
II.2.2	Formulation générale	74
II.2.3	Différentes corrections possibles	75
II.2.3.1	À partir des BIR mesurées sur nos propres oreilles	75
II.2.3.2	À partir des BIR mesurées sur un autre individu	76
II.2.4	Matériel disponible	78
II.3	Réalisation, traitements et exploitation des mesures	78
II.3.1	Réalisation des mesures	79
II.3.2	Traitements sur les mesures	79
II.3.2.1	Décomposition des filtres	80
II.3.2.2	Compensation de l'effet de l'enceinte de mesure	80
II.3.2.3	Inversion de la réponse impulsionnelle d'un casque d'écoute	82
II.3.2.4	Calcul et inversion de la HRTF en champ diffus ou pondéré	84
II.4	Optimisation du mode de reproduction	84
II.4.1	Méthodologie	84
II.4.2	Test préliminaire	85
II.4.3	Choix des méthodes d'égalisation	86
II.4.4	Comparaison objective	87
II.4.4.1	Dans une chambre anéchoïque	88
II.4.4.2	Dans l'habitacle	91
II.4.5	Modalités du test d'écoute	95
II.4.6	Protocole expérimental	98
II.4.7	Analyse statistique classique	101
II.4.7.1	Impressions générales	101
II.4.7.2	Analyse des distributions des jugements de dissemblance	101
II.4.7.3	Analyse de variance	106
II.4.7.4	Ordonnancement des stimuli	106
II.4.7.5	Conclusion	108
II.4.8	Analyse multidimensionnelle	109
II.4.8.1	Introduction	109
II.4.8.2	Analyse sur la moyenne des sujets	110
II.4.8.3	Analyse prenant en compte chaque sujet	115
II.4.8.4	Conclusion	119
II.5	Conclusion	119
III	Amélioration de l'écoute embarquée	121
III.1	Introduction	121
III.2	Correction du timbre	122
III.2.1	Cadre de l'étude	122
III.2.2	Configurations envisagées	122
III.2.3	Réglage du Spatialisateur®	124
III.2.4	Protocole expérimental	125
III.2.5	Analyse objective	126
III.2.5.1	Configurations idéales	126
III.2.5.2	Configurations réelles	127

III.2.6	Analyse statistique classique	130
III.2.6.1	Analyse de variance	130
III.2.6.2	Analyse des distributions des jugements de dissemblance	131
III.2.7	Analyse multidimensionnelle	137
III.2.8	Conclusions du test d'écoute	143
III.2.9	Implantation en véhicule	144
III.2.9.1	Correction optimisée pour le conducteur	145
III.2.9.2	Influence d'une correction optimisée pour un autre passager	148
III.3	Correction de l'image sonore	148
III.3.1	Cadre de l'étude	148
III.3.2	Alignement des sources sonores existantes	150
III.3.3	Principe de création de sources sonores virtuelles sur haut-parleurs	151
III.3.3.1	Principe des système transauraux	152
III.3.3.2	Cas d'un dispositif de restitution symétrique	154
III.3.3.3	Cas d'un dispositif de restitution non symétrique	157
III.3.3.4	Limitation de la largeur de bande	160
III.3.4	Validation en champ libre	162
III.3.4.1	Dispositif de restitution symétrique	163
III.3.4.2	Dispositif de restitution non symétrique	170
III.3.4.3	Conclusion	173
III.3.5	Validation en véhicule	176
III.3.5.1	Calcul des filtres transauraux	176
III.3.5.2	Influence de la position de l'auditeur	181
III.3.5.3	Propositions d'amélioration	184
III.4	Augmentation de la sensation d'espace	188
III.4.1	Position du problème	188
III.4.1.1	Approche classique	188
III.4.1.2	Approche originale appliquée aux petites salles	191
III.4.2	Implantation en véhicule	195
III.4.2.1	Ajout de premières réflexions latérales	195
III.4.2.2	Retard et atténuation sur les haut-parleurs arrières	195
III.4.2.3	Augmentation du niveau en basses fréquences	195
III.4.2.4	Décorrélant des signaux alimentant les haut-parleurs	196
III.4.2.5	Ajout de réverbération	200
III.5	Synthèse des différents traitements	200
III.6	Réalisation d'une plate-forme de tests	203
III.6.1	Dispositif de traitements	203
III.6.2	Dispositif de restitution	206
III.6.3	Couplage	207
Conclusion		209
A Matériel de mesure		213
A.1	Le microphone à directivité variable	213
B Méthodes d'analyse statistique		217
B.1	Analyse statistique classique	217
B.1.1	La méthode des comparaisons par paires et le modèle de Bradley-Terry	217
B.1.2	L'analyse de variance	219
B.1.3	La technique du bootstrap	220
B.2	Méthodes multidimensionnelles	223

B.2.1	L'analyse en composantes principales	223
B.2.2	L'analyse MDSCAL	225
B.2.3	L'analyse INDSCAL	226
B.2.4	L'analyse EXSCAL	230
C	Filtrage numérique	233
C.1	Filtrage linéaire des signaux à temps discret	233
C.1.1	Définitions	233
C.1.2	Équation linéaire aux différences à coefficients constants	234
C.1.3	Propriétés	234
C.1.4	La cellule d'ordre deux	237
C.1.5	Structures des filtres	238
C.1.6	Synthèse d'un filtre non récursif	239
C.1.7	Synthèse d'un filtre récursif	240
C.1.8	Lissage fréquentiel	243
C.1.9	La technique du warping	245
C.2	Transformée de Fourier à court terme	249
D	Inversion de réponses impulsionnelles	251
D.1	Inversion de réponses impulsionnelles	251
D.1.1	Cadre théorique	251
D.1.1.1	Présentation du problème	251
D.1.1.2	Inversion généralisée	257
D.1.1.3	Régularisation d'un problème mal posé	258
D.1.2	Application à l'inversion de réponses impulsionnelles acoustiques	264
D.1.2.1	Approche directe	264
D.1.2.2	Techniques de régularisation	273
D.1.2.3	Conclusion	279
E	Réverbération artificielle	283
E.1	Le Spatialisateur®	283
E.1.1	Description détaillée de l'algorithme	284
E.1.2	Normalisation	287
E.1.3	Simplifications	288
	Bibliographie	291

Introduction

LA QUALITÉ ACOUSTIQUE de l'habitacle d'un véhicule automobile suscite un intérêt croissant. L'habitacle d'un véhicule est aujourd'hui le premier environnement d'écoute de musique ou de radio. Comme pour toute situation d'écoute, les propriétés du champ sonore du lieu de reproduction influent de manière décisive sur la qualité d'écoute finale. Cependant, la nature particulière des habitacles automobiles provoque des distorsions majeures sur l'écoute. Par ailleurs, les spécifications d'un système audio embarqué requièrent un certain nombre de compromis qui sont bien différents de ceux habituellement admis dans une installation domestique, principalement la consommation en puissance, l'encombrement, la masse et la position des haut-parleurs, et surtout le coût financier. En conséquence, l'étude des phénomènes acoustiques et électroacoustiques à l'intérieur de l'habitacle est une tâche décisive pour améliorer les conditions d'écoute.

Les principales *distorsions* perceptibles de la reproduction sonore en habitacle automobile sont la modification de la réponse fréquentielle du système stéréophonique (colorations et équilibre entre zones de fréquence), et la déformation de sa réponse spatiale (déséquilibre gauche/droite et instabilité de la localisation des événements sonores). Ces effets dépendent de la géométrie et des revêtements des parois, ainsi que de l'emplacement des haut-parleurs et de l'auditeur. Par ailleurs, la présence de bruit pendant l'écoute du signal audio est un aspect spécifique, et son influence peut affecter différents paramètres de la perception auditive. La marge de manœuvre sur les paramètres de conception étant limitée par les critères de sécurité et de coût intervenant dans la conception d'un véhicule, il apparaît opportun de développer des solutions s'appuyant sur un traitement de *correction électronique* des signaux sonores.

Cette thèse se donne pour objectifs l'étude des méthodes de caractérisation de la qualité de reproduction d'un signal audio à l'intérieur d'un véhicule, et la recherche de corrections visant l'amélioration de cette qualité par l'utilisation de techniques de traitement du signal. Elle fait suite aux études déjà entreprises par la Direction de la Recherche de RENAULT sur l'aide à la conception audio en véhicule [100], et destinées à fournir au bureau d'étude des outils pour faciliter le développement d'un système audio. Car, dans l'industrie automobile actuelle, si le développement des composants des interfaces électroacoustiques est du ressort des fournisseurs, RENAULT se doit de *spécifier* ces différentes fonctions.

Il faut garder à l'esprit que les spécifications ne peuvent avoir pour objet unique le système audio ou l'environnement acoustique, mais la combinaison des deux : l'amplificateur et les haut-parleurs ont une profonde influence sur la qualité de la reproduction sonore, de même que les caractéristiques acoustiques du compartiment telles que sa géométrie, le revêtement des parois, les objets qui s'y trouvent, et surtout la position et l'orientation des haut-parleurs. La caractérisation d'un système de reproduction sonore nécessite la prise en compte des propriétés du lieu d'écoute. Cependant, un habitacle de véhicule ne peut être assimilé à une salle traditionnelle. En premier lieu, l'habitacle constitue un volume extrêmement

réduit. Bien que la réponse acoustique de l'habitacle montre un phénomène de décroissance de l'énergie, les hypothèses communément admises pour l'analyse du champ réverbéré ne sont ici plus respectées, en tous cas sur une grande plage de fréquences. Par ailleurs, les sources sont intégrées à des parois qui ont une masse et une rigidité faible, et par conduction solidienne, celles-ci entrent en vibration ; on ne peut donc pas les considérer comme des parois à réaction locale. De plus, l'ensemble des sources et des parois se trouvent au voisinage immédiat de l'auditeur. Aussi, les théories classiquement utilisées pour modéliser les propriétés d'une salle ne sont plus adaptées. Sans chercher à modéliser le comportement acoustique de l'habitacle, nous nous attacherons cependant à la caractérisation de la qualité du point d'écoute, première étape nécessaire pour envisager des méthodes d'optimisation. Cette caractérisation *objective* sera menée avec des outils classiques comme l'acquisition de réponses impulsionnelles, et l'analyse de la distribution temps-fréquence ou de la distribution spatiale de l'énergie. Dans un second temps, en nous fondant sur l'hypothèse de linéarité des phénomènes observés, et sans chercher l'optimisation de l'habitacle lui-même, nous étudierons les possibilités de correction de la qualité d'écoute en traitant le signal diffusé par le système électroacoustique.

La caractérisation objective ne peut se suffire à elle-même et l'auditeur reste le maillon final de la chaîne. La pertinence des mesures doit donc être validée par des tests d'écoute subjectifs. La nécessité de comparer de façon immédiate différents habitacles de véhicule ou différents traitements correctifs impose le développement préalable d'une plate-forme de restitution en laboratoire. Celle-ci doit permettre de restituer aussi fidèlement que possible la qualité acoustique d'un habitacle. La méthode la plus appropriée est basée sur la captation et la synthèse de réponses binaurales de véhicules intégrant, le cas échéant, les traitements correctifs. Du point de vue de l'analyse perceptive, l'effet d'habitacle que l'on peut observer s'éloigne de l'acoustique des salles par les ordres de grandeur de certains phénomènes, principalement la distribution modale et le taux de décroissance de l'énergie. Cependant, on peut supposer que les *indices perceptifs* qui vont être jugés par l'auditeur restent identiques (précision de localisation, sensation d'enveloppement, balance spectrale, etc.), même si, compte tenu du contexte particulier, ses *attentes subjectives* peuvent être différentes (nécessité accrue de la sensation d'enveloppement pour « effacer » subjectivement les limites de l'habitacle). Sans consacrer une étude spécifique à la caractérisation perceptive dans le contexte d'un habitacle, nous nous attacherons à la perception de certaines caractéristiques de la distribution temps-fréquence et aux principaux facteurs liés à l'image spatiale, tels que la précision de localisation de la source et la sensation d'enveloppement. Les facteurs cognitifs qui déterminent la qualité d'un système audio embarqué sont sans doute plus importants que ceux existant pour une installation domestique ; la qualité réelle et la qualité présupposée d'un équipement audio doivent être couplées avec l'apparence visuelle et le prestige du véhicule. Dans notre étude cependant, nous adopterons une démarche exclusivement basée sur la perception auditive, sans traiter les éventuelles interactions avec les aspects visuels et cognitifs.

La première partie de cette étude est consacrée à la caractérisation objective de la chaîne audio et de la qualité acoustique de l'habitacle. La propagation d'un signal sonore dans une salle est décrite comme une transmission à travers un système linéaire et invariant dans le temps, entièrement caractérisé par sa *réponse impulsionnelle* causale. Nous rappellerons qu'il existe plusieurs techniques permettant d'obtenir cette réponse impulsionnelle. La plus connue est l'utilisation de séquences de longueur maximale [25], qui sont des signaux déterministes dont l'autocorrélation est celle d'un signal aléatoire. Une autre technique, voisine de la spectrométrie à filtrage décalé, consiste à utiliser une sinusoïde de fréquence glissante [258] ; une nouvelle méthode a récemment fait son apparition [80], basée également sur l'utilisation d'un balayage fréquentiel, mais permettant d'obtenir la réponse impulsionnelle à l'aide d'une seule opération.

Ces techniques ont en commun qu'il est nécessaire d'effectuer une lecture et un enregistrement synchrones, c.-à-d. avec le même dispositif. Or cette contrainte est gênante dans le contexte automobile, puisqu'il est nécessaire d'accéder aux connexions électriques des haut-parleurs. Une solution plus pratique serait de stocker le signal de mesure sur un disque compact, et d'utiliser le lecteur de l'autoradio ;

l'enregistrement du signal capté par le microphone de mesure serait alors directement effectué sur le disque dur d'un ordinateur. En outre, il devient possible d'inclure dans la mesure la fonction de transfert de l'amplificateur de l'autoradio. Malheureusement, le dispositif de lecture et le dispositif d'enregistrement ont une fréquence d'échantillonnage différente. Il est donc nécessaire, d'une part, d'*estimer* la différence de fréquence d'échantillonnage entre les deux dispositifs, d'autre part, de *resynchroniser* le signal enregistré par rapport au signal émis, avant d'effectuer les traitements adéquats. Une procédure automatique sera développée à cette occasion, et nous verrons qu'il est possible de récupérer les retards de propagation relatifs lorsque plusieurs haut-parleurs sont mesurés de manière séquentielle.

Une campagne de mesures dans différents véhicules sera effectuée afin de constituer une base de données destinée simultanément à dresser un paysage de l'existant et à vérifier l'adéquation des procédures d'analyse. Cette base de données comportera des véhicules de marque RENAULT, mais aussi de marque concurrentes. Pour chaque véhicule, les mesures seront effectuées aux quatre places principales, pour chaque haut-parleur et avec trois systèmes de prise de son différents : microphone omnidirectionnel BRÜEL & KJÆR, tête artificielle HEAD ACOUSTICS et microphone SOUNDFIELD.

Nous nous intéresserons ensuite à l'analyse de ces réponses impulsionnelles acoustiques. Étant donné les différences d'ordre de grandeur de l'étalement temporel des réponses impulsionnelles mesurées dans un habitacle de véhicule par rapport à celles mesurées dans des salles de concert, les indices classiques de l'acoustique des salles sont inapplicables. Nous nous concentrerons donc sur l'*analyse temps-fréquence* de ces réponses impulsionnelles [139, 310]. Outre le temps de réverbération et le spectre initial de puissance, il est possible de caractériser fréquemment le son direct, le champ réverbéré, et le rapport entre les deux. De plus, nous présenterons plusieurs indices permettant de rendre compte de la qualité de la restitution sonore dans le domaine temporel et dans le domaine spatial.

Ces travaux de caractérisation doivent permettre de dégager des pistes concernant les traitements à mettre en œuvre pour améliorer l'écoute d'un signal audio dans les habitacles de voiture.

Une seconde partie de l'étude sera consacrée à la validation d'une procédure d'auralisation de la qualité acoustique de l'habitacle permettant de mener des tests subjectifs sur la base de stimuli auditifs reconstruits à partir de la mesure de réponses impulsionnelles binaurales dans le véhicule.

Pour des raisons pratiques, le choix du dispositif de restitution s'est porté sur un casque d'écoute : on parle de *simulation binaurale* [205]. Sous certaines conditions restrictives, le rendu acoustique peut être exact : le signal de pression arrivant aux tympans de l'individu et émis par le casque est identique à celui qui lui serait parvenu s'il s'était trouvé dans le champ acoustique réel. Malheureusement, les conditions nécessaires à cette exactitude sont très réductrices, puisqu'elles supposent de mesurer les réponses impulsionnelles binaurales de chaque individu. A partir de mesures réalisées avec un dispositif permettant d'obtenir des mesures reproductibles, en l'occurrence un mannequin acoustique, nous chercherons à optimiser le système de prise de son et de restitution en mode binaural, afin qu'il reste valide pour le plus grand nombre d'individus [178].

Afin de déterminer la méthode de reproduction respectant le plus fidèlement possible, et de manière individuelle, les conditions d'écoute naturelle en véhicule, nous mènerons une évaluation objective et perceptive des différentes solutions envisagées. Nous réaliserons un test d'écoute *in situ* qui nous permettra de comparer, d'une part, l'écoute du champ sonore *réel* de l'habitacle, et d'autre part, le champ sonore *reconstruit* à partir de mesures de réponses impulsionnelles binaurales et diffusé par un casque d'écoute ; différents traitements seront appliqués sur ces réponses impulsionnelles binaurales, dans le but d'obtenir une restitution la plus fidèle possible. Grâce à une analyse statistique multidimensionnelle des résultats du test d'écoute, nous verrons qu'il est possible de déterminer avec précision le traitement le plus adéquat.

La troisième partie de cette thèse est consacrée à la mise en place de traitements électroniques visant à améliorer la qualité de l'écoute dans un habitacle de voiture. Plusieurs aspects seront abordés :

le contenu fréquentiel, l'image stéréophonique et l'impression d'espace.

L'analyse temps-fréquence d'une réponse impulsionnelle acoustique mesurée dans un habitacle de véhicule automobile montre que le temps de réverbération est extrêmement faible, de l'ordre de 50 ms à partir de 500 Hz ; et par conséquent, au-delà de cette fréquence, le support temporel de la réponse impulsionnelle est très étroit. La question se pose donc de la stratégie à adopter pour mener l'égalisation. Il peut paraître *a priori* inutile de distinguer la réponse spectrale liée au son direct de celle qui est liée à l'habitacle. Pour s'en assurer, un test d'écoute psychoacoustique sera réalisé à partir de réponses impulsionnelles synthétiques permettant de contrôler de manière fine les contenus spectraux du son direct et de l'effet d'habitacle. L'enjeu est de savoir si, dans le contexte de réponses impulsionnelles courtes, nous sommes capables de distinguer, à spectre total constant, des variations de répartition temps-fréquence. En effet pour les salles de taille classique, il est démontré que les contributions du son direct et de la réverbération sont perceptivement indépendantes : cette indépendance est liée à la distribution temporelle de la réponse impulsionnelle, mais aussi à la distribution spatiale, le son direct étant localisé mais la réverbération diffuse. Nous montrerons que c'est bien la *distribution spatiale* qui permet, dans le cas d'une écoute en véhicule, de maintenir la distinction entre le son direct et l'effet d'habitacle, qui doivent par conséquent être égalisés séparément.

Nous nous intéresserons, dans un second temps, à la correction de l'image sonore. Étant donnée la position de l'auditeur par rapport au système électroacoustique, il est difficile d'avoir une perception correcte de l'espace sonore, en tout cas celui qu'a voulu créer l'ingénieur du son ou le technicien metteur en ondes. Il s'agit donc ici de recréer des conditions d'écoute qui permettent de percevoir fidèlement l'image sonore (aussi bien dans son ensemble que pour chacune de ses composantes), en restaurant les conditions idéales d'écoute de la stéréophonie. Pour cela, nous utiliserons la technique du PanPot binaural-transaural, et nous simulerons deux sources virtuelles à $\pm 30^\circ$. Plus généralement, nous ouvrons la voie à la possibilité de diffuser des messages autrement que de manière frontale, en donnant la possibilité d'accéder à l'ensemble de la sphère auditive ; cette demande est récurrente dans les interfaces homme-machine. La synthèse binaurale (c.-à-d. la simulation de la localisation de sources sonores en trois dimensions un utilisant les propriétés fréquentielles du codage directionnel de l'oreille humaine) permet une reproduction sonore tridimensionnelle sur deux haut-parleurs à condition que le signal binaural soit décodé au travers d'une matrice de filtrage inverse ; cette technique, dite *transaurale*, a pour effet d'*annuler* l'onde acoustique transmise de chaque haut-parleur vers l'oreille opposée [46]. La principale difficulté provient du fait que cette technique est théoriquement étudiée pour être écoutée en chambre anéchoïque. Même si la réverbération naturelle d'un habitacle automobile peut être considérée comme faible, elle n'est pas complètement absente, et la présence de réflexions prononcées nuit considérablement à la qualité de la reproduction. Il faut donc *inverser* l'acoustique de l'habitacle, ce qui a pour effet de détruire virtuellement les chemins acoustiques entre les haut-parleurs physiques et les oreilles de l'auditeur ; il devient alors possible de créer d'autres sources sonores virtuelles. L'étape inversion consiste à calculer un filtre qui, convolué avec la réponse impulsionnelle de départ, fournira un signal le plus proche possible d'une impulsion-unité. Malheureusement, nous verrons qu'il n'existe pas de solution mathématique exacte, car celle-ci exhibe des filtres non causaux à réponse impulsionnelle infinie [224] ; il faut donc avoir recours à une *approximation*. Une difficulté supplémentaire provient du fait que la position de l'auditeur n'est pas symétrique par rapport au système électroacoustique à inverser ; les schémas classiques et simplificateurs ne peuvent donc plus s'appliquer.

Enfin, nous chercherons à augmenter la sensation d'espace, pour le moins réduite dans un aussi faible volume. Nous ferons la distinction entre la sensation d'espace associée à la source sonore, encore appelée *largeur apparente de la source*, et la sensation d'espace associée à la salle, encore appelée *enveloppement sonore* [215]. Comme il est démontré dans la littérature, la largeur apparente de la source est liée à la présence de réflexions précoces latérales ; alors que l'enveloppement sonore est lié au rapport énergétique entre l'avant et l'arrière, au rapport énergétique entre le son direct et le champ réverbéré et au temps de réverbération. Plusieurs solutions seront testées pour augmenter aussi bien la largeur apparente de la source que l'enveloppement sonore.

Tous les traitements présentés sont implantés de manière logicielle dans un environnement temps réel dédié à l'audio ; comme l'auditeur demeure le seul juge, il nous a semblé important de ne pas négliger cet aspect. La *pertinence perceptive* des différents traitements peut être évaluée directement à l'intérieur d'un véhicule, par le biais de comparaisons avec le système d'origine ; mais le dispositif peut également être connecté à une plate-forme de simulation en laboratoire, ce qui permet d'effectuer des comparaisons auditives détachées de toute influence visuelle.

Page blanche

CHAPITRE I

Caractérisation objective

I.1 Introduction

La première étape de cette thèse est orientée vers la caractérisation objective de la chaîne audio et de la qualité acoustique de l'habitacle. Cette phase sera l'occasion de faire un bilan des techniques de mesure de réponse impulsionnelle au regard de certaines contraintes pratiques liées au contexte de mesure. En particulier, nous étudierons et démontrerons la possibilité de réaliser une mesure asynchrone, c'est-à-dire avec un dispositif d'enregistrement découplé du dispositif de lecture. Ce travail s'est poursuivi chez Renault par un stage qui a permis l'écriture d'une procédure de mesure automatique (détection de dérive d'horloge d'échantillonnage, conversion de fréquence d'échantillonnage, et récupération des temps de propagation relatifs entre plusieurs sources mesurées de manière séquentielle).

Un ensemble d'indices objectifs sera ensuite calculé à partir des réponses impulsionnelles mesurées dans un grand nombre de véhicules avec trois capteurs différents : microphone omnidirectionnel, mannequin acoustique et microphone à directivité variable. Les mesures effectuées avec un microphone omnidirectionnel seront analysées à l'aide d'une procédure permettant de calculer une distribution temps-fréquence. La recherche d'informations sur la distribution directionnelle de l'énergie à l'intérieur de l'habitacle a recours soit aux mesures effectuées avec la tête artificielle soit aux mesures directionnelles accessibles avec le microphone à directivité variable.

Si l'on tente de caractériser l'effet d'habitacle par un temps de réverbération associé, on observe des valeurs extrêmement faibles (de 150 à 50 ms variant peu avec la fréquence) et différant peu entre véhicules. Il faut cependant souligner que, en dessous d'une certaine fréquence, l'analyse temps-fréquence est délicate, car on ne rentre plus dans le cadre du formalisme sous-entendu : la densité fréquentielle de modes est faible (à cause du volume réduit) et, par conséquent, les isofréquences du relief de décroissance sont parfois peu assimilables à des exponentielles décroissantes.

Outre le temps de réverbération et le spectre initial de puissance, il est possible de caractériser fréquemment le son direct, le champ réverbéré et le rapport entre les deux. En d'autres termes, on peut séparer l'information de timbre entre la puissance rayonnée par la source, et la caractéristique propre de l'habitacle. Nous verrons que les spectres de l'énergie du son direct et du champ réverbéré peuvent présenter de fortes variations.

I.2 Définitions

L'ensemble des transformations subies par un signal sonore depuis son émission jusqu'à sa réception peut être vu comme une transmission d'ondes sonores dans un *canal acoustique* [96]. On suppose que

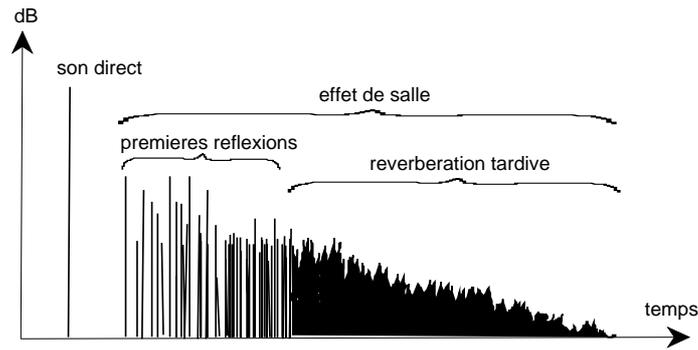


FIGURE I.1 – Echogramme schématisique d'une réponse impulsionnelle de salle

celui-ci se comporte comme un système linéaire invariant dans le temps. L'équation de convolution entre le signal émis $x(t)$, la réponse impulsionnelle de ce système $h(t)$ et le signal reçu $y(t)$ s'écrit

$$y(t) = x(t) * h(t) = \int_0^{+\infty} h(u)x(t-u) du. \quad (\text{I.1})$$

Le comportement du canal acoustique est donc complètement déterminé par sa réponse impulsionnelle causale $h(t)$. Si le système n'est pas linéaire, c.-à-d. si l'on désire prendre en compte les éventuelles distorsions harmoniques, on ne peut plus appliquer tel quel le principe de la convolution [76].

Dans un lieu clos, la présence de parois et d'obstacles entraîne l'existence de trajets multiples de durées différentes entre la source et le récepteur : c'est le phénomène de *réverbération*. Il se traduit par la superposition du *son direct* et des contributions des autres trajets : le son direct n'ayant subi aucune réflexion, les contributions des autres trajets formant l'*effet de salle*. La réponse impulsionnelle rassemble l'ensemble de ces contributions ; elle varie avec les positions de la source et du récepteur, à l'intérieur du même espace : la qualité acoustique du canal acoustique n'est pas la même en tout point de l'espace. Le canal acoustique est ainsi défini par un ensemble de paramètres que l'on regroupe en trois catégories :

- *la salle*, définie par la géométrie et les matériaux des murs et des obstacles ;
- *la source*, définie par sa position, son diagramme de rayonnement, son orientation et son étendue ;
- *le récepteur* (microphone ou oreille), défini par sa position, son diagramme de directivité et son orientation.

La figure I.1 présente l'échogramme type $h^2(t)$ d'une réponse impulsionnelle enregistrée dans une salle. On distingue les contributions du son direct et de l'effet de salle. Dans l'effet de salle, on distingue encore les *réflexions précoces*, séparées dans le temps, et un « continuum » perçu comme une traînée sonore, que l'on appelle *réverbération tardive*. Elle peut être considérée comme perceptivement caractéristique de la salle et de la puissance rayonnée par la source, indépendamment des positions et orientations de la source et du récepteur. On remarquera que l'énergie de la réverbération tardive décroît linéairement en décibels.

Prenons le cas d'une salle de concert, comme par exemple la Philharmonie de Berlin, et considérons la réponse impulsionnelle entre un haut-parleur situé sur la scène face au chef d'orchestre, et un micro placé dans la salle, face au haut-parleur, à une distance de 16 mètres (fig. I.2(a)) ; les premières réflexions sont facilement identifiables sur l'échogramme, et on peut considérer que la réverbération débute environ 100 ms après le son direct ; le temps de réverbération moyen est supérieur à 2 secondes. Les ordres de grandeur de ces quantités sont radicalement modifiés dans un habitacle automobile ; ainsi, dans une berline de moyenne gamme comme la *Laguna*, entre le haut-parleur avant-gauche et un micro omnidirectionnel placé approximativement au centre de la tête d'un conducteur (fig. I.2(b)), il est impossible d'identifier les premières réflexions, et on peut considérer que la réverbération débute entre 5 et 10 ms après le son direct (temps à partir duquel débute la décroissance exponentielle de l'énergie) ; le temps de réverbération moyen est de 50 ms environ.

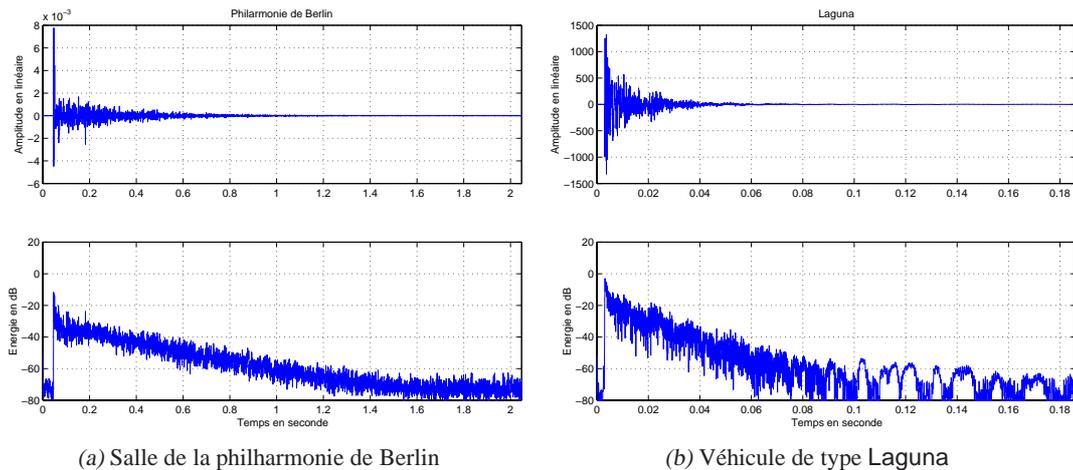


FIGURE I.2 – Echogrammes de réponses impulsionnelles de salles très différentes !

I.3 Mesure de réponses impulsionnelles

I.3.1 Présentation des différentes techniques

I.3.1.1 Principe général

La méthode la plus intuitive pour mesurer la réponse impulsionnelle est l'envoi d'une impulsion à travers le canal acoustique (on suppose celui-ci linéaire). Mais on se heurte à plusieurs difficultés techniques.

Une onde de pression strictement impulsionnelle reste impossible à générer. Quand bien même on arrive à émettre une onde de pression quasi-impulsionnelle (à l'aide d'un éclateur à gaz ou d'un pistolet), il se pose des problèmes de reproductibilité à l'identique, et surtout de distribution uniforme de l'énergie sur une large bande de fréquence. Et si on émet un signal impulsif à l'aide d'un haut-parleur, l'impulsion doit nécessairement avoir une amplitude élevée, afin de garantir un rapport signal/bruit suffisant ; se pose alors le problème des non-linéarités dans les haut-parleurs, problème exacerbé par la distribution non uniforme de l'énergie, puisque la limitation de la linéarité est imposée par la bande de fréquence où il y a le plus d'énergie.

Une autre approche consiste à exciter le système avec du bruit. Comme l'excitation est appliquée durant une longue durée, l'énergie délivrée au système pour une amplitude donnée est bien supérieure ; le problème de dynamique est donc résolu. De plus, il est facile d'assurer une uniformité de la distribution énergétique. La réponse impulsionnelle est alors déduite par intercorrélacion entre le bruit d'entrée et le signal de sortie [285]. Si le signal d'excitation $x(n)$ est un bruit blanc, la réponse impulsionnelle $h(n)$ est fournie par la fonction d'intercorrélacion $C_{xy}(n)$ entre le signal d'entrée $x(n)$ et le signal de sortie $y(n)$ du système mesuré : en effet, en régime stationnaire, on peut écrire

$$C_{xy}(n) = C_{xx}(n) * h(n), \quad (\text{I.2})$$

où $*$ représente l'opération de convolution. En pratique, l'intercorrélacion $C_{xy}(n)$ peut être estimée soit par la méthode du périodogramme moyenné, soit par celle du corrélogramme [195, 261] ; dans les deux cas, un temps d'observation relativement long est nécessaire pour que la fonction d'intercorrélacion estimée se rapproche de la réponse impulsionnelle réelle.

Une méthode astucieuse peut être mise en œuvre grâce aux séquences de longueur maximales, ou MLS (Maximum Length Sequences), dont l'autocorrélacion possède une propriété spécifique [309].

Mais la déconvolution fonctionne avec n'importe quel signal, car il suffit de connaître les signaux d'entrée et de sortie. Pour faire une mesure correcte, il suffit que le signal d'entrée ait de l'énergie à toutes

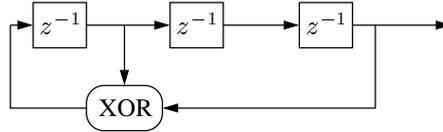


FIGURE I.3 – Registre à décalage binaire rebouclé de longueur $k = 3$ capable de générer une séquence de longueur maximale $L = 7$

les fréquences. Il est donc également possible d'utiliser comme signal d'excitation un sinus glissant. Un avantage de cette approche est que l'on peut choisir combien de temps passer à chaque fréquence, ce qui permet d'envoyer plus d'énergie aux fréquences où il y a un bruit de fond important. Cette approche s'apparente à la spectrométrie à filtrage décalé, ou TDS (Time Delay Spectrometry) [55, 259, 308].

I.3.1.2 Les séquences de longueur maximale

La solution la plus courante est l'utilisation d'un signal d'excitation composé de séquences pseudo-aléatoires [25, 146]. Les séquences pseudo-aléatoires sont des séquences de longueur maximale ; il s'agit de séquences périodiques de longueur $L = 2^k - 1$, avec k entier positif : leur longueur n'est donc pas une puissance de deux. Ces signaux sont aisément produits à l'aide d'un registre à décalage binaire rebouclé comportant k mémoires, tel que celui présenté figure I.3 [323]. Avec k mémoires, on peut générer $2^k - 1$ états différents, d'où l'appellation « séquences de longueur maximale » (on ne compte pas l'état où toutes les mémoires sont nulles, puisque le système reste figé). La séquence produite varie en fonction de la position du rebouclage (mais pas en fonction de l'initialisation des cellules mémoires, puisqu'on effectue une périodisation) ; il est possible de générer des séquences à l'aide de plusieurs rebouclages. La position du ou des rebouclages influence le comportement de la séquence, particulièrement lorsque le système n'est pas parfaitement linéaire ; Vanderkooy [309] suggère que certaines séquences sont meilleures que d'autres, et propose des positions préférentielles de rebouclages.

L'état 1 du registre à décalage est mis au niveau logique -1, et l'état 0 est mis au niveau +1 (la séquence est maintenant symétrique par rapport au niveau 0). L'avantage le plus important de ces séquences est que leur fonction d'autocorrélation *circulaire* est presque exactement une impulsion (mis à part une légère erreur continue). En d'autres termes, leur densité spectrale de puissance est uniforme (sauf pour le continu) : bien qu'une séquence de longueur maximale soit un signal *déterministe*, elle partage les propriétés d'un *bruit blanc*, d'où l'appellation « pseudo-aléatoire ».

On a l'habitude d'utiliser des séquences symétriques, c.-à-d. comportant $(L + 1)/2$ termes égaux à -1 et $(L - 1)/2$ termes égaux à $+1$; la somme des L échantillons de la séquence vaut alors -1 .

Bien que la période L des signaux ne soit pas une puissance de deux (ce qui interdit l'utilisation de l'algorithme de FFT), la corrélation circulaire peut être calculée grâce à un algorithme rapide utilisant la transformée de Hadamard sur des matrices de Sylvester [5, 6, 24, 25, 37]. L'opération d'autocorrélation est alors extrêmement simple, et ne requiert que des additions et des soustractions. Le coût de calcul est environ de $2.5L \log_2 L$, avec L la longueur de la séquence.

Considérons un signal d'entrée $x(n)$ périodique de période L . La corrélation entre le signal d'entrée et le signal de sortie $y(n)$ peut s'écrire sous la forme d'une intercorrélacion circulaire

$$C_{xy}^L(n) = \frac{1}{L} \sum_{m=0}^{L-1} x(m)y(m+n) \text{ mod } L, \quad (\text{I.3})$$

ou encore à l'aide d'un simple changement d'indice

$$C_{xy}^L(n) = \frac{1}{L} \sum_{m=0}^{L-1} x(m-n)y(m) \text{ mod } L = x(-n) \otimes y(n), \quad (\text{I.4})$$

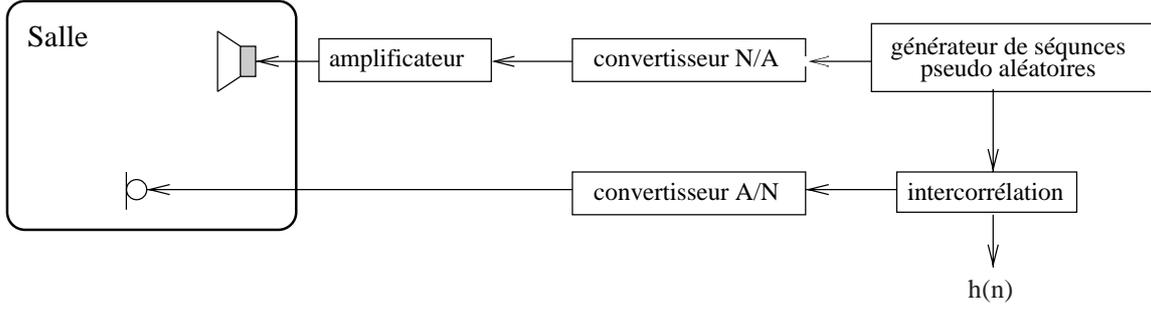


FIGURE I.4 – Principe de la mesure d'une réponse impulsionnelle avec des séquences pseudo-aléatoires

où \otimes représente l'opération de convolution circulaire.

Erreur d'offset de l'autocorrélation On peut montrer que la fonction d'autocorrélation circulaire d'une séquence de longueur maximale de période L , notée toujours C_{xx}^L , s'écrit

$$C_{xx}^L(n) = \frac{1}{L} \sum_{m=0}^{L-1} x(m-n)x(m) = \frac{L+1}{L} \delta_L(n) - \frac{1}{L}. \quad (\text{I.5})$$

On vérifie alors que la fonction d'intercorrrelation circulaire, définie par $C_{xy}^L = C_{xx}^L * h$, s'écrit

$$C_{xy}^L(n) = \frac{1}{L} \sum_{m=0}^{L-1} x(m-n)y(m) = \frac{L+1}{L} h(n) - \frac{1}{L} \sum_{m=0}^{L-1} h(m); \quad (\text{I.6})$$

où $\frac{1}{L} \sum_{m=0}^{L-1} h(m)$ représente la composante continue de la réponse impulsionnelle. D'autre part, on peut montrer que

$$\sum_{m=0}^{L-1} C_{xy}^L(m) = \frac{1}{L} \sum_{m=0}^{L-1} h(m), \quad (\text{I.7})$$

d'où la réponse impulsionnelle $h(n)$ peut être calculée exactement à partir de l'intercorrrelation circulaire $C_{xy}^L(n)$, c.-à-d.

$$h(n) = \frac{L}{L+1} \left[C_{xy}^L(n) + \sum_{m=0}^{L-1} C_{xy}^L(m) \right]. \quad (\text{I.8})$$

Ainsi, à la composante continue près, la fonction de corrélation entrée-sortie est la réponse impulsionnelle cherchée, périodisée tous les L échantillons (cf. fig. I.4). La réponse impulsionnelle peut donc être calculée exactement sur une durée finie d'une période. Comme une MLS est en fait un signal déterministe, la reproductibilité de l'opération est ainsi garantie au bruit ambiant près [25].

Mais on peut montrer que ce calcul est inutile [270]. En effet, en définissant l'intercorrrelation circulaire $C_{xy}^L(n)$ non pas sous la forme donnée à l'équation (I.3), c.-à-d. avec une division par L , mais plutôt avec une division par $L+1$, on obtient

$$\begin{aligned} C_{xy}^L(n) &= h(n) - \frac{1}{L+1} \sum_{m=0}^{L-1} h(m) \\ &= h(n) - \frac{1}{L} \sum_{m=0}^{L-1} h(m) + \frac{1}{L(L+1)} \sum_{m=0}^{L-1} h(m). \end{aligned} \quad (\text{I.9})$$

Le deuxième terme de la dernière équation représente la composante continue de la réponse impulsionnelle. De même pour le troisième terme, mais elle est atténuée d'une valeur $L+1$. Le deuxième terme

retire donc la composante continue du premier terme, et le troisième ajoute une très faible valeur (ceci d'autant plus que la plupart des systèmes audio ont une réponse sans composante continue). Il n'est donc pas nécessaire de procéder au calcul de l'équation (I.8). De plus, la composante continue peut être décalée par de la distorsion d'ordre paire [309].

Une autre façon de procéder peut être de considérer une MLS ainsi décalée [270]

$$x_0(n) = x(n) - 1, \quad (\text{I.10})$$

et de calculer le signal de sortie en supposant le signal d'entrée égal à $x_0(n)$, c.-à-d.

$$\begin{aligned} y(n) &= x_0(n) \otimes h(n) \\ &= x(n) \otimes h(n) - \sum_{m=0}^{L-1} h(n-m) \end{aligned} \quad (\text{I.11})$$

Si l'on calcule la réponse impulsionnelle à l'aide de l'équation (I.4), on obtient, en notant $\sum_{m=0}^{L-1} h(m) = H(0)$ la TFD de $h(m)$,

$$\begin{aligned} C_{xy}^L(n) &= x(-n) \otimes (x(n) \otimes h(n) - H(0)) \\ &= (x(-n) \otimes x(n)) \otimes h(n) - x(-n) \otimes H(0) \\ &= \sum_{m=0}^{L-1} C_{xx}^L(n)h(n-m) - \frac{H(0)}{L+1} \sum_{m=0}^{L-1} x(m) \end{aligned} \quad (\text{I.12})$$

Puisque les MLS originales sont symétriques, la sommation de $x(m)$ sur L échantillons vaut toujours -1 ; et le premier terme est transformé grâce à l'équation (I.6), l'intercorrélant étant définie avec une normalisation par $L+1$ et non par L . On obtient alors

$$\begin{aligned} C_{xy}^L(n) &= h(n) - \frac{1}{L+1} \sum_{m=0}^{L-1} h(n) + \frac{H(0)}{L+1} \\ &= h(n) - \frac{H(0)}{L+1} + \frac{H(0)}{L+1} \\ &= h(n) \end{aligned} \quad (\text{I.13})$$

c.-à-d. exactement la réponse impulsionnelle cherchée. Donc en ajoutant une composante continue de -1 à la MLS émise, on peut retrouver directement la réponse impulsionnelle avec la bonne composante continue.

Longueur de la séquence et temps de mesure Pratiquement, lorsqu'on injecte une MLS dans un système linéaire, le signal de sortie résulte d'une convolution avec le signal d'entrée. Or la propriété désirable de l'autocorrélation des MLS ne voit le jour que sous couvert d'une opération de convolution circulaire. Le principe consiste donc à émettre au moins deux périodes de la MLS dans le système à mesurer, et à ne pas enregistrer la première. D'autre part, il faut que la longueur de la MLS soit toujours supérieure à la longueur de la réponse impulsionnelle à mesurer; dans le cas contraire les termes supplémentaires se superposent aux premiers puisque la MLS est périodique. Si la réponse impulsionnelle est de longueur infinie, il faut alors s'assurer que ses échantillons au-delà de la longueur de la MLS soient de faible amplitude.

Plaçons-nous dans le cas de la mesure d'une réponse impulsionnelle de salle. Il est nécessaire que la séquence entière, supposée de longueur suffisante, ait totalement excité la salle, et qu'un état stable soit atteint; quand on ne prend pas cette précaution, on effectue le calcul de l'intercorrélant sur un enregistrement qui ne contient pas toute l'information, d'où la présence d'un bruit de fond nettement supérieur (cf. fig. I.5¹). Et comme la longueur de la séquence doit être supérieure au temps de réverbéra-

1. Ces mesures ont été effectuées en introduisant une MLS (stockée sous forme de fichier informatique) dans le Spatialisateur®, et en enregistrant la sortie (également sous forme de fichier informatique) sur le même ordinateur; l'intercorrélant circulaire entre les deux fichiers a été calculée en temps différé sous Matlab®. La dynamique maximale escomptée est donc bien supérieure à celle obtenue avec une chaîne de mesure traditionnelle.

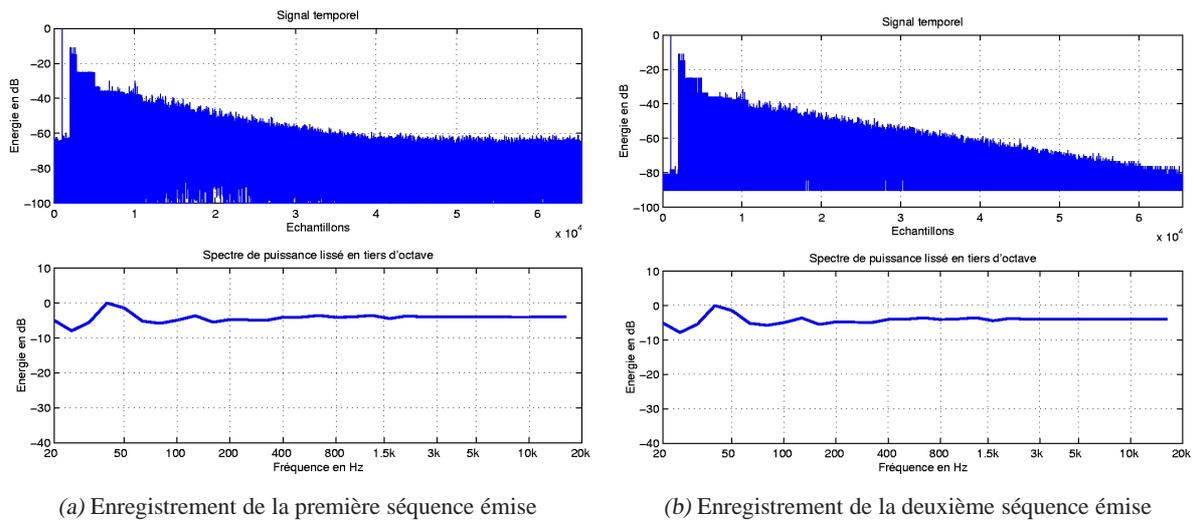


FIGURE I.5 – Illustration de la nécessité d'attendre que l'état stable du système à mesurer soit atteint avant de calculer l'intercorrélacion

tion, l'information contenue dans le début de la séquence a disparu. Il n'y a donc d'autre possibilité que d'émettre au moins deux fois la séquence, et de ne pas enregistrer la première.

Amélioration de la dynamique Pour améliorer le rapport signal sur bruit de la mesure, il est conseillé d'émettre un grand nombre de fois la séquence, et de les moyenner avant de calculer l'intercorrélacion [25, 270] (à condition que le système à mesurer et la fréquence d'horloge soient stables dans le temps). L'idée est que la réponse désirée s'additionne de manière cohérente, alors que le bruit s'additionne de manière incohérente : l'amélioration du rapport signal/bruit est de l'ordre de \sqrt{N} , où N est le nombre de périodes moyennées. À l'origine, ce principe était utilisé pour la mesure de haut-parleurs à l'aide de véritables impulsions [19]. L'inconvénient de ce principe est que le temps nécessaire pour réaliser la mesure est plus long.

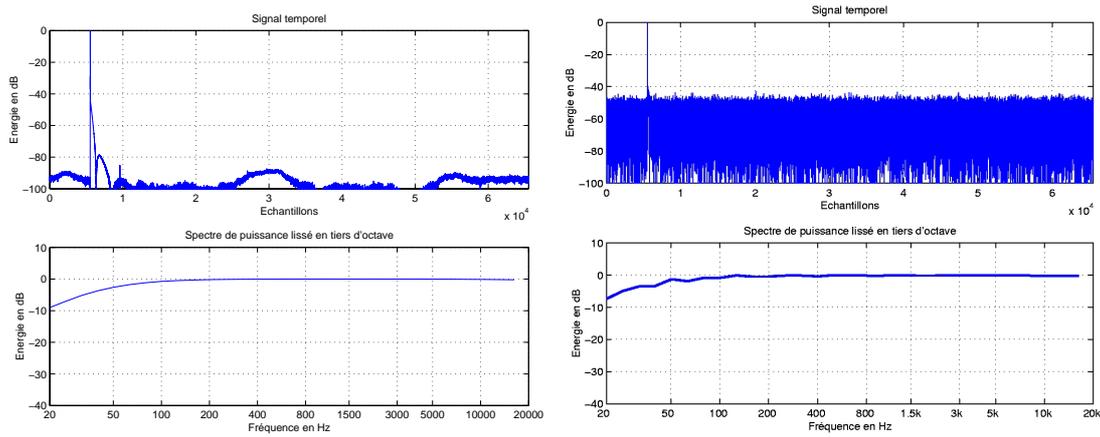
Immunité au bruit Par la loi des grands nombres, on peut montrer que la densité de probabilité du signal reçu $y(n)$, c.-à-d. une MLS filtrée par le canal acoustique, est une gaussienne de moyenne nulle [270].

BRUIT STATIONNAIRE

La salle dont on mesure le canal acoustique n'est pas en général totalement silencieuse ; elle présente en général un bruit de fréquence faible correspondant aux bruits extérieurs filtrés par les parois de la salle. Ceci est encore plus vrai dans le cas d'un véhicule au repos, puisqu'il n'est pas toujours possible de réaliser les mesures en plaçant le véhicule dans une chambre anéchoïque. Pour faire face à cette imperfection, on peut préaccentuer les fréquences basses des MLS émises, puis égaliser de nouveau le signal à la sortie [211, 270].

On peut ainsi filtrer les MLS émises par un filtre de bruit rose. Non seulement l'immunité aux bruits est améliorée en basses fréquences, mais en plus le signal d'excitation présente une densité de probabilité quasi gaussienne dont le spectre en tiers d'octave est idéal pour l'évaluation subjective d'un système audio.

Il est également bon d'atténuer les fréquences élevées du signal d'excitation émis par les haut-parleurs afin d'éviter une surchauffe (parfois fatale) des tweeters [210]. La préaccentuation peut être



(a) Réponse impulsionnelle calculée à l'aide de MLS en reliant la sortie du convertisseur N/A à l'entrée du convertisseur A/N (sans passer par un canal acoustique)

(b) Idem, en rajoutant un bruit blanc en sortie du convertisseur A/N (rapport signal/bruit réglé à 10 dB seulement)

FIGURE I.6 – Immunité au bruit ambiant lors de la mesure de réponses impulsionnelles par MLS

effectuée à l'aide de la transformée de Hadamard inverse. Cette technique peut aussi être utilisée comme méthode de compensation des atténuations fréquentielles dues aux distorsions de certains composants du canal acoustique.

Mais l'avantage de la technique de mesure de réponses impulsionnelles à l'aide de séquences de longueur maximale est qu'elle est relativement insensible au bruit ambiant, pourvu que celui-ci soit stationnaire et non corrélé (cf. fig. I.6). La dynamique maximale de la réponse impulsionnelle calculée avec cette technique est théoriquement de 90 dB (c.-à-d. la dynamique d'un signal numérique stocké sur 16 bits avec un bit de signe). Pratiquement, la dynamique utile n'excède pas 70 dB au mieux. En effet, le bruit ambiant n'est jamais négligeable, et le niveau sonore d'émission des MLS ne peut jamais être au moins 90 dB supérieur au niveau de bruit ; de plus, la mesure par MLS est assez sensible aux variations locales du champ sonore dans la salle, qui n'est jamais véritablement stable, et les variations thermiques jouent aussi un rôle perturbateur.

BRUIT TRANSITOIRE

Les bruits transitoires sont bien maîtrisés par cette méthode de mesure [270]. Supposons un bruit transitoire $x_t(n)$ en entrée, on obtient en sortie, d'après l'équation (I.4),

$$r(n) = x_t(n) \otimes x(-n) \tag{I.14}$$

que l'on peut considérer comme la sortie du filtre de réponse impulsionnelle $x_t(n)$ attaqué par $x(-n)$. Si le bruit transitoire en question est une impulsion $x_t(n) = \delta(n)$, alors on obtient en sortie

$$r(n) = \frac{1}{L+1} x(-n); \tag{I.15}$$

tout bruit transitoire (clic) sera donc disséminé en sortie sur toute la réponse impulsionnelle calculée.

Non-linéarités L'effet de faibles non-linéarités dans une mesure MLS peut être modélisé par un signal à phase aléatoire, représentant la somme de tous les ordres des produits d'intermodulation présents

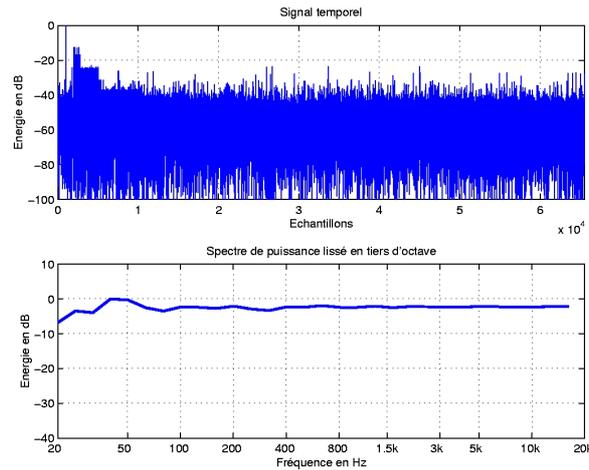


FIGURE I.7 – Influence d'un niveau d'émission trop élevé : saturation de la chaîne de mesure et diminution drastique de la dynamique de la réponse impulsionnelle

à la sortie du canal [270, 309]. On obtient de ce fait un signal aléatoire stationnaire. Son énergie est donc distribuée uniformément sur toute la période de la réponse impulsionnelle calculée. Comme le canal acoustique est un système transitoire, une partie très importante de l'énergie de sa réponse impulsionnelle est concentrée au début de la réponse calculée. On peut ainsi séparer la réponse du canal des non-linéarités en ne considérant que les premiers échantillons de la réponse.

GÉNÉRALITÉS

En analysant la présence de non-linéarités dans le canal acoustique [65, 151, 268, 309], on peut montrer que les distorsions d'ordre faible sont à l'origine de la présence de petites impulsions irrégulières observables surtout dans la partie stabilisée de la réponse impulsionnelle, c.-à-d. dans le bruit de fond. Ces irrégularités ne ressemblent donc pas à du bruit. Pour une MLS et un canal acoustique donnés, la disposition de ces impulsions est fixe. Ce motif peut donc être considéré comme une « empreinte digitale » de la MLS attaquant ce canal. On peut atténuer ces irrégularités à l'aide d'une moyenne de différentes réponses impulsionnelles obtenues par des MLS différentes.

Des non-linéarités peuvent être apportées par des saturations au niveau des transducteurs, ou par la conversion numérique \rightarrow analogique lors de la génération du signal. L'influence sur la dynamique de la réponse calculée peut même être catastrophique (cf. fig. I.7).

Pour éliminer l'influence de ces impulsions, il suffit de considérer une réponse impulsionnelle la plus longue possible, par l'utilisation d'une MLS de grande période. La réponse impulsionnelle se stabilisant rapidement, on peut tronquer cette réponse et ne garder que les premiers échantillons, là où la réponse n'a pas atteint le niveau de bruit de fond. Mais on peut également montrer qu'il faut trouver un niveau optimal du signal d'excitation. Si l'amplitude est trop grande, on observe ces distorsions ; et si l'amplitude est trop faible, le bruit de fond entache la mesure de la réponse impulsionnelle. Il ne faut donc pas hésiter à augmenter le nombre de périodes de MLS émises, la mesure moyenne ayant tendance à diminuer le niveau de ce bruit centré.

LES SÉQUENCES INVERSEMENT RÉPÉTÉES

Une amélioration de la méthode de mesure par MLS est l'utilisation de séquence inversement répétées, ou IRS (Inverse Repeated Sequences) [65], définies par

$$x(n) = \begin{cases} x(n) & \text{pour } n \text{ pair} \\ -x(n) & \text{pour } n \text{ impair} \end{cases}, \text{ avec } 0 \leq n < L, \quad (\text{I.16})$$

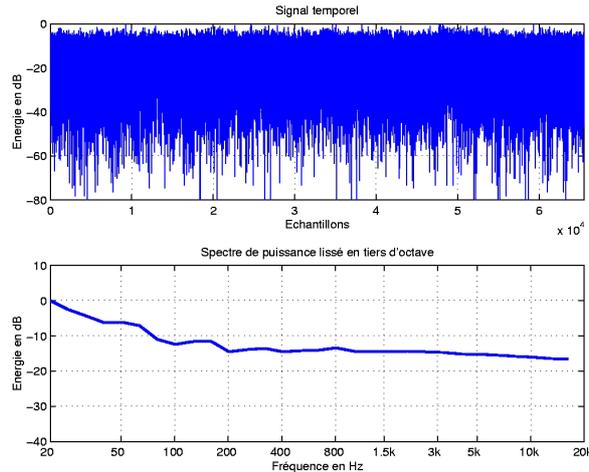


FIGURE I.8 – Réponse impulsionnelle calculée avec un appareil numérique pour la lecture, et un autre pour l'enregistrement, par la méthode des MLS; l'écart relatif entre les deux fréquences d'horloge n'est pourtant que de 2%

où le signal $x(n)$ est une MLS de période L . L'autocorrélation de ce signal est donnée par

$$C_{xx}(n) = \delta(n) - \frac{(-1)^n}{L+1} - \delta(n-L), \text{ avec } 0 \leq n < L. \quad (\text{I.17})$$

En calculant la réponse impulsionnelle par la même méthode que pour les MLS, on obtient en fait deux fois la réponse impulsionnelle : une positive et une négative. L'intérêt de cette méthode est qu'elle présente une très importante immunité face à la distorsion. En effet, elle élimine totalement les non-linéarités d'ordre pair. En revanche, elle est plus complexe à mettre en œuvre, et son défaut majeur est que l'on ne peut plus utiliser la transformée rapide de Hadamard.

Synchronisation des dispositifs d'émission et de réception Il est absolument nécessaire que les signaux d'émission et de sortie soient *parfaitement* synchronisés, c.-à-d. alignés sur le *même* signal d'horloge. En d'autres termes, cela veut dire que l'émission et l'enregistrement doivent être effectués par la même carte électronique. Un décalage d'horloge extrêmement faible suffit à diminuer le rapport signal/bruit de la réponse calculée dans des proportions très importantes, voire à rendre la mesure totalement inexploitable (cf. fig. I.8) ! Ce défaut a été étudié par Vanderkooy [309] qui montre qu'une variation temporelle du canal acoustique ou une gigue de l'échantillonnage va effectivement induire un bruit uniformément réparti sur la réponse impulsionnelle. Plus précisément, Svensson [303] met en évidence, par modélisation, par simulation et par mesures réelles, les effets d'un ralentissement, d'une accélération, voire d'une oscillation de l'échelle temporelle de la réponse du canal acoustique. Concrètement, on peut discerner deux type de variations temporelles.

Les variations temporelles à long terme sont les variations observables d'une période à l'autre de la MLS, voire à une plus grande échelle ; c'est donc une évolution lente de l'échelle temporelle. Dans le cas qui nous intéresse, c.-à-d. la non-synchronisation des horloges des deux convertisseurs, on peut ainsi prévoir une variation lente du rapport entre les deux fréquences d'échantillonnage.

Les variations temporelles à court terme sont les variations à l'échelle d'une seule période de MLS. Cela peut donc inclure la gigue de l'échantillonnage, c.-à-d. une régularité non parfaite des instants d'échantillonnage ou d'interpolation. Ce phénomène peut également se présenter en cas de non-synchronisation des deux dispositifs.

Dans les deux cas, on observe un bruit particulier qui croît de 6 dB par octave, lorsque l'on considère la différence entre le spectre idéal et le spectre réel. Ce bruit peut être atténué en moyennant plusieurs mesures. Est également présente une composante de bruit à 12 dB par octave entre le spectre des systèmes

variant et non variant ; cette deuxième composante ne peut pas être atténuée en moyennant plusieurs mesures.

Conclusion La mesure de réponses impulsionnelles acoustiques à l'aide de séquences de longueur maximale est une technique efficace qui permet d'atteindre un rapport signal/bruit confortable, et ce de manière rapide et reproductible.

De nombreuses réalisations commerciales ont vu le jour, utilisant soit une carte électronique spécialement dédiée (par exemple le logiciel MLSSA [269, 270]), soit une carte électronique dévolue au multimédia (par exemple le logiciel AURORA [77, 80]).

I.3.1.3 Balayage fréquentiel

Cette approche plus ancienne [123] propose d'envisager non pas la simple réponse impulsionnelle du système, mais l'évolution temporelle de la réponse fréquentielle. On dispose ainsi d'une mesure des différentes réflexions et de leurs caractéristiques spectrales. Dans la spectrométrie à filtrage décalé, un haut-parleur irradie le canal acoustique d'une sinusoïde dont la fréquence varie linéairement par rapport au temps. Le signal capté par le microphone attaque un filtre variable passe-bande de bande passante très étroite. La bande passante évolue également linéairement au cours du temps, à la même vitesse que le signal d'excitation. Ainsi, le filtre ne sélectionnera que les signaux d'un retard fixé, en fonction du retard entre le signal d'excitation et le filtre.

Le filtrage de sélection s'effectue simplement en multipliant le signal reçu par un signal similaire à celui émis, mais retardé d'un délai correspondant au délai à mesurer pour la partie réelle de la réponse fréquentielle, et du même délai en quadrature de phase pour la partie imaginaire. Le résultat est alors filtré par un filtre passe-bas pour ne garder que la composante continue [308].

Cette méthode présente quelques difficultés à mettre en œuvre. Le calcul précis de la bande passante la plus étroite du canal acoustique nécessite une connaissance parfaite de la fonction de transfert. Inversement, la fonction de transfert ne peut être connue exactement que si la bande passante la plus étroite du canal acoustique est mesurée de manière précise. De plus, cette bande passante peut en fait correspondre à un sous-système de réponse en amplitude très faible.

Distribution de Wigner Pour plus de clarté dans la compréhension des phénomènes, introduisons la transformée de Wigner [39–41, 44, 86, 87, 126, 193]. Toute fonction du temps $x(t)$ qui possède une transformée de Fourier $X(f)$ peut aussi être décrite comme une fonction à la fois du temps et de la fréquence, définie par

$$W_x(t, f) = \int_{-\infty}^{+\infty} x\left(t + \frac{\tau}{2}\right) x^*\left(t - \frac{\tau}{2}\right) e^{-2j\pi f\tau} d\tau, \quad (\text{I.18a})$$

$$W_X(t, f) = \int_{-\infty}^{+\infty} X\left(f + \frac{\nu}{2}\right) X^*\left(f - \frac{\nu}{2}\right) e^{2j\pi f\nu} d\nu. \quad (\text{I.18b})$$

La réponse impulsionnelle d'un système linéaire et invariant dans le temps est la description fondamentale d'un système dans le domaine temporel. La transformée de Fourier de la réponse impulsionnelle, connue sous le nom de fonction de transfert, contient la réponse en module et en phase du système pour une entrée purement sinuïdale. La distribution de Wigner de la réponse impulsionnelle est une description fondamentale du système à la fois en temps et en fréquence.

La distribution de Wigner du signal impulsion-unité $\delta(t - t_0)$ est donnée par

$$W_\delta(t, f) = \delta(t - t_0), \quad (\text{I.19})$$

elle est représentée dans le plan temps-fréquence par une droite parallèle à l'axe fréquentiel passant par le point $t = t_0$.

La distribution de Wigner d'une exponentielle complexe $e^{2j\pi f_0 t}$ est donnée par

$$W_e(t, f) = \delta(f - f_0), \quad (\text{I.20})$$

elle est représentée dans le plan temps fréquence par une droite parallèle à l'axe temporel passant par le point $f = f_0$.

Le signal modulé linéairement en fréquence, et utilisé pour la mesure de la réponse d'un système, est une fonction dont la fréquence instantanée varie linéairement avec le temps, c.-à-d.

$$c(t) = e^{j\pi at^2} \quad \text{avec } f_i(t) = \frac{1}{2\pi} \frac{d}{dt}(\pi at^2) = at; \quad (\text{I.21})$$

la distribution de Wigner de ce signal est donnée par

$$W_c(t, f) = \delta(f - at), \quad (\text{I.22})$$

elle est représentée dans le plan temps-fréquence par une droite faisant un angle $\theta = \arctan a$ avec l'axe temporel.

Transformation des coordonnées Puisqu'un signal dont la fréquence varie linéairement avec le temps possède une distribution de Wigner qui relie linéairement le temps et la fréquence, ce signal peut être utilisé pour réaliser une transformation linéaire des coordonnées de la distribution de Wigner. La théorie de la transformation des coordonnées de la distribution de Wigner est identique à celle de la transformation des coordonnées de la fonction d'ambiguïté [249].

Par exemple, la distribution de Wigner du signal $p(t) = x(t)e^{j\pi at^2}$ est, grâce à la propriété de la multiplication de la distribution de Wigner,

$$W_p(t, f) = W_x(t, f) \overset{f}{*} \delta(f - at) = W_x(t, f - at), \quad (\text{I.23})$$

où $\overset{f}{*}$ représente l'opération de convolution par rapport à la fréquence, et $W_x(t, f)$ la distribution de Wigner de $x(t)$. Cela conduit à une transformation du plan qui peut être représentée sous forme matricielle par

$$P(a) = \begin{pmatrix} 1 & 0 \\ a & 1 \end{pmatrix}. \quad (\text{I.24})$$

De manière similaire, la distribution de Wigner du signal $q(t) = x(t) * e^{j\pi t^2/b}$ est donnée par

$$W_q(t, f) = W_x(t, f) \overset{t}{*} \delta(f - t/b) = W_x(t - bf, f); \quad (\text{I.25})$$

la transformation correspondante du plan peut être représentée sous forme matricielle par

$$Q(b) = \begin{pmatrix} 1 & b \\ 0 & 1 \end{pmatrix}. \quad (\text{I.26})$$

Donc en multipliant ou en convoluant une fonction avec un signal à modulation de fréquence linéaire, il est possible de réaliser une transformation des coordonnées de sa distribution de Wigner.

Mesure d'un système par balayage fréquentiel La mesure d'un système linéaire à l'aide d'un signal à balayage de fréquence linéaire permet de déterminer sa réponse en fréquence et en phase. En pratique, le signal introduit dans le système est réel et de durée finie. Le module et la phase du signal de sortie décrivent la réponse en module et en phase du système étudié, relativement au signal d'entrée : la réponse en module et en phase sont générés comme une fonction linéaire du temps. On peut considérer que cette technique a effectué une transformée de Fourier de la réponse impulsionnelle du système, et a donc inversé les rôles du temps et de la fréquence.

La théorie de la transformation des coordonnées de la distribution de Wigner $W_h(t, f)$ de la réponse impulsionnelle $h(t)$ du système, résumée précédemment, permet de fournir une description de cette technique de mesure. Cette description est basée sur le fait que la distribution de Wigner de la transformée de Fourier $X(f)$ d'une fonction $x(t)$, exprimée comme une fonction du temps, peut s'écrire (cf. éq. (I.18a) et (I.18b))

$$W_X(t, f) = W_x(-f, t); \quad (\text{I.27})$$

la matrice de la transformation correspondante du plan est donnée par

$$L = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}, \quad (\text{I.28})$$

elle correspond à une rotation des coordonnées de la distribution de Wigner de $x(t)$ de 90° (dans le sens horaire). Et on peut montrer que si f correspond à la réponse impulsionnelle h d'un système, alors la transformée de Fourier de $h(t)$ s'écrit

$$H(f) = B \left[\left(h(t) * e^{j\pi t^2} \right) e^{-j\pi t^2} \right] * e^{j\pi t^2}, \quad (\text{I.29})$$

avec B une constante complexe arbitraire. La première opération est une convolution, elle décrit la sortie d'un système dont le signal d'entrée est ici un signal modulé en fréquence ; le signal de sortie doit être multiplié puis convolué par deux autres signaux modulés en fréquence afin de reproduire la fonction de transfert exacte $H(f)$. La mesure par balayage fréquentiel linéaire peut donc être décrite par une rotation de 90° de la distribution de Wigner de la réponse impulsionnelle du système.

En général, la pente de la modulation de fréquence linéaire du signal d'entrée vaut une valeur fixée k exprimée en Hz par seconde, donc homogène à des sec^{-2} ou Hz^2 . De ce fait, les deux autres opérateurs de modulation de fréquence linéaire doivent avoir la même valeur de la pente de la modulation de manière à maintenir la rotation de 90° . La distribution de Wigner $W_h(t, f)$ de la réponse impulsionnelle subit alors une rotation de 90° ainsi que deux facteurs d'échelle, respectivement $1/k$ et k dans les nouvelles coordonnées t et f , soit

$$W_{h_L}(t, f) = W_h \left(\frac{f}{k}, -kt \right); \quad (\text{I.30})$$

la matrice de la transformation correspondante du plan est alors donnée par

$$L = \begin{pmatrix} 0 & \frac{1}{k} \\ -k & 0 \end{pmatrix}. \quad (\text{I.31})$$

On peut alors montrer [259] que

$$H(kt) = \sqrt{k} e^{-j\pi/4} \left[\left(h(t) * e^{j\pi kt^2} \right) e^{-j\pi kt^2} \right] * e^{j\pi kt^2}; \quad (\text{I.32})$$

il faut bien sûr appliquer la transformée de Fourier inverse pour obtenir la réponse impulsionnelle.

Techniques approximatives Historiquement, la mesure par balayage fréquentiel linéaire ne mettait en œuvre que la première ou/et la deuxième opération décrite dans l'équation (I.32).

Les premiers systèmes ne comportaient qu'un seul balayage en entrée, et l'amplitude du signal de sortie était déterminée par filtrage, ou par calcul de la valeur RMS. Il était donc nécessaire de mettre en œuvre des techniques non linéaires sur le signal de sortie pour déterminer approximativement l'amplitude en fonction du temps.

Une amélioration de cette première approche a consisté en l'utilisation de détecteurs d'amplitude hétérodynes (ou synchrones) : la fréquence de sortie était multipliée par un autre signal sinusoïdal de même fréquence, et le produit des deux signaux était filtré par un filtre passe-bas. Comme le signal d'entrée était un balayage fréquentiel, le dispositif hétérodyne de multiplication utilisait également un balayage fréquentiel, identique au premier. Pour un balayage linéaire, cela veut dire que le signal de sortie était multiplié par un signal à modulation linéaire de fréquence ; cette procédure est décrite par la deuxième opération de l'équation (I.32). Afin d'obtenir une mesure précise, la vitesse de balayage devait être petite devant la bande passante du système étudié, c.-à-d. $k\pi \ll \mathcal{B}^2$, avec \mathcal{B} la bande passante du système étudié. Cette valeur est en fait difficile à estimer, puisque d'une part la fonction de transfert exacte est requise pour déterminer \mathcal{B} , et d'autre part \mathcal{B} est indispensable pour calculer la fonction de transfert. Des méthodes heuristiques devaient donc être employées dans le but d'assurer des résultats fiables.

Technique exacte La troisième opération de modulation linéaire de fréquence de l'équation (I.32) complète le processus, et effectue une rotation de 90° ainsi qu'une dilatation de la distribution de Wigner de la réponse impulsionnelle. Cette dernière opération retire le facteur de modulation de fréquence linéaire de la réponse impulsionnelle, puisque $e^{-j\pi kt^2} * e^{j\pi kt^2} = \delta(t)$. Grâce à cette amélioration, la vitesse de balayage n'est contrainte à aucune limitation, et la transformée de Fourier exacte de la réponse impulsionnelle est obtenue comme une fonction du temps.

On peut vérifier que la procédure de mesure par balayage fréquentiel linéaire permet de séparer les différentes composantes de la réponse impulsionnelle d'un système. Pour cela, le deuxième et le troisième opérateur de modulation de fréquence doivent être retardés respectivement par τ et $-\tau$, c.-à-d.

$$\begin{aligned} h_L(t) &= \left[\left(h(t) * e^{j\pi kt^2} \right) e^{-j\pi k(t-\tau)^2} \right] * e^{j\pi k(t+\tau)^2} \\ &= \sum_{i=1}^N H_i(kt) e^{-j\pi k(t_i-\tau)} \end{aligned} \quad (I.33)$$

Pour $\tau = t_i$, la τ -ième réponse du sous-système est ramenée autour de 0 Hz. Donc si la fonction est filtrée par un filtre passe-bas, les autres sous-systèmes qui ne sont pas autour de 0 Hz sont atténués, et la τ -ième fonction de transfert est sélectionnée. La procédure complète est donnée par

$$\begin{aligned} H(kt) e^{2j\pi kt\tau} * w(t) &= C \left[\left(h(t) * e^{j\pi kt^2} \right) e^{-j\pi k(t-\tau)^2} \right] * e^{j\pi k(t+\tau)^2} * w(t) \\ &\approx H_\tau(kt) \end{aligned} \quad (I.34)$$

avec $w(t)$ la réponse impulsionnelle du filtre passe-bas.

Il est donc démontré que la mesure par balayage fréquentiel linéaire permet de séparer les différentes composantes d'une réponse impulsionnelle grâce à un filtrage passe-bas, à condition que ces composantes ne se recouvrent pas. Cette procédure est une extension de la spectrométrie à filtrage décalé [123]. Elle élimine toute contrainte sur la vitesse de balayage d'un signal d'entrée de durée très longue constitué d'une modulation linéaire en fréquence.

Interprétation graphique Pour une compréhension plus aisée de l'influence des trois opérateurs de modulation de fréquence, et notamment le mécanisme de génération de la transformée de Fourier de

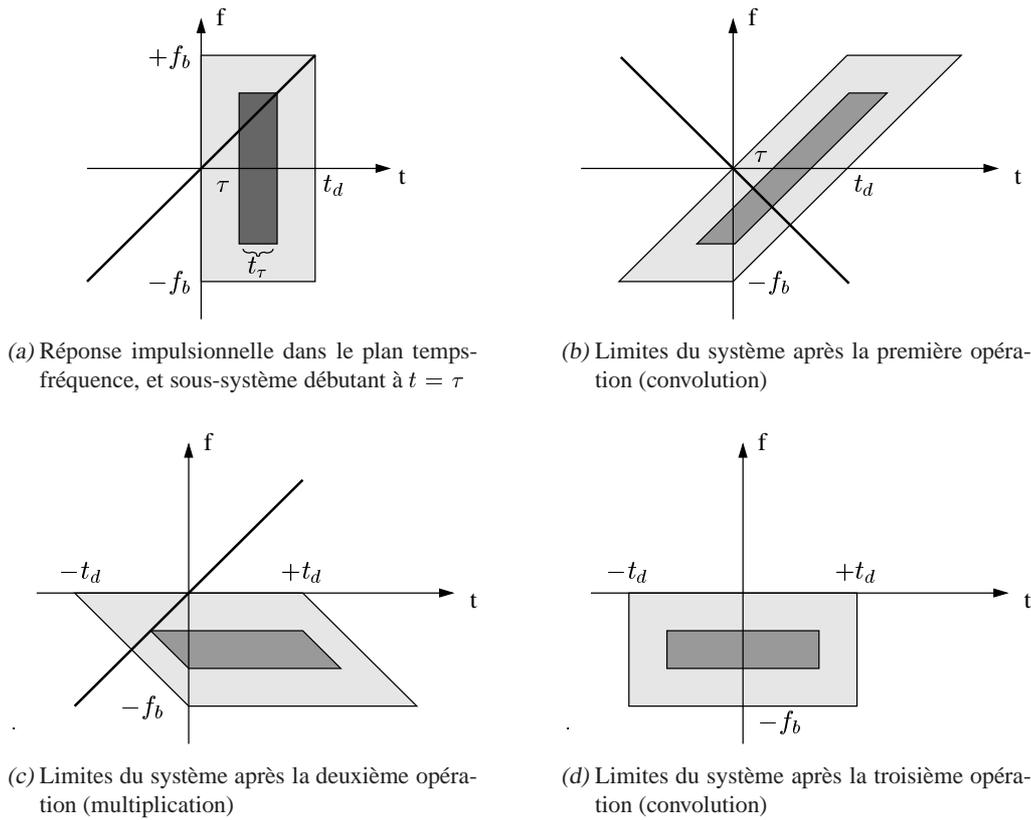


FIGURE I.9 – Représentation graphique des opérations effectuées lors de la mesure d'un système à l'aide d'un balayage fréquentiel linéaire

la réponse impulsionnelle, on peut représenter l'ensemble des opérations dans le plan temps-fréquence. Cette description montre comment les opérateurs retardés produisent une modulation de la fonction de transfert.

Considérons une réponse impulsionnelle $h(t)$ d'un système présentant des réflexions multiples. La durée totale t_d de la réponse est définie comme le temps après lequel l'énergie de la réponse n'est plus significative. De même, la largeur de bande totale f_b est définie comme la bande hors de laquelle l'énergie n'est plus significative. Ces limites sont arbitraires, et de plus elles sont mutuellement exclusives, puisqu'un signal à temps limité ne peut être à bande limité, et *vice versa*. Cependant, on définit les limites approximatives du système en temps et en fréquence, c.-à-d. $t \in [0; t_d]$ et $f \in [-f_b; f_b]$: comme représenté sur la figure I.9(a), cela définit dans le plan temps-fréquence un rectangle contenant la majeure partie de l'énergie du système.

Supposons aussi que le système complet contienne un sous-système qui débute à $t = \tau$.

La première opération est l'injection dans le système du signal modulé en fréquence (de fréquence instantanée kt), ce qui correspond à une opération de convolution temporelle entre les distributions de Wigner de la réponse impulsionnelle du système et du signal modulé en fréquence (cf. éq. (I.20) et (I.25)). Pour chaque fréquence f , la distribution de Wigner est convoluée avec un signal impulsion unité situé au temps $t = f/k$. Donc, la distribution de Wigner de la réponse impulsionnelle est décalée dans le temps de f/k comme le montre la figure I.9(b), et devient

$$W_1(t, f) = W_h\left(t - \frac{f}{k}, f\right). \quad (\text{I.35})$$

La deuxième opération consiste à multiplier le résultat de la première opération avec un signal modulé en fréquence (de fréquence instantanée $-kt$), ce qui correspond à une convolution fréquentielle entre les distributions de Wigner correspondantes (cf. éq (I.20) et (I.23)), donc à un décalage en fréquence

$$W_2(f, t) = W_1(t, k + kt); \quad (\text{I.36})$$

on obtient la distribution de Wigner schématisée figure I.9(c).

La troisième opération consiste à nouveau en une convolution avec le premier signal modulé en fréquence (de fréquence instantanée kt), c.-à-d.

$$W_3(t, f) = W_2\left(t - \frac{f}{k}, f\right); \quad (\text{I.37})$$

on obtient la distribution de Wigner schématisée figure I.9(d). Cette opération est nécessaire afin de déformer la distribution dans le sens des temps négatifs ; elle permet de retrouver la forme correcte de la distribution de Wigner de la réponse impulsionnelle du système, mais avec une rotation de 90° et deux facteurs d'échelle, soit $W_3(t, f) = W_h\left(-\frac{f}{k}, kt\right)$.

À ce stade, on obtient, à une constante près, la transformée de Fourier $H(kt)$, qu'il faut inverser si on veut retrouver un signal temporel.

Mise en œuvre pratique Grâce à la troisième opération, il est possible d'utiliser un signal d'excitation modulé en fréquence dont la vitesse de balayage est illimitée. Mais comme on veut mesurer une réponse impulsionnelle de durée t_d et occupant un spectre limité à f_b , le plus simple est de choisir une vitesse de balayage $k = f_b/t_d$. La sortie du canal acoustique s'étendra alors sur la zone temporelle $[-t_d; 2t_d]$ (cf. fig. I.9(b)). Le signal d'excitation, théoriquement de durée infinie, peut donc être tronqué à l'intervalle $[-2t_d; 2t_d]$ [258]. Le deuxième signal modulé en fréquence, utilisé pour le produit temporel, peut se limiter à l'intervalle $[-t_d; 2t_d]$ (cf. fig. I.9(b)). Par convolution avec le troisième signal modulé en fréquence, la sortie du système occupera un intervalle temporel $[-t_d; t_d]$; le troisième signal doit théoriquement occuper l'intervalle $[-3t_d; 2t_d]$. Il est cependant préférable de mener cette dernière opération dans le domaine fréquentiel en manipulant des transformées de Fourier ; le signal modulé en fréquence peut alors occuper un intervalle fréquentiel $[-f_b; 0]$ et un intervalle temporel $[-t_d; 0]$.

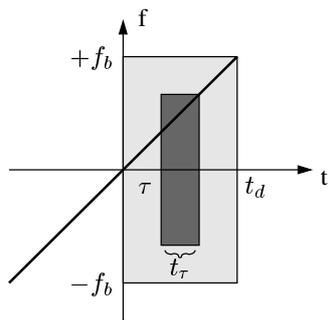
L'échantillonnage en sortie pourra se faire à une fréquence $f_s \geq 2f_b$, car le contenu fréquentiel du signal est limité à f_b après chaque opération.

Et, comme nous l'avons déjà mentionné, il faut procéder à une transformée de Fourier inverse avant de retrouver un signal temporel correspondant à la réponse impulsionnelle.

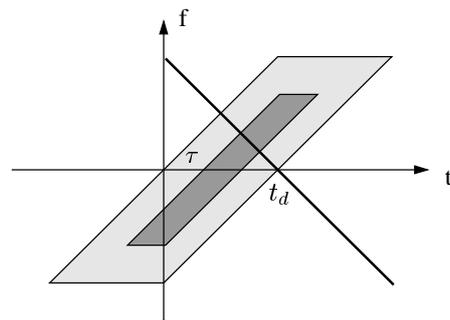
Comme dans le cas de la mesure par MLS, il est possible de moyenniser plusieurs mesures identiques. Si l'on répète N fois la mesure, on obtient une amélioration du rapport signal/bruit d'un facteur N [258].

Mais à l'inverse de la mesure par MLS, il n'est pas nécessaire d'émettre au moins une fois le balayage fréquentiel avant de réaliser l'enregistrement. Par contre, l'enregistrement doit être effectué sur une durée au moins égale à celle du balayage, la durée totale tenant compte du temps de réverbération supposé à chaque fréquence (on peut considérer un signal modulé linéairement en fréquence comme une impulsion dont chaque composante fréquentielle est retardée d'une durée proportionnelle à cette fréquence).

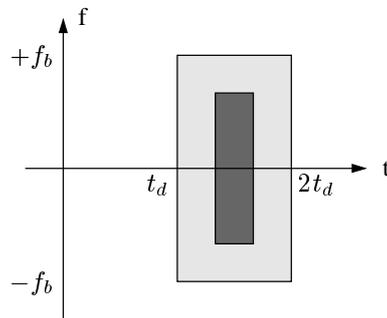
Simplification du dispositif On peut montrer que la mesure d'un système par balayage fréquentiel peut être réduite à deux opérations : l'émission dans le système d'un signal modulé en fréquence, et la convolution par le signal modulé inverse. On obtient alors directement un signal temporel, qui est la réponse impulsionnelle décalée du système mesuré (cf. fig. I.10(c)) ; ce dispositif est proposé par Farina [80]. Cette méthode se rapproche beaucoup de la mesure d'une réponse impulsionnelle à l'aide de MLS ; dans les deux cas, le calcul de la réponse impulsionnelle à partir du signal enregistré ne nécessite qu'une



(a) Réponse impulsionnelle dans le plan temps-fréquence, et sous-système débutant à $t = \tau$



(b) Limites du système après la première opération (convolution)



(c) Limites du système après la deuxième opération (convolution)

FIGURE I.10 – Représentation graphique des opérations effectuées lors de la mesure simplifiée d'un système à l'aide d'un balayage fréquentiel linéaire, tel que décrit par Farina [80]

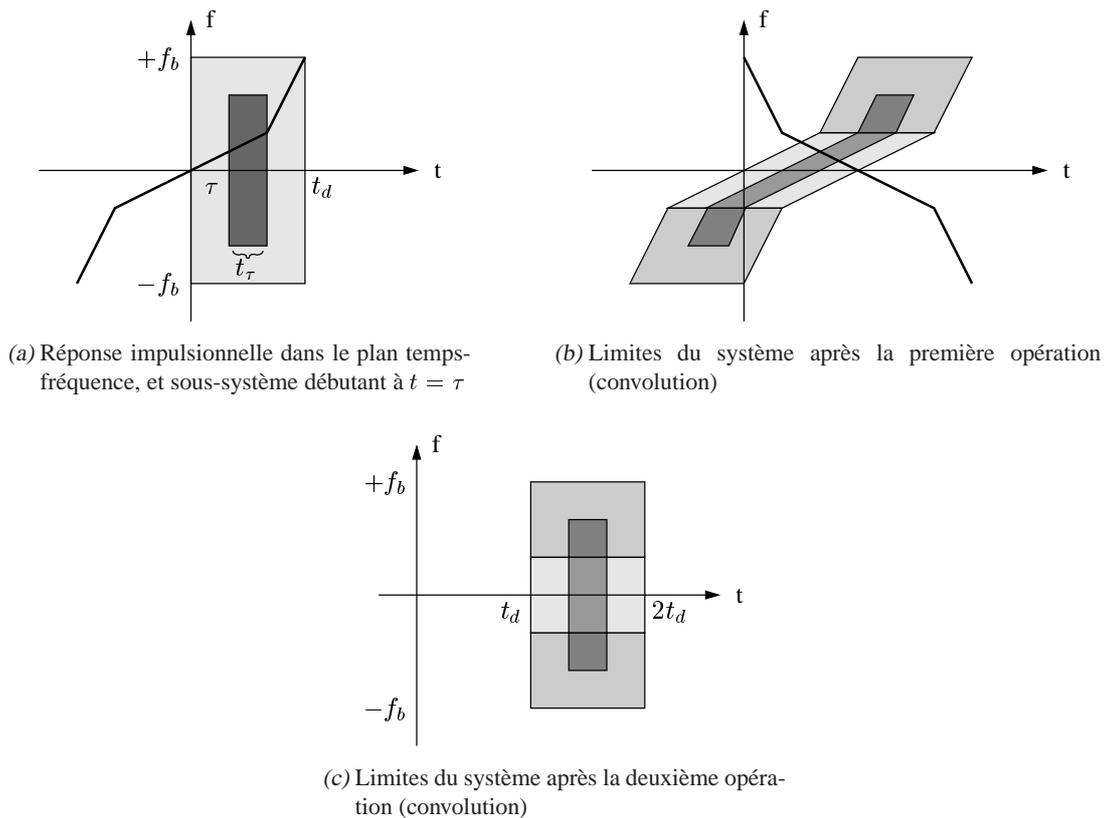


FIGURE I.11 – Représentation graphique des opérations effectuées lors de la mesure simplifiée d'un système à l'aide d'un balayage fréquentiel non linéaire, tel que décrit par Farina [80]

seule opération (convolution ou intercorrélacion) qui peut, en outre, être effectuée par un algorithme rapide.

En outre, il propose également d'effectuer un balayage fréquentiel non linéaire. Généralement, on accepte de passer plus de temps dans les basses fréquences pour obtenir un meilleur rapport signal/bruit. Dans le plan temps-fréquence, le signal enregistré possède alors une énergie dilatée en basses fréquences (cf. fig. I.11(b)). Pour compenser cette différence d'énergie, il faut moduler en *amplitude* le signal de démodulation, la modulation d'amplitude étant égale à l'inverse de la modulation en fréquence.

Synchronisation des dispositifs d'émission et de réception À l'inverse de la mesure par MLS, il n'est pas absolument nécessaire que les signaux d'émission et de sortie soient *parfaitement* synchronisés. En l'absence de synchronisation parfaite, la vitesse de glissement de fréquence du signal émis et du signal avec lequel on multiplie le signal reçu ne seront pas exactement les mêmes. Formellement, l'équation (I.32) est multipliée par une exponentielle complexe (c.-à-d. la réponse en fréquence d'un filtre passe-tout) dont la phase est linéaire. Cela se traduit dans le plan temps-fréquence par une rotation globale de la figure (cf. fig. I.12); mais en tout état de cause, le rapport signal/bruit est conservé (cf. fig. I.13), ainsi que les principaux indices acoustiques.

I.3.1.4 Comparaison des méthodes

En ce qui concerne les non-linéarités, la mesure par balayage fréquentiel permet d'obtenir un meilleur rapport signal/bruit que la mesure par MLS non préaccentuées si la fonction de transfert est bien plate. Par contre, dans le cas de la présence de résonances marquées, la mesure par MLS se révélera plus avan-

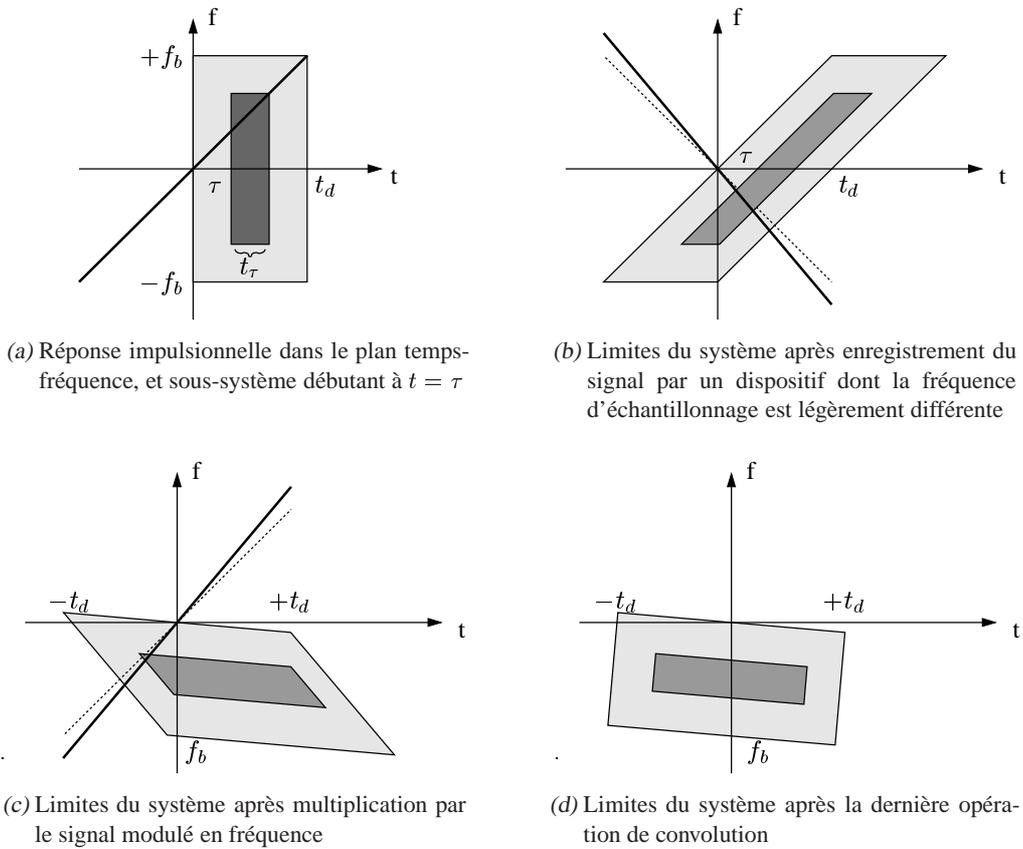


FIGURE I.12 – Représentation graphique des opérations effectuées lors de la mesure d'un système par balayage fréquentiel, avec défaut de synchronisation des horloges d'échantillonnage entre le dispositif d'émission et le dispositif de réception

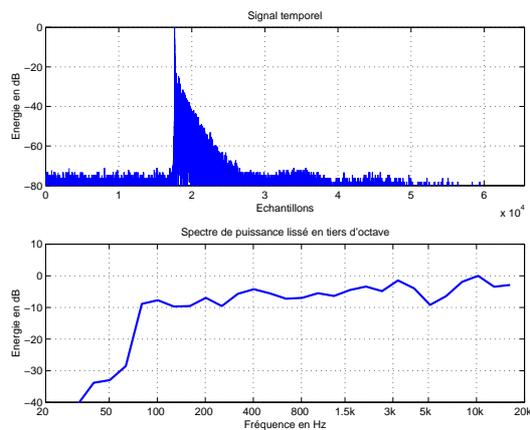


FIGURE I.13 – Réponse impulsionnelle calculée avec un appareil numérique pour la lecture, et un autre pour l'enregistrement, par la méthode du balayage fréquentiel ; l'écart relatif entre les deux fréquences d'horloge est de 2%

tageuse en terme de rapport signal/bruit [258]. En tout état de cause, la mesure par MLS préaccentuées se révèle plus efficace qu'une mesure par balayage fréquentiel.

D'autre part, pour une même résolution fréquentielle, une mesure par balayage fréquentiel nécessitera un temps de calcul beaucoup plus long qu'une mesure par MLS, et est très complexe à implanter, que ce soit au niveau électronique qu'au niveau informatique.

Le système simplifié de mesure par balayage fréquentiel proposé par Farina [80] est par contre très perspécace, car il ne nécessite qu'une seule opération après réception du signal, opération pouvant être effectuée grâce à un algorithme rapide. Ce système conjugue les avantages de la mesure par MLS (rapidité) et par balayage fréquentiel (rapport signal/bruit, immunité aux non-linéarités et à la non-synchronisation des dispositifs d'émission et de réception).

La mesure par MLS nécessite l'émission d'une séquence « à blanc » précédant la première séquence utile à la mesure ; alors que tout système par balayage fréquentiel nécessite au contraire l'enregistrement d'une tranche supplémentaire et consécutive de signal, liée au temps de réverbération.

I.3.1.5 Utilisation dans le contexte automobile

Un dispositif nécessitant une lecture et un enregistrement synchrones n'est pas souple d'utilisation dans un contexte automobile, puisqu'il est nécessaire de démonter l'autoradio pour accéder aux connexions électriques vers les haut-parleurs ; or quand bien même cette opération est techniquement faisable (ce qui n'est pas toujours le cas), il faut encore repérer le fil exact relié au haut-parleur que l'on désire mesurer, et y connecter la sortie du dispositif de lecture. D'autre part, il est impossible d'inclure l'amplificateur de l'autoradio dans le chaîne de mesure, puisque celui-ci ne possède pas d'entrée/sortie externe.

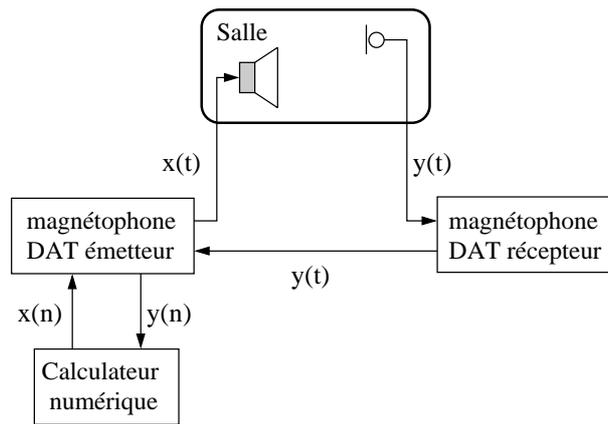
Nous désirons élaborer un nouveau dispositif qui permettra d'effectuer une mesure de réponse impulsionnelle acoustique dans un contexte automobile, de manière souple et asynchrone, c.-à-d. avec le lecteur de disque compact comme dispositif de lecture, et un ordinateur muni d'une bonne carte son comme dispositif d'enregistrement. Il est tout à fait envisageable de stocker les MLS sur cédérom, de diriger le signal lu vers le haut-parleur choisi, et d'enregistrer directement le signal en numérique sur l'ordinateur.

Cependant on se heurte à une difficulté majeure : le dispositif de lecture et le dispositif d'enregistrement possèdent des fréquences d'horloge différentes. Nous verrons comment il est possible de contourner ce problème ; et nous verrons qu'il est même possible de retrouver les retards de propagation relatifs entre les différents haut-parleurs mesurés dans une même session.

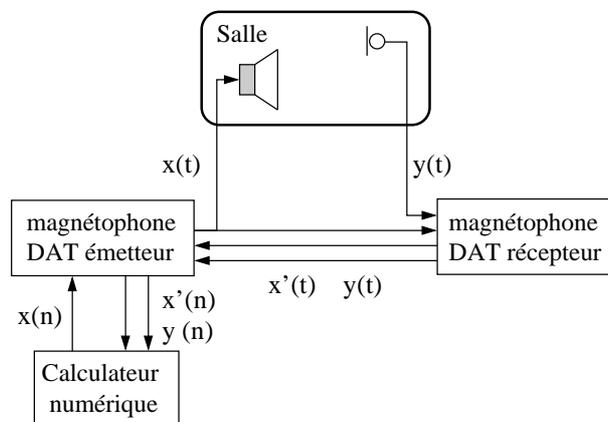
I.3.2 Dispositifs découplés

I.3.2.1 Première approche

La conservation de la durée de la période (mesurée en nombre de points) entre l'émission et la réception est un facteur déterminant pour la réussite de la mesure. Une approche originale est développée par Jot [138]. Il suggère d'utiliser deux magnétophones numériques (un pour l'émission, un pour la réception), et de transférer à nouveau le signal enregistré du récepteur vers l'émetteur (cf. fig. I.14(a)). Le fait de transférer $y(t)$ du récepteur vers l'émetteur permet de s'affranchir de l'écart entre les fréquences des deux horloges d'échantillonnage indépendantes. La réponse impulsionnelle est calculée par intercorrélation d'une période des deux signaux du dispositif émetteur (la séquence lue et le signal enregistré). Le succès de la mesure ne dépend que de la stabilité des horloges, en supposant que cette stabilité est supérieure à l'incertitude sur leur fréquence. Mais la référence temporelle, permettant de mesurer la durée entre le haut-parleur et le microphone, est perdue.



(a) Transfert du signal enregistré vers le dispositif émetteur



(b) Transfert du signal émis et du signal enregistré vers le dispositif émetteur

FIGURE I.14 – Dispositifs imaginé par Jot [138] pour s'affranchir de l'écart des fréquences d'échantillonnage entre les deux dispositifs

Jot propose alors une amélioration de ce système, consistant à enregistrer le signal source sur le magnétophone récepteur, et à transférer les deux signaux vers le magnétophone émetteur (cf. fig. I.14(b)). La précision de la réponse impulsionnelle est améliorée par une étape de déconvolution supplémentaire. En effet, la réponse impulsionnelle $h(t)$ est calculée en déconvoluant la réponse impulsionnelle C_{xy} (incluant le canal acoustique) par la réponse impulsionnelle $C_{xx'}$ (trajet direct). Cette déconvolution peut même être effectuée au moyen de l'algorithme de FFT, car les deux réponses C_{xx} et C_{xy} ayant un support plus court que la période L , on peut les prolonger par un échantillon nul. Un avantage supplémentaire de cette déconvolution est qu'elle restitue l'instant initial dans la réponse impulsionnelle comme instant d'émission de l'impulsion par le haut-parleur, ce qui permet de mesurer la durée du trajet entre le haut-parleur et le microphone. Mais l'inconvénient de cette procédure est que, pour la mesure d'une réponse impulsionnelle stéréophonique, elle nécessite trois canaux d'enregistrements.

Ces deux dispositifs ne sont pas applicables dans un contexte automobile, puisque nous désirons utiliser le lecteur de disque compact de l'autoradio comme dispositif de lecture ; or celui-ci n'est pas en mesure d'enregistrer un signal.

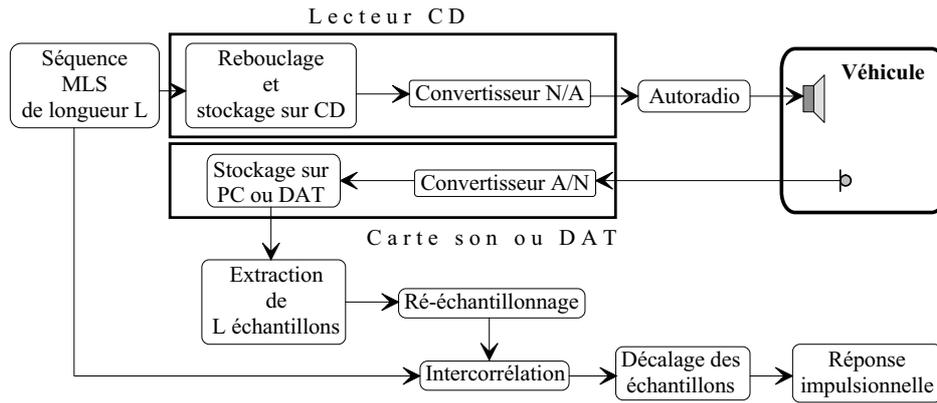


FIGURE I.15 – Proposition d'une nouvelle installation destinée à rendre la lecture des MLS indépendante de leur enregistrement

Si on veut utiliser le lecteur de disque compact de l'autoradio comme dispositif d'émission des MLS, il faut stocker sur cédérom un signal audio composé d'un grand nombre de MLS mises bout à bout, puis planer la lecture du disque. Pour une mesure des deux haut-parleurs avant ou arrière, la sélection du haut-parleur diffusant le signal peut être effectuée séquentiellement à l'aide des deux canaux gauche et droite, en ayant au préalable réglé de manière adéquate le dispositif de balance avant/arrière sur l'autoradio ; si l'on veut mesurer les quatre réponses impulsionnelles, il faut procéder en trois temps : enregistrer les deux haut-parleurs avant, puis modifier le réglage de la balance avant/arrière, et enregistrer les deux haut-parleurs arrière. L'enregistrement du signal diffusé par le haut-parleur sera effectué sur le disque dur d'un ordinateur via une carte de numérisation de bonne qualité comportant deux canaux parfaitement synchronisés. Cet enregistrement doit être rééchantillonné ; puis il faut extraire kL échantillons, et les moyenner pour obtenir L échantillons, avant de calculer l'intercorrélacion circulaire à l'aide d'un algorithme rapide utilisant la transformée de Hadamard. Ces kL échantillons peuvent être choisis à n'importe quel endroit du signal enregistré. Comme on a perdu la synchronisation avec la lecture, on a également perdu le retard absolu de propagation entre la source et le récepteur. Suivant l'endroit où ont été extraits les échantillons, on verra apparaître un retard de propagation différent, mais en aucun cas significatif de la réalité physique. En sortie du calcul de l'intercorrélacion circulaire, le signal est bien entendu rebouclé sur lui-même, mais il suffit de décaler les échantillons pour le rendre causal. Le schéma de principe de ce dispositif est présenté figure I.15.

Pour mener à bien ce processus, il faut effectuer deux opérations :

1. Quantifier très précisément la différence de fréquence d'échantillonnage entre les deux dispositifs.
2. Rééchantillonner le signal enregistré pour que sa nouvelle fréquence d'échantillonnage soit strictement identique à celle du dispositif de lecture.

L'estimation de la différence de fréquence d'échantillonnage doit être effectuée à chaque fois que l'on change de dispositif de lecture ou d'enregistrement, selon le schéma de la figure I.16. Mais si l'on fait l'hypothèse qu'une dérive d'horloge est possible sur le même appareil au cours du temps, il peut être judicieux de l'estimer avant chaque mesure. Nous avons rencontré un tel phénomène sur les lecteurs de cédérom de certains ordinateurs !

I.3.2.2 Estimation du décalage des fréquences d'échantillonnage

Nous présentons ici deux méthodes qui permettent d'estimer le décalage de fréquence d'échantillonnage entre deux dispositifs électroniques. Pour un panorama plus complet, le lecteur se reportera à [180, 219].

L'idée consiste à faire émettre par le premier dispositif un signal connu (en l'occurrence une sinusoïde

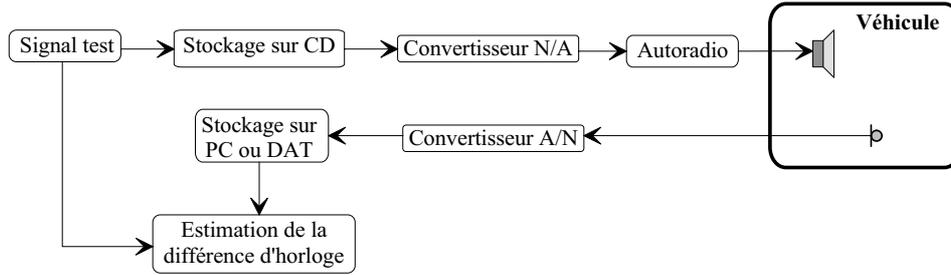


FIGURE I.16 – Schéma synoptique de l'installation destinée à estimer la différence de fréquence d'échantillonnage entre le dispositif de lecture et le dispositif d'enregistrement

de fréquence connue), à enregistrer ce signal par le deuxième dispositif, et à le comparer avec le signal émis.

Passages par zéro Cette méthode est décrite dans [36, 48, 312]. Le principe consiste à mesurer les instants où le signal reçu passe par l'origine des ordonnées. Cette mesure peut être précise grâce à l'utilisation d'interpolations. En évaluant la distance entre ces instants de passage par zéro et les instants théoriques déterminés par la fréquence de la sinusoïde d'entrée, on peut mesurer la distorsion de l'échelle fréquentielle.

Si t est le temps théorique, et τ le temps mesuré, appelons f la fonction exprimant l'un par rapport à l'autre, c.-à-d. $t = f(\tau)$.

Si on appelle f_{lec} la fréquence d'échantillonnage du dispositif de lecture, f_{enr} la fréquence d'échantillonnage du dispositif d'enregistrement, et f_{sinus} la fréquence de la sinusoïde injectée, on peut écrire

$$\frac{f_{\text{enr}}}{f_{\text{lec}}} = \frac{\tau_k - \tau_0}{k} \frac{f_{\text{sinus}}}{f_{\text{lec}}}, \quad (\text{I.38})$$

avec k le numéro de la période de la sinusoïde, ce qui peut se récrire

$$\tau_k = \tau_0 + k \frac{f_{\text{enr}}}{f_{\text{sinus}}}, \quad (\text{I.39})$$

soit finalement

$$f_{\text{enr}} = \frac{\tau_k - \tau_0}{k} f_{\text{sinus}}. \quad (\text{I.40})$$

Si τ_k ne correspond pas à un nombre entier d'échantillons, on procède à une interpolation linéaire de la sinusoïde enregistrée entre les deux échantillons les plus proches.

On peut également mesurer les passages par zéros de la dérivée du signal reçu [292]. On montre que la valeur de la dérivée aux points intermédiaires

$$\bar{\tau}_i = \frac{\tau_{i+1} + \tau_i}{2} \quad (\text{I.41})$$

est égale au rapport de l'incrément de temps théorique sur l'incrément mesuré, soit

$$\frac{df}{d\tau}(\bar{\tau}_i) = \frac{T/4}{\Delta\tau} \quad \text{avec } \Delta\tau = \tau_{i+1} - \tau_i, \quad (\text{I.42})$$

où T est la période de la sinusoïde. Les autres valeurs de $\frac{df}{d\tau}$ peuvent être déterminées par interpolation linéaire ; cette interpolation n'étant pas très fiable, on peut moyenner la fonction $\frac{df}{d\tau}$ ainsi obtenue pour éliminer le bruit introduit. Les instants effectifs d'échantillonnage sont donc :

$$t_k = t_{k-1} + \Delta\tau f(k\Delta\tau). \quad (\text{I.43})$$

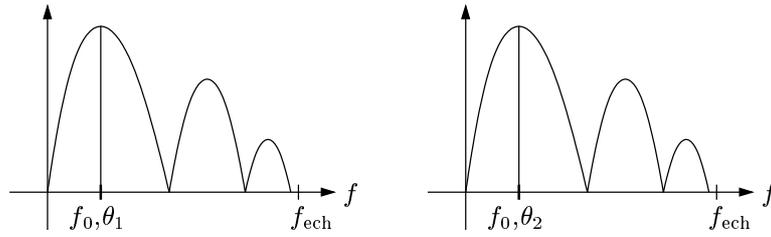


FIGURE I.17 – Transformée de Fourier à temps continu de deux tranches successives d'une sinusoïde : fréquence fondamentale continue et phase associée

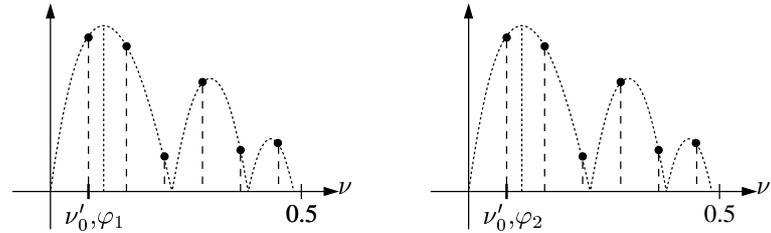


FIGURE I.18 – Transformée de Fourier discrète de deux tranches successives d'une sinusoïde échantillonnée : fréquence fondamentale discrète et phase associée

Evaluons la sensibilité au bruit de cette mesure [312]. Soit σ_n l'écart type d'un bruit additif ; l'écart type des écarts d'instant d'échantillonnage, noté σ_τ , est alors égal à l'écart type du bruit divisé par la valeur absolue de la dérivée du signal échantillonné aux points de passage par zéro, c.-à-d.

$$\sigma_\tau = \frac{\sigma_n}{A\omega_{\text{sinus}}}, \quad (\text{I.44})$$

avec A et ω_{sinus} l'amplitude et la pulsation de la sinusoïde injectée.

Rotation de phase Il est également possible de procéder à deux TFD sur deux tranches successives du signal enregistré, ces tranches devant être les plus grandes possibles. Le signal acquis correspondant théoriquement à une sinusoïde, les deux TFD présentent chacune un pic à la fréquence réduite la plus proche de la fréquence correspondante de la sinusoïde ; mais la résolution fréquentielle sera toujours insuffisante pour détecter des différences infimes de fréquences. En revanche, si la phase associée à la raie varie, cela sera le signe d'une fréquence instantanée différente de la fréquence du pic.

DÉTERMINATION DE LA FRÉQUENCE RÉDUITE D'UNE SINUSOÏDE ÉCHANTILLONNÉE

Supposons que l'on procède à deux transformées de Fourier continues sur deux tranches successives de taille T secondes d'une sinusoïde continue de fréquence f_0 . Chaque transformée présente ici un pic à la fréquence f_0 , convolué par la transformée de Fourier de la fenêtre de troncature. Soit θ_1 et θ_2 la phase associée au pic de la première et de la deuxième transformée de Fourier (cf. fig. I.17).

Procédons maintenant à la TFD de ces deux même tranches de la sinusoïde échantillonnée à la fréquence f_{ech} . On suppose que ces tranches sont de longueur $N = Tf_{\text{ech}}$, où N est une puissance de 2. Les deux TFD présentent alors un pic à la fréquence réduite ν_0 , où $N\nu_0$ est l'entier le plus proche du rationnel Nf/f_{ech} ; ce pic représente en fait l'échantillon de la transformée de Fourier analogique à la fréquence $f'_0 = \nu'_0 f_{\text{ech}}$. Soit φ_1 et φ_2 la phase du pic associée respectivement à la première et à la deuxième TFD (cf. fig. I.18). La phase de ce pic est égale à la phase du pic de la transformée de Fourier continue à la fréquence f_0 , ajoutée d'un biais correspondant à la transformée de Fourier de la fenêtre de troncature. Si l'on suppose que la sinusoïde est d'amplitude constante, ce biais est le même entre les

deux tranches, puisque la TFD de la fenêtre de troncature est constante. La différence entre les phases $\varphi_2 - \varphi_1$ sera donc égale à la différence entre les phases de chaque pic des deux transformées de Fourier continues $\theta_2 - \theta_1$. Cette déviation de la phase provient en fait de la différence entre la fréquence exacte de la sinusoïde f_0 et la fréquence $f'_0 = \nu_0 f_{\text{ech}}$ correspondant à un échantillon fréquentiel de la TFD.

Supposons que la fréquence de la sinusoïde corresponde à une fréquence réduite de la TFD, donc en fait $f_0 = f'_0 = f_{\text{ech}}(n-1)/N$, avec n entier tel que $n \leq N/2 + 1$. Supposons également que l'espacement entre le début de ces deux tranches soit de N échantillons ; cet espacement est proportionnel à la période de la sinusoïde. Sa phase est donc exactement la même dans les deux tranches.

Dans le cas qui nous intéresse, la fréquence de la sinusoïde est différente de la fréquence f' du pic de la TFD. La différence de phase $\theta_2 - \theta_1$ correspond alors à la différence des fréquences, c.-à-d. $\theta_2 - \theta_1 = 2\pi T(f_0 - f'_0)$. Connaissant la fréquence réduite ν_0 du pic de la TFD, ainsi que cette différence de phase, on en déduit la fréquence précise de la sinusoïde

$$f_0 = \nu'_0 f_{\text{ech}} + \frac{\varphi_2 - \varphi_1}{2\pi T}, \quad (\text{I.45})$$

ou sa fréquence réduite

$$\nu_0 = \nu'_0 + \frac{1}{N} \frac{\varphi_2 - \varphi_1}{2\pi}. \quad (\text{I.46})$$

Il est bien entendu souhaitable que la taille de la TFD soit la plus grande possible, afin de garantir une précision importante de l'estimation de la fréquence du pic ν'_0 et de sa phase associée ; d'autre part, il est également souhaitable que l'espacement entre les deux trames ne soit pas trop grand afin que θ_1 et θ_2 ne diffèrent pas de plus de 2π . Il s'agit donc d'un compromis entre le type de défaut de fréquence auquel on s'attend, la fréquence de la sinusoïde de référence, la longueur et l'espacement entre les trames.

Pour plus de clarté, considérons un exemple dans lequel la fréquence de la sinusoïde de référence est de 1000 Hz, et la fréquence d'échantillonnage de 48 kHz. Si on calcule la TFD d'une tranche de 512 échantillons, on observera une raie principale sur le 11^{ème} point de la TFD. Cependant, la résolution fréquentielle de 93,75 Hz est nettement insuffisante pour détecter des très faibles différences de fréquence. En revanche, si on procède à une TFD glissante tous les 48 points, et si la phase associée à la raie varie, ce sera le signe d'une fréquence différente.

APPLICATION À NOTRE ÉTUDE

Le signal numérique de référence consiste en une sinusoïde de fréquence réduite ν_{lec} . Cette sinusoïde passe au travers d'un convertisseur numérique/analogique de fréquence d'échantillonnage f_{lec} : la sinusoïde analogique est donc de fréquence $f = \nu_{\text{lec}} f_{\text{lec}}$. Celle-ci est numérisée par le convertisseur analogique/numérique à la fréquence d'échantillonnage f_{enr} : le signal obtenu est donc de fréquence réduite

$$\nu_{\text{enr}} = \nu_{\text{lec}} \frac{f_{\text{lec}}}{f_{\text{enr}}}, \quad (\text{I.47})$$

d'où on peut déduire facilement le rapport des deux fréquences d'échantillonnage

$$\frac{f_{\text{lec}}}{f_{\text{enr}}} = \frac{\nu_{\text{enr}}}{\nu_{\text{lec}}}; \quad (\text{I.48})$$

ν_{lec} est parfaitement connu car c'est la fréquence réduite de la sinusoïde de référence, et ν_{enr} est calculé avec précision grâce à la formule (I.46).

Comparaison entre ces deux méthodes En pratique, le signal enregistré lors de l'émission de la sinusoïde de référence comporte non seulement une composante fréquentielle à la fréquence théorique de la sinusoïde, mais également diverses harmoniques dues à des phénomènes de distorsion et d'interférences.

La méthode des passages par zéro n'est pas robuste à cette dégradation du signal. En effet, les composantes fréquentielles supplémentaires ne garantissent plus un passage par zéro du signal aux instants prévus. La mesure du rapport de fréquences d'échantillonnage n'est précise qu'au dixième de Hz près, ce qui n'est pas satisfaisant. On peut contourner cette difficulté en filtrant le signal enregistré par un filtre passe-bande très étroit et centré sur la fréquence de la sinusoïde ; le signal obtenu est alors quasiment sinusoïdal, et on obtient une précision de l'ordre du millièème de Hz, ce qui est tout à fait satisfaisant.

La méthode par examen de la rotation de phase n'est pas sensible à ces composantes fréquentielles parasites. En effet, à partir du moment où la sinusoïde reçue est d'amplitude supérieure à celle des autres composantes fréquentielles (ce qui est heureusement vrai en pratique), ces composantes n'influent pas sur le calcul. Il n'est donc pas nécessaire de procéder à un filtrage du signal enregistré. On obtient une précision de l'ordre du millièème de Hz.

Cette deuxième méthode est donc plus simple à implanter que la première. Cependant, elle nécessite le calcul de deux TFD de taille très grande, ce qui est coûteux en temps de calcul.

Implantation logicielle et mise en œuvre Ces deux méthodes ont été implantées sous forme de modules externes pour le logiciel CoolEdit Pro™² [135, 136] fonctionnant sous Windows™ [179]. Ce choix a été motivé par le fait que les outils de mesure et d'analyse proposés par Farina [80] sont également implantés sous la même forme. Nous avons donc rajouté nos propres algorithmes au sein d'une plate-forme déjà utilisée chez Renault.

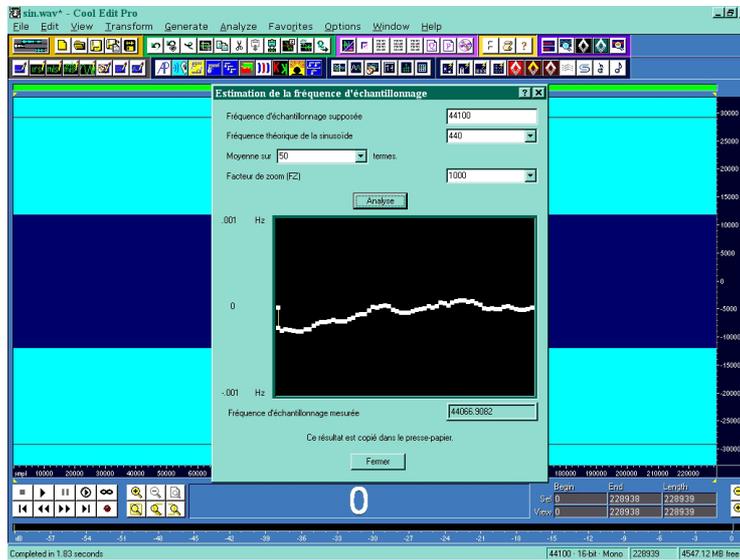
La mise en œuvre est très simple. Nous avons gravé sur un disque compact plusieurs pistes audio comportant des sinusoïdes de différentes fréquences. La procédure consiste à lire ce disque par l'intermédiaire de l'autoradio, et à enregistrer directement sur l'ordinateur de mesure le signal diffusé à l'intérieur de l'habitacle. L'enregistrement est contrôlé par le logiciel CoolEdit Pro™. Une fois le signal acquis sur plusieurs secondes, il suffit de faire appel à la fonction correspondante pour calculer la différence de fréquence d'échantillonnage entre les deux dispositifs.

Nous procédons à un filtrage passe-bande très étroit avant de faire appel à la première méthode (passages par zéro), comme présenté sur la figure I.19(a) ; puis nous introduisons la fréquence d'échantillonnage supposée, la fréquence de la sinusoïde de référence, et le nombre de termes sur lequel sera calculé la moyenne. Une fois le calcul lancé, le résultat est affiché sous forme de fréquence d'échantillonnage apparente. De plus, un graphique présente l'évolution de l'estimation de la fréquence d'échantillonnage moyennée en fonction du nombre de passages par zéro observés ; la variation est inférieure au millièème de Hz.

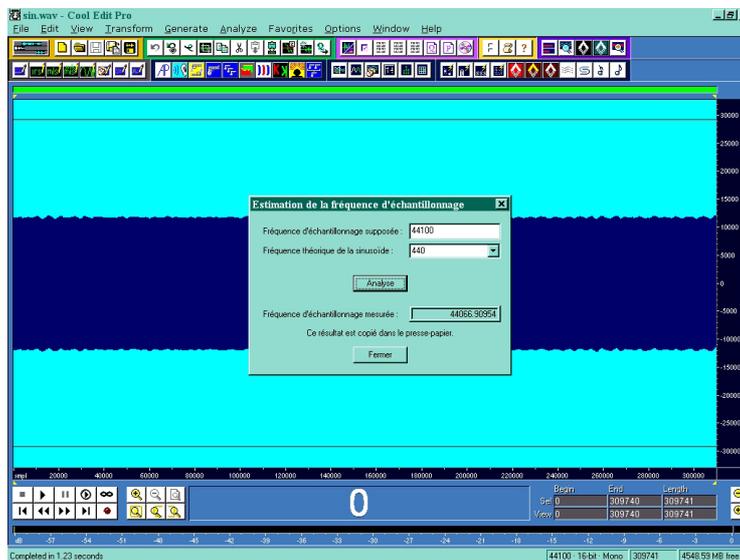
La deuxième méthode peut être exécutée directement sur le signal enregistré (cf. fig. I.19(b)) ; il est juste nécessaire d'indiquer au programme la fréquence d'échantillonnage supposée et la fréquence de la sinusoïde de référence. Le calcul dure quelques secondes, et la comparaison avec le résultat fourni par la première méthode montre que les deux estimations concordent à un millièème de Hz près.

Pour les deux méthodes, le résultat affiché dans la fenêtre correspond à la fréquence d'échantillonnage apparente du dispositif d'enregistrement, en supposant que la fréquence d'échantillonnage du dispositif de lecture est de 44100 Hz. Pour l'exemple donné figure I.19, la fréquence d'échantillonnage calculée est de 44066,90954 Hz ; cela veut dire qu'il faut augmenter la fréquence d'échantillonnage de tout signal lu par le lecteur de disque compact de l'autoradio et enregistré par l'ordinateur : la nouvelle fréquence d'échantillonnage doit être de 44133,11531 Hz.

2. <http://www.syntrillium.com/cep/index.html>



(a) Calcul de la fréquence de la sinusoïde enregistrée à l'aide de la méthode des passages par zéro



(b) Calcul de la fréquence de la sinusoïde enregistrée à l'aide de la méthode de la rotation de phase

FIGURE I.19 – Présentation de l'implantation des deux méthodes de calcul de la fréquence d'une sinusoïde sous forme de modules externes pour le logiciel CoolEdit Pro™

I.3.2.3 Modification de la fréquence d'échantillonnage

Une fois estimée le décalage de fréquence entre le dispositif de lecture et le dispositif d'enregistrement, il faut procéder à la modification de la fréquence d'échantillonnage du signal enregistré.

Principe Le changement de fréquence d'échantillonnage trouve une solution simple lorsque les deux fréquences sont dans un rapport P/Q , où P et Q sont des entiers. Pour passer d'une fréquence F_1 à une fréquence $F_2 = F_1 \cdot P/Q$, on suréchantillonne d'un facteur P , puis on sous-échantillonne d'un facteur Q [243]. Il est nécessaire de commencer par suréchantillonner pour ne pas perdre de contenu spectral.

Pour un changement de fréquence en rapport non rationnel α , on peut chercher la fraction rationnelle la plus proche, et procéder comme ci-dessus. Pour cela, on a intérêt à utiliser le développement en fractions rationnelles [62] où

$$\alpha = a_0 + \frac{1}{a_1 + \frac{1}{a_2 + \dots}}, \quad (\text{I.49})$$

car les fractions rationnelles obtenues en tronquant le développement réalisent en général de meilleures approximations que les fractions décimales.

Si cette approximation n'est pas suffisante, on peut réaliser le changement de fréquence par implantation d'une approximation du filtre passe-bas idéal (sinus cardinal). Cela consiste à reconstruire le signal analogique, puis à l'échantillonner à la nouvelle fréquence. En effet, si $x(n)$ représente le signal original échantillonné à la fréquence F_1 et $y(n)$ le signal après conversion, on remarque que cette conversion résulte de deux étapes (on suppose que l'on passe d'une fréquence basse à une fréquence plus élevée) :

1. Calcul du signal continu correspondant à $x(n)$, c.-à-d.

$$z(t) = x(n) * \text{sinc}_{F_1}(t) = \sum_{i=-\infty}^{+\infty} x(i) \frac{\sin \pi F_1 (t - i/F_1)}{\pi F_1 (t - i/F_1)}, \quad (\text{I.50})$$

où $\text{sinc}_{F_1}(t)$ représente la réponse impulsionnelle du filtre passe-bas idéal (non causal et non stable) de fréquence de coupure $F_1/2$.

2. Échantillonnage de $z(t)$ à la fréquence F_2 , c.-à-d.

$$y(n) = z(n/F_2) = \sum_{i=-\infty}^{+\infty} x(i) \frac{\sin \pi F_1 (n/F_2 - i/F_1)}{\pi F_1 (n/F_2 - i/F_1)}, \quad (\text{I.51})$$

où on a éliminé le filtrage passe-bas puisque par hypothèse $F_2 > F_1$. Ainsi, pour calculer $y(n)$, on a besoin de la réponse impulsionnelle du filtre de reconstruction aux instants $t = n/F_2 - j/F_1$. L'idée consiste alors à effectuer une approximation des valeurs continues de la réponse impulsionnelle par interpolation linéaire des valeurs stockées dans une table. Une fois que la réponse impulsionnelle du filtre d'interpolation a été calculée (par exemple par la méthode de la fenêtre (cf. § C.1.6)), on l'échantillonne à une fréquence F_f et on la stocke dans une table ; on choisit la fréquence F_f suffisamment grande pour que l'erreur générée par l'interpolation linéaire soit inférieure au pas d'échantillonnage [296]. Les échantillons $y(n)$ sont ensuite directement obtenus par la formule (I.51) dans laquelle la sommation est limitée par la taille de la table, et les valeurs continues de la réponse impulsionnelle sont interpolées à partir des valeurs fournies par la table.

Dans le cas de signaux de durée finie (ou dans le cas d'un échantillonnage non uniforme [42]), on peut effectuer des interpolations de Lagrange [265], c.-à-d.

$$z_k^L(t) = \prod_{\substack{i=-N_1 \\ i \neq k}}^{N_2} \frac{t - t_i}{t_k - t_i}, \quad (\text{I.52})$$



FIGURE I.20 – Boîte de dialogue concernant la modification de la fréquence d'échantillonnage d'un signal, réalisée sous forme de module externe pour le programme CoolEdit Pro™

où $t_i = i/F_1$ sont les instants d'échantillonnage. Bien sûr, ces fonctions n'assurent pas une reconstitution parfaite, mais on peut remarquer que, pour un échantillonnage uniforme avec $N_1 \rightarrow \infty$ et $N_2 \rightarrow \infty$, on retrouve la formule (I.50).

On peut cependant montrer qu'il n'est pas nécessaire de reconstruire intégralement le signal analogique. Ecrivons le temps continu à partir de la période d'échantillonnage initiale T_1 , soit

$$t = nT_1 + t_1, \quad (\text{I.53})$$

où n et t_1 sont le quotient et le reste de la division euclidienne de t par T_1 . Mettons-nous dans le cas non idéal, c.-à-d. avec un filtre de reconstruction $g(t)$ différent d'un sinus cardinal. L'équation (I.50) devient

$$\hat{z}(t) = \sum_{i=-\infty}^{+\infty} x(i)g(t - iT_1), \quad (\text{I.54})$$

ou encore

$$\hat{z}(nT_1 + t_1) = \sum_{m=-\infty}^{\infty} x((n - m)T_1)g_m(t_1) \quad \text{avec} \quad g_m(t_1) = g_m(mT_1 + [t]_{\text{mod } T_1}). \quad (\text{I.55})$$

Cette équation peut s'implanter à l'aide d'un filtre transverse à coefficients variables mais périodiques [265]. Si maintenant on échantillonne ce signal analogique avec une période T_2 , on obtient

$$\hat{z}(kT_2) = \hat{z}(lT_2 + t_2) = \sum_{m=-\infty}^{+\infty} x((l - m)T_2)g_m(t_2), \quad (\text{I.56})$$

avec l et t_2 le quotient et le reste de la division euclidienne de kT_2 par T_1 . Il n'est pas nécessaire de reconstruire le signal analogique ; il faut juste calculer en temps réel les coefficients $\{g_m(t_2)\}$ pour chaque échantillon de sortie.

Implantation et mise en œuvre Cette méthode a également été implantée sous forme de module externe pour le logiciel CoolEdit Pro™ [179]. Il faut indiquer au programme la fréquence d'échantillonnage du signal (en général celle déterminée à l'aide d'une des deux méthodes d'estimation de la fréquence d'échantillonnage décrites dans le paragraphe précédent), et la fréquence d'échantillonnage désirée du signal (en général 44100 Hz) ; en outre, il est nécessaire d'indiquer l'horizon d'observation, c.-à-d. la longueur du filtre de reconstruction (cf. fig. I.20).

On se doit d'établir un compromis entre le temps de calcul et la précision du résultat. Ce compromis est représenté par l'horizon d'observation. Si on choisit un horizon petit, le calcul sera rapide, mais

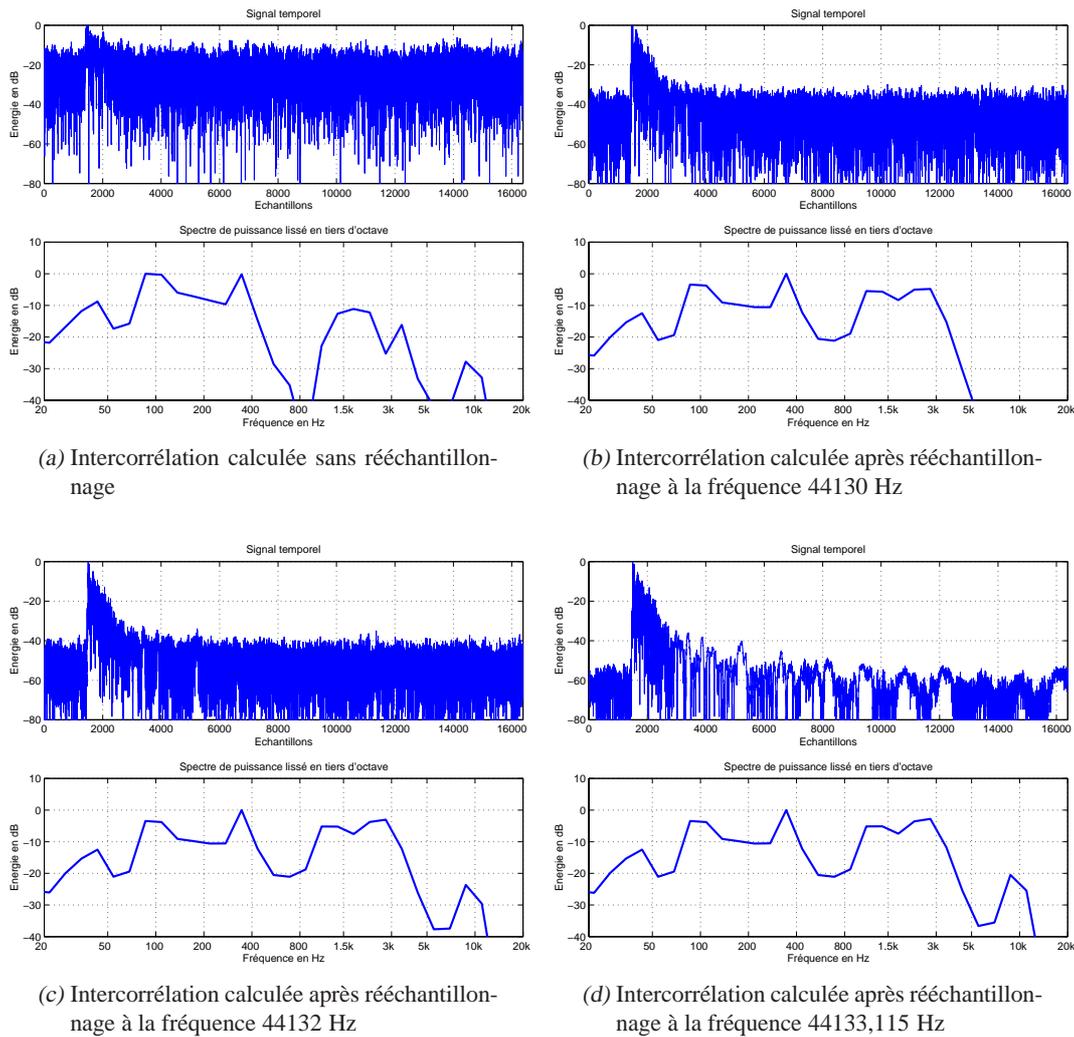


FIGURE I.21 – Influence de la précision du rééchantillonnage sur le calcul de la réponse impulsionnelle

quelque peu inexact ; si au contraire on choisit un horizon grand, l'attente du résultat sera récompensée par un rééchantillonnage proche de l'idéal. D'après les observations que nous avons réalisées, un horizon d'observation d'une dizaine d'échantillons est suffisant pour obtenir une précision du rééchantillonnage de l'ordre du millième de Hz ; le temps de calcul est fonction de la rapidité du processeur et de la longueur du signal à traiter.

Nous avons observé l'influence de la précision du rééchantillonnage sur le rapport signal/bruit de la réponse impulsionnelle (cf. fig. I.21). Il est clair que plus la précision du rééchantillonnage est importante, meilleur est le rapport signal/bruit.

I.3.2.4 Prise en compte des retards relatifs de propagation

Pour chaque haut-parleur du véhicule, on désire connaître non seulement la fonction de transfert du système total considéré (chaîne audio et habitacle), mais également l'information temporelle, et en particulier le temps de propagation de l'onde acoustique depuis le haut-parleur jusqu'au capteur de mesure.

Toute méthode synchrone est capable d'obtenir les différentes informations temporelles absolues, puisque le signal acquis contient l'information temporelle de synchronisation : son échelle temporelle



FIGURE I.22 – Descriptif schématique du protocole de mesure séquentielle des quatre haut-parleurs de l'habitacle

est définie en fonction de l'instant où est émise l'impulsion par l'émetteur. Les délais de propagation peuvent donc être aisément calculés.

Difficultés soulevés par la méthode asynchrone Lors d'une mesure par une méthode asynchrone, il n'est pas possible d'obtenir ces informations temporelles absolues. En effet, le dispositif récepteur ne communiquant pas avec le dispositif émetteur, il ne peut pas situer dans le signal qu'il reçoit l'instant zéro d'émission de l'impulsion.

Lors d'une mesure par MLS, la réponse impulsionnelle est rebouclée sur elle-même, et il se peut qu'elle ne soit pas causale ; comme une MLS est un signal périodique, il suffit d'effectuer un décalage circulaire pour rendre causale la réponse impulsionnelle. Lors d'une mesure par balayage fréquentiel, il n'y a pas de rebouclage, et la réponse impulsionnelle calculée est toujours causale. Mais dans les deux cas, aucune indication ne peut être donnée quant à un quelconque « instant zéro ».

Malgré tout, on peut envisager de récupérer les informations temporelles *relatives* entre différentes mesures effectuées *séquentiellement* : en effet, si le système récepteur est asynchrone avec le système émetteur, il l'est aussi avec lui-même d'une mesure à l'autre. Donc si on veut conserver le même référentiel pour différentes mesures, il est nécessaire que celles-ci soient effectuées au sein d'une seule acquisition. On peut alors considérer la réponse impulsionnelle d'un des haut-parleurs comme étant le référentiel temporel, puis comparer les autres réponses impulsionnelles comparativement à cette réponse. On obtient ainsi les retards relatifs entre les différents haut-parleurs ; cette information est nécessaire et suffisante pour notre étude, que ce soit pour l'analyse objective de ces réponses impulsionnelles, ou pour l'auralisation du système audio du véhicule.

Mais nous faisons face à une nouvelle difficulté : l'autoradio n'est pas capable d'alimenter indépendamment les quatre haut-parleurs, puisque le signal de mesure inscrit sur le disque compact est un signal stéréophonique. Il est donc nécessaire d'utiliser la balance manuelle avant/arrière de la voiture, et de la faire fonctionner durant l'acquisition. La solution proposée est d'émettre d'abord les signaux de mesure sur l'un puis l'autre des haut-parleurs avant, puis de laisser un temps défini au manipulateur pour régler la balance vers l'arrière, puis d'envoyer les signaux de mesure sur l'un puis l'autre des haut-parleurs arrière. Dans un souci de vérification du bon fonctionnement du processus, on peut continuer l'enregistrement en laissant d'abord un nouveau temps de manipulation pour le changement de balance vers l'avant, puis en émettant de nouveau le signal de mesure par le premier haut-parleur mesuré (cf. fig. I.22). On devra alors vérifier que la première et la dernière séquence sont identiques et n'ont aucun retard l'un par rapport à l'autre. L'ensemble de ces signaux émis doit donc occuper une seule piste audio du disque compact, et doit être enregistré en une seule opération. Il est ensuite nécessaire de découper, dans le signal ainsi acquis, les seules parties qui nous intéressent et ceci de manière totalement automatique.

La caractéristique asynchrone de la mesure soulève un autre problème. Dans le signal enregistré, il n'est pas aisé de déterminer à partir de quel échantillon commence véritablement la séquence qui nous intéresse. Ce problème diffère suivant la méthode de mesure considérée.

MÉTHODE DES MLS ASYNCHRONES

On se sert de la propriété de périodicité des MLS, et de la circularité du calcul de l'intercorrélation. Supposons que l'on extraie L échantillons du signal reçu, entre le haut-parleur n° 1 et le micro, à partir

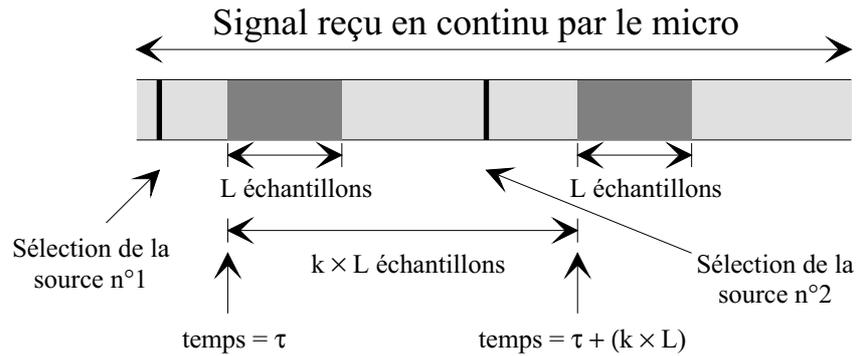


FIGURE I.23 – Principe de découpage du signal reçu en vue de la récupération des délais relatifs entre plusieurs sources sonores mesurées de manière séquentielle à l'aide de MLS

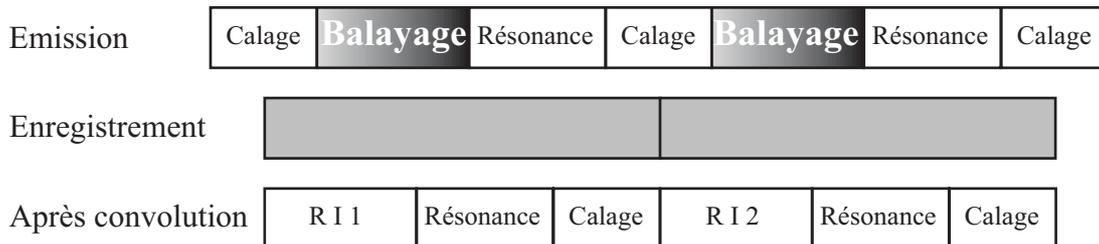


FIGURE I.24 – Principe de découpage du signal reçu lors de la mesure à l'aide d'un balayage fréquentiel

d'un temps τ tel que $\tau \geq L + 1$ échantillons après le début de la réception (on rappelle qu'il faut que la séquence entière de longueur L soit passée au moins une fois par toutes les réflexions). On observe un délai de propagation absolu appelé δ sur la réponse impulsionnelle calculée ; ce délai de propagation est tout à fait arbitraire, puisque si on avait choisi d'extraire L échantillons sur le signal reçu à partir d'un temps $\tau' \neq \tau$, on aurait obtenu un autre délai de propagation $\delta' \neq \delta$. Néanmoins, si on extrait L échantillons sur le signal reçu à partir d'un temps $\tau'' = \tau + (k \times L)$ échantillons, avec k entier naturel, le délai de propagation est identique. Donc, si on extrait L échantillons sur le signal reçu entre le haut-parleur n° 2 et le micro, à partir d'un temps τ'' , on observera un délai de propagation δ'' , lui aussi complètement arbitraire, mais qui, par différence avec δ , donnera le délai de propagation *relatif* existant entre les couples {haut-parleur n° 1 ; micro} et {haut-parleur n° 2 ; micro}. Pratiquement, il faut extraire L échantillons tous les multiples de L (cf. fig. I.23).

MÉTHODE DU BALAYAGE FRÉQUENTIEL ASYNCHRONE

Les sinusoïdes de fréquence glissante ne présentent pas le caractère cyclique des MLS et il n'est pas nécessaire d'émettre un balayage d'excitation avant le balayage de mesure ; par contre, l'enregistrement présente deux parties : une partie correspondant à l'émission du balayage fréquentiel, et une partie correspondant à la résonance de l'habitacle. Entre chaque balayage émis, il doit donc y avoir un moins ce silence de résonance.

Pour faire face au caractère asynchrone, il est possible d'ajouter un autre silence correspondant à une durée de calage (ou plutôt une tolérance temporelle) en début d'acquisition. Si on laisse un silence fixe avant chaque balayage émis, il suffira au manipulateur de sélectionner le signal débutant un peu avant le premier balayage enregistré. Pour chaque balayage émis, il faut faire l'acquisition d'une longueur égale à la durée du balayage, augmentée de la durée de la résonance et de la durée de calage ; cette dernière est répartie au début et à la fin du signal enregistré. Après la convolution avec le balayage inverse, on obtiendra la réponse impulsionnelle augmentée d'un silence précédant et succédant cette réponse impulsionnelle. Une partie du silence précédant la réponse impulsionnelle, de longueur égale à la durée

FIGURE I.25 – Boîte de dialogue permettant la mesure automatique de huit réponses impulsionnelles binaires grâce à une mesure asynchrone à l'aide de MLS, réalisée sous forme de module externe pour le programme CoolEdit Pro™

FIGURE I.26 – Boîte de dialogue permettant la mesure automatique de huit réponses impulsionnelles binaires grâce à une mesure asynchrone à l'aide d'un balayage fréquentiel, réalisée sous forme de module externe pour le programme CoolEdit Pro™

du balayage, est due à la convolution avec le balayage inverse, et peut donc être supprimée (cf. fin du § I.3.1.3). On obtient donc finalement la réponse impulsionnelle, augmentée de la durée de la résonance et du temps de calage (cf. fig. I.24).

Implantation et mise en œuvre Les deux procédures de mesure ont été implantées dans le logiciel CoolEdit Pro™ [179]. L'une et l'autre des deux techniques suppose que le décalage de fréquence d'échantillonnage ait été préalablement estimé. La boîte de dialogue concernant la mesure à l'aide de MLS est présentée figure I.25, et celle concernant la mesure à l'aide d'un balayage fréquentiel est présentée figure I.26. Pour une question de temps, il n'a pas été possible de mener à terme la procédure de mesure par balayage fréquentiel ; le processus s'arrête une fois que les séquences pertinentes ont été extraites du signal enregistré, puis rééchantillonné ; il manque l'étape de convolution par le balayage fréquentiel inverse. À la fin de la procédure, huit fichiers sont stockés sur disque ; leur nom est déterminé suivant les informations fournies dans la boîte de dialogue, en accord avec une norme en vigueur chez Renault. Comme les fichiers sont codés suivant le format WAV, ils sont composés d'échantillons entiers et normés.

TABLEAU I.1 – Retards d'arrivée de l'onde directe en échantillons, entre les quatre haut-parleurs et les deux oreilles d'un mannequin, comparés entre quatre méthodes de mesure

Haut-parleur/oreille	WG	WD	XG	XD	YG	YD	ZG	ZD
MLSSA	128	138	109	128	212	184	217	177
AURORA MLS synchrone	125	135	127	147	209	182	255	215
AURORA balayage synchrone	128	137	98	126	209	188	209	170
AURORA MLS asynchrone	128	134	109	129	209	180	214	177

Il faut donc déterminer l'amplitude maximale de l'ensemble des réponses impulsionnelles calculées, puis diviser toutes ces réponses par cette amplitude ; il en résulte un coefficient multiplicatif appelé facteur de cadrage : nous avons écrit ce facteur de cadrage à l'intérieur du fichier, après les données.

I.3.2.5 Validation

Pour valider notre protocole de mesure, nous avons effectué une série de mesures avec plusieurs dispositifs, et nous avons comparé les temps d'arrivée de l'onde directe entre les quatre haut-parleurs (WXYZ) du véhicule et les deux oreilles (GD) d'un mannequin placé en position conducteur dans un véhicule de type Safrane. Les dispositifs utilisés étaient les suivants :

- le logiciel MLSSA ;
- le logiciel AURORA avec la mesure par MLS ;
- le logiciel AURORA avec la mesure par balayage fréquentiel ;
- notre dispositif de mesure asynchrone par MLS.

Le système MLSSA nécessite une carte électronique propriétaire, alors que le système AURORA et notre dispositif ne nécessitent qu'une carte électronique standard dédiée au multimédia, mais fonctionnant en *full duplex* de manière parfaitement synchronisée : nous utilisons une carte de marque TURTLEBEACH, modèle FIJI.

Les résultats sont rassemblés dans le tableau I.1. Le système MLSSA faisant office d'étalon, on remarque que l'erreur commise par notre protocole sur le temps d'arrivée de l'onde directe est inférieure à 5 échantillons, soit un dixième de milliseconde environ. On peut conclure que notre protocole de mesure de réponses impulsionnelles par méthode MLS asynchrone est fiable.

D'autre part, les résultats obtenus avec AURORA présentent des différences notoires par rapport aux résultats obtenus avec MLSSA. Cela peut être du au mauvais fonctionnement du *full duplex* ou à une mauvaise synchronisation du logiciel CoolEdit Pro™: en effet, lors d'une mesure en boucle (c.-à-d. en injectant le signal de sortie de la carte dans l'entrée de celle-ci), nous avons constaté un retard de plusieurs échantillons.

Nous avons également comparé les réponses impulsionnelles obtenues avec les quatre systèmes. Comme la mesure asynchrone contient la réponse de l'amplificateur de l'autoradio, il faut la déconvoluer avant d'effectuer la comparaison. La figure I.27(a) présente les premiers échantillons des réponses obtenues avec ces quatre systèmes ; on constate que les premières réflexions sont correctement restituées. D'autre part, comme le montre la figure I.27(b), le rapport signal/bruit de la mesure effectuée par la méthode asynchrone est presque identique à la mesure de référence effectuée avec le système MLSSA.

Enfin, nous avons comparé les réponses en fréquences globales ; elles s'avèrent identiques (cf. fig. I.28).

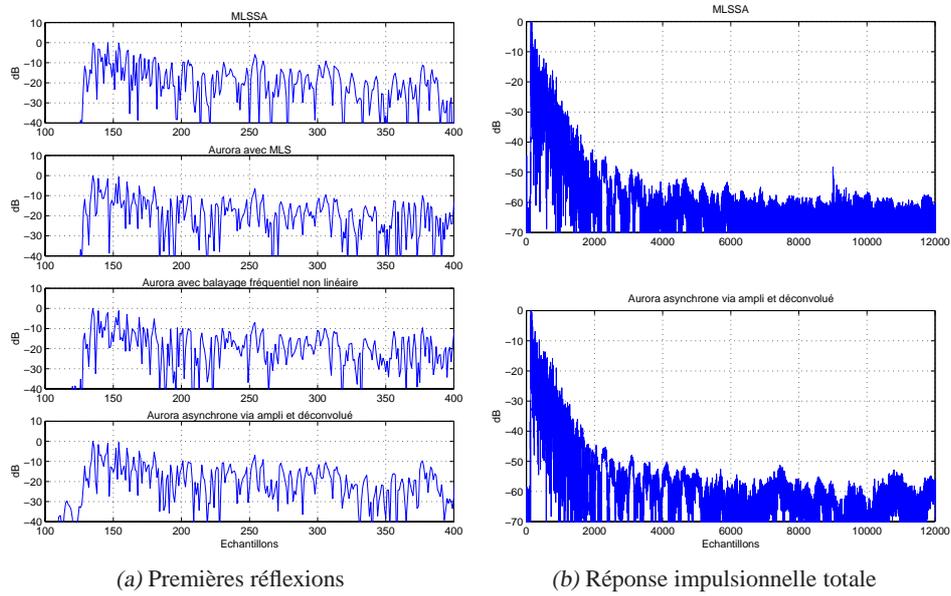


FIGURE I.27 – Comparaison du contenu temporel d'une réponse impulsionnelle obtenue avec plusieurs systèmes de mesure

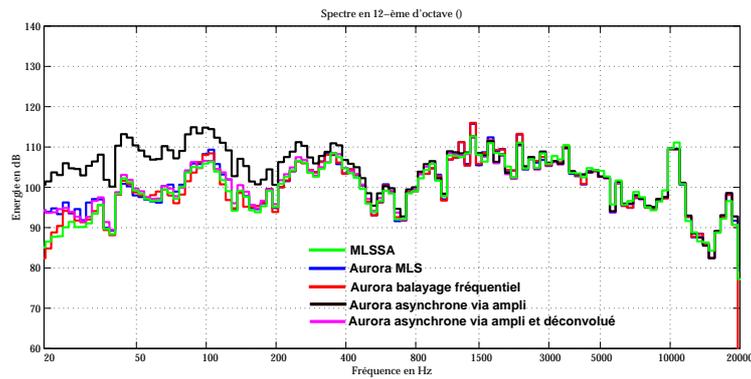


FIGURE I.28 – Comparaison des réponses en fréquence d'une réponse impulsionnelle obtenue avec plusieurs systèmes de mesure

Grâce à cette étude, nous disposons d'une méthode de mesure aynchrone des réponses impulsionnelles acoustiques de l'habitacle à l'aide de séquences de longueur maximale. La procédure est relativement simple à mettre en œuvre, puisqu'il suffit d'insérer un disque compact dans le lecteur de l'autoradio, de lancer la lecture et d'enregistrer le signal audio directement sur le disque dur d'un ordinateur. La seule contrainte préalable est de connaître avec précision la différence des fréquences d'échantillonnage des dispositifs de lecture et d'enregistrement. Tous les algorithmes ont été implantés sous forme de modules externes pour le logiciel CoolEdit Pro™; cette étape a fait l'objet d'un stage d'élève ingénieur [179]. Une fois le signal enregistré, le programme se charge du rééchantillonnage, du découpage des tranches de signal intéressantes, et du calcul de l'intercorrélacion avec le signal émis (correspondant à une copie informatique du signal présent sur le disque compact).

Nous avons pu comparer un ensemble des réponses calculées grâce à notre système avec le même ensemble de réponses mesurées avec le logiciel MLSSA. Tant au niveau du contenu temporel que du contenu spectral, les résultats concordent parfaitement.

Nous avons enfin pu montrer qu'il était possible d'envisager une mesure asynchrone à l'aide d'un balayage fréquentiel tout en récupérant les retards relatifs de chacun des haut-parleurs. Dans le cas d'une mesure synchrone, cette méthode se révèle même supérieure à la méthode des MLS (en terme de rapport signal/bruit et d'immunité aux non-linéarités) ; elle est même peu sensible à la différence des fréquences d'échantillonnage entre plusieurs dispositifs, à condition que cette différence ne soit pas trop importante, de sorte que le rééchantillonnage n'est pas nécessaire (mais la réponse en phase est dégradée).

I.4 Description de la réponse impulsionnelle

Les analyses perceptives démontrent que, dans le cas d'une salle de taille classique, on peut décomposer l'effet de salle en réflexions précoces et réverbération tardive (cf. § I.2). Nous présenterons dans un premier temps ces deux entités principales. Puis nous nous intéresserons à l'analyse temps-fréquence de réponses impulsionnelles.

I.4.1 Considérations générales

I.4.1.1 Premières réflexions

Les premières réflexions sont isolées les unes des autres et ne dépendent que des positions de la source et du récepteur. Leur influence sur notre perception de l'espace sonore peut être reliée objectivement à la forme de la salle et aux positions de la source et du récepteur dans la salle. L'impression de clarté est évaluée par l'importance du rapport de l'énergie des réflexions précoces à l'énergie de la réverbération tardive. D'autre part, la distribution temporelle des réflexions précoces et du son direct, observée dans le domaine fréquentiel, traduit une certaine « coloration » apportée par la salle [85, 137, 181]. Enfin le retard de la première réflexion par rapport au son direct peut être relié à la dimension subjective de la salle [85].

Elles jouent donc un rôle important sur le plan perceptif, et sont facilement décrites par la théorie géométrique. Leur date d'arrivée, leur amplitude et leur direction de provenance peuvent être calculées par des modèles géométriques tels que la méthode des rayons, des cônes, ou des images [138, 172].

I.4.1.2 Réverbération tardive

Les réflexions finissent par se mélanger d'une manière telle qu'il devient impossible de les différencier : on parle de *réverbération*. Le signal temporel correspondant peut être modélisé par un processus aléatoire gaussien dans le domaine temporel et spatial, et il est lié à la notion de champ diffus.

Champ diffus Lors de la décroissance du champ réverbéré, on considère que le champ est *diffus* à un endroit donné si la distribution directionnelle de l'énergie incidente est uniforme sur tout intervalle de

temps de durée courte par rapport au temps de réverbération [172] ; le champ sonore est alors isotrope. Cela nécessite une densité temporelle d'échos suffisante, et restreint donc l'application de la notion de champ diffus à la réverbération tardive. Une discussion intéressante sur la notion de perte d'individualité des particules sonores est effectuée par Kuttruff [172, p. 123].

La validité de l'hypothèse de champ diffus dépend de la géométrie de la salle et des caractéristiques acoustiques des parois. De manière générale, on peut montrer qu'un champ diffus est réalisé au bout d'un certain temps dans une salle aux parois parfaitement réfléchissantes et diffusantes (c.-à-d. renvoyant l'énergie incidente dans toutes les directions) [172, 257]. Pratiquement, dans la réponse impulsionnelle d'un canal acoustique, le champ diffus n'est atteint qu'approximativement, mais d'autant plus rapidement que les parois de la salle sont diffusantes.

Néanmoins, même si les frontières d'un espace clos ne produisent des réflexions que partiellement diffuses, leur contribution à l'établissement d'un champ sonore diffus peut être considérable, puisque pour chaque réflexion, l'énergie sonore spéculaire est convertie en énergie non spéculaire. Le rapport de l'énergie non spéculaire sur l'énergie totale est d'autant plus élevé que l'absorption des parois est faible [172, p. 120].

La notion de champ diffus permet de calculer simplement à chaque fréquence le temps de réverbération et le niveau de l'énergie réverbérée. La partie tardive d'une réponse impulsionnelle utilisable dans une opération d'auralisation peut donc être fournie par un processus aléatoire gaussien non stationnaire possédant les bonnes propriétés de décroissance exponentielle de l'énergie en fonction de la fréquence [72, 138, 212]. Le caractère gaussien provient de l'application du théorème de la limite centrale [277], puisque la densité temporelle d'échos est très importante (la superposition de nombreux phénomènes aléatoires indépendants et de même nature tend vers une loi normale lorsque le nombre de processus tend vers l'infini).

Temps de réverbération et courbe de décroissance La *temps de réverbération* d'une salle est le temps nécessaire pour que l'énergie reçue décroisse de 60 dB par rapport à son niveau en régime établi [131]. Sa définition stricte suppose que le signal émis est un bruit blanc, et on appelle *courbe de décroissance* la variation de l'énergie reçue en fonction du temps écoulé après arrêt de l'émission du bruit ; il s'agit en fait du carré de la réponse impulsionnelle $h^2(t)$, appelé échogramme, ou ETC (Energy Time Curve). Représentée sur une échelle logarithmique (cf. fig. I.29(a)), cette courbe a une allure exponentielle décroissante. La pente de la droite, sur laquelle on peut aisément mesurer un temps de réverbération, est indépendante de la position de la source et du récepteur : elle est uniquement représentative de la salle. Cependant, cette courbe de décroissance est perturbée par le caractère doublement aléatoire de la réverbération, à la fois dans le domaine temporel et dans le domaine spatial. La solution naturelle est d'effectuer une moyenne d'un grand nombre de courbes de décroissance. Mais Schroeder a montré qu'il existait un moyen de calculer directement la moyenne d'une infinité de réalisations [282] : la courbe de décroissance théorique, obtenue en moyennant une *infinité* de réalisations, s'obtient par intégration rétrograde de l'échogramme $h^2(t)$. Elle se déduit d'une seule mesure de la réponse impulsionnelle $h(t)$ par la relation

$$\text{EDC}_h(t) = \int_t^{+\infty} h^2(t) dt, \quad (\text{I.57})$$

avec $\text{EDC}_h(t)$ la courbe de décroissance énergétique intégrée, ou EDC (Energy Decay Curve). Cette courbe de décroissance est donc l'énergie résiduelle après l'instant t dans la réponse impulsionnelle $h(t)$.

On peut ainsi calculer une moyenne statistique au moyen d'une seule réalisation, au lieu de multiplier les mesures de la même expérience pour obtenir une moyenne plausible ; l'intégration permet de minimiser le caractère aléatoire de la réponse impulsionnelle en champ diffus.

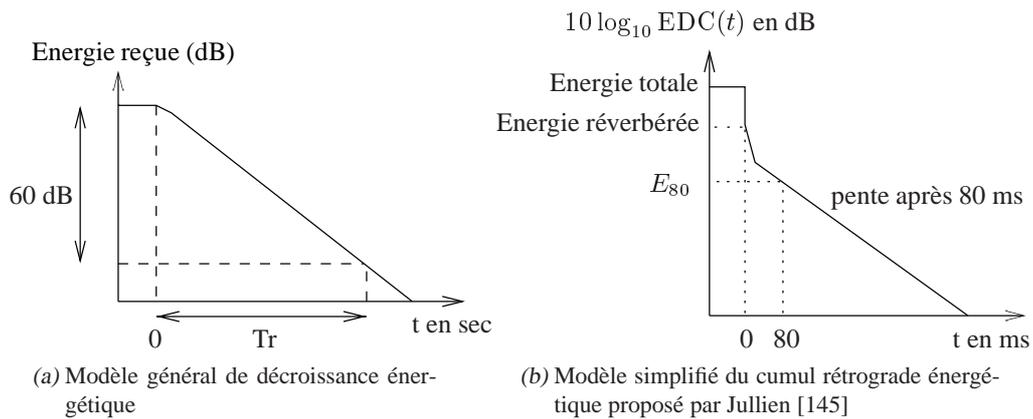


FIGURE I.29 – Courbe de décroissance énergétique

De plus, il est possible d'effectuer une moyenne spatiale de l'échogramme correspondant à différentes positions du récepteur avant d'effectuer le cumul rétrograde (il est également possible d'effectuer une moyenne spatiale de plusieurs cumuls rétrogrades) [155] ; cette opération accentue encore le lissage.

Dans le cas idéal où on effectue un calcul sur une version non bruitée de la réponse impulsionnelle, Jullien [145] propose un modèle simplifié de la courbe de décroissance intégrée. Sur la figure I.29(b), on peut constater que certains paramètres importants peuvent être déduits, comme l'énergie totale, l'énergie réverbérée, l'énergie restante après τ ms, et le temps de réverbération tardif après τ ms. La décroissance exponentielle débute souvent bien avant la limite de 80 ms, proposée par Jullien [145], qui est une simple convention.

Si le son direct et les toutes premières réflexions ont une amplitude importante, l'EDC chute brutalement. Cela se produit lorsque l'on se situe près de la source, et/ou que celle-ci est dirigée vers le récepteur et qu'elle est directive. Cette chute s'amplifie en hautes fréquences car la directivité de la source est en général plus grande en hautes fréquences qu'en basses fréquences ; de plus le son direct parvient de face au récepteur, par définition, tandis que la réverbération tardive, sous l'hypothèse de champ diffus, peut être assimilée à une infinité d'ondes planes provenant de toutes les directions possibles [314].

Comme on peut le voir sur la figure I.30, le cumul énergétique lisse considérablement la courbe représentant l'échogramme.

Mais en pratique, la courbe de décroissance intégrée est fortement affectée par le niveau de bruit de fond non négligeable présent sur toute mesure de la réponse impulsionnelle, comme le montre la figure I.30(b) ; la perte de dynamique est très importante. Il peut être mal aisé de déterminer un temps de réverbération par régression linéaire sur cette courbe, alors qu'on ne connaît même pas la dynamique maximale utilisable. De plus, la présence du bruit de fond perturbe profondément la convolution avec un signal ; d'après Xiang [330], le résultat fait apparaître un surplus subjectif de réverbération (il y a convolution du bruit de fond avec le signal). Pour remédier à ce problème, il faudrait connaître la « durée utile » de la réponse impulsionnelle ; or celle-ci varie avec la fréquence, puisque les fréquences élevées décroissent plus vite que les fréquences basses à cause des propriétés visco-thermiques de l'air [30]. Une procédure de restauration des réponses impulsionnelles mesurées a été proposée dans ce sens [113, 139, 310].

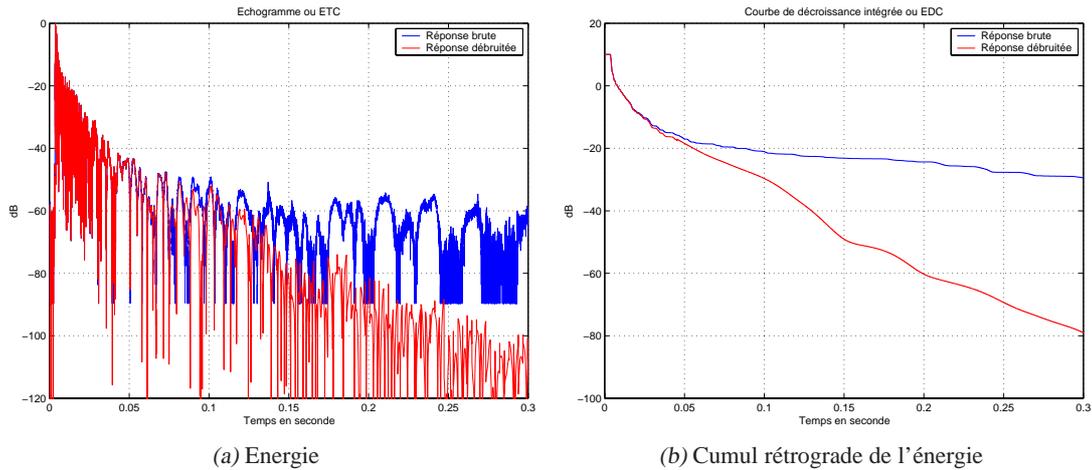


FIGURE I.30 – Echogramme et courbe de décroissance intégrée calculés à partir d'une réponse impulsionnelle mesurée dans un véhicule

I.4.2 Description temps-fréquence de la réverbération

On s'intéresse ici à la double description du phénomène de réverbération, ainsi qu'au modèle statistique qui lui est associé.

I.4.2.1 Analyses temps-fréquence

La description physique de la réverbération fait appel à une représentation conjointe temps-fréquence de la réponse impulsionnelle du canal acoustique, car toutes les grandeurs caractéristiques de la distribution énergétique dans la réponse, ainsi que le temps de réverbération, dépendent de la fréquence. L'idée la plus naturelle consiste à visualiser les courbes de variation fréquentielle des indices objectifs classiques tels que le temps de réverbération précoce, la clarté, la force sonore, etc. [22, 145, 181]. Mais cette solution ne constitue pas une représentation exhaustive, ni une représentation synthétique de l'effet de salle [138]. On est donc conduit à reprendre le problème à la base : il faut chercher une description globale du phénomène de réverbération, à la fois en temps et en fréquence. Nous appellerons $\rho_h(t, f)$ une distribution temps-fréquence en énergie.

Spectrogramme Un choix possible pour $\rho_h(t, f)$ est le spectre de Fourier à court terme (ou *spectrogramme*), défini comme le carré du module de la transformée de Fourier à court terme, (cf. § C.2) c.-à-d.

$$\text{SFCT}(i, f) = |\text{TFCT}(i, f)|^2 = \left| \int_{i \Delta t}^{(i+1) \Delta t} w(t - i \Delta t) h(t) e^{-2j\pi f t} dt \right|^2, \quad (\text{I.58})$$

où Δt représente la longueur de chaque trame i , et w la fenêtre de pondération (ou fenêtre d'analyse) [119, 239]. Ainsi, pour chaque fréquence, on peut calculer l'ensemble des indices choisis. La figure I.31 présente un exemple de représentation à trois dimensions du spectrogramme d'une réponse impulsionnelle non bruitée. Il s'agit d'une sorte de « déploiement fréquentiel » de l'échogramme. Pour des commodités d'affichage, les résolutions temporelles et fréquentielles ont été considérablement réduites grâce à un lissage avec fenêtres de Hann.

La lisibilité de cette représentation est cependant troublée par des fluctuations erratiques, telles que celles constatées sur l'échogramme (ou ETC) $h^2(t)$ au § I.4.1.2. La même solution est donc employée pour lisser la représentation : conservant l'approche de Schroeder [282], on va effectuer une sommation rétrograde de l'énergie à chaque fréquence [138].

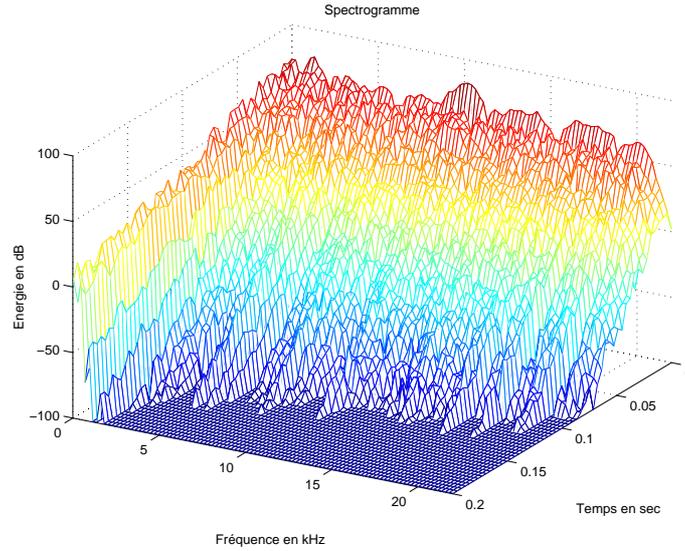


FIGURE I.31 – Spectrogramme d'une réponse impulsionnelle non bruitée analysée avec fenêtres de Blackman-Harris d'ordre 4 de durée 11,6 ms, et un pas d'avancement de 2,9 ms

Relief de décroissance L'idée intuitive est de disposer d'une représentation temps-fréquence telle que chaque évolution temporelle d'un canal fréquentiel puisse être considérée comme une courbe de décroissance énergétique (ou EDC, explicitée à l'équation I.57) évaluée localement au voisinage de la fréquence centrale du canal. On parle bien de *voisinage*, car une représentation temps-fréquence ne permet jamais d'isoler une fréquence particulière à cause du fenêtrage temporel qui se traduit par une convolution fréquentielle et un lissage de l'échantillonnage de la transformée de Fourier à temps discret.

Ainsi, si on remplace une réponse impulsionnelle $h(t)$ par une distribution temps-fréquence en énergie $\rho_h(t, f)$, on peut définir le relief de décroissance énergétique, ou EDR (Energy Decay Relief), par la quantité

$$\text{EDR}_{h,\rho}(t, f) = \int_t^{+\infty} \rho_h(\tau, f) d\tau. \quad (\text{I.59})$$

Dans le cas où la distribution temps-fréquence est la transformée de Fourier à court terme, on est tout naturellement conduit à une représentation obtenue par le calcul de spectres cumulatifs, en partant de la fin de la réponse impulsionnelle. Le relief de décroissance s'exprime donc par

$$\text{EDR}(i, f) = \sum_{k=i}^{k_{\max}} \text{SFCT}(k, f), \quad (\text{I.60})$$

avec la définition du spectre de Fourier à court terme donné à l'équation (I.58). La figure I.32 montre le résultat d'un cumul rétrograde sur le spectrogramme de la figure I.31. On voit nettement à quel point le cumul améliore la lisibilité de la représentation. Toujours pour des commodités de visualisation, les résolutions temporelles et fréquentielles ont été réduites dans les mêmes proportions et dans les mêmes conditions que pour la figure I.31.

L'interprétation d'une représentation telle que celle de la figure I.32 est facilitée par le modèle simplifié de courbe de décroissance proposé par Jullien (cf. fig. I.29(b)). Le relief de décroissance fournit donc, dans chaque bande de fréquences, l'énergie totale et l'énergie de la réverbération tardive. On peut en particulier en déduire, toujours dans chaque bande de fréquences, le temps de réverbération, et remonter au spectre de puissance initial tel que défini au § I.4.2.2.

De manière plus générale, le choix de la distribution ρ_h influe tout naturellement sur le relief de décroissance. Rien ne garantit que le spectrogramme fournisse un résultat correct. Polack [257] préfère

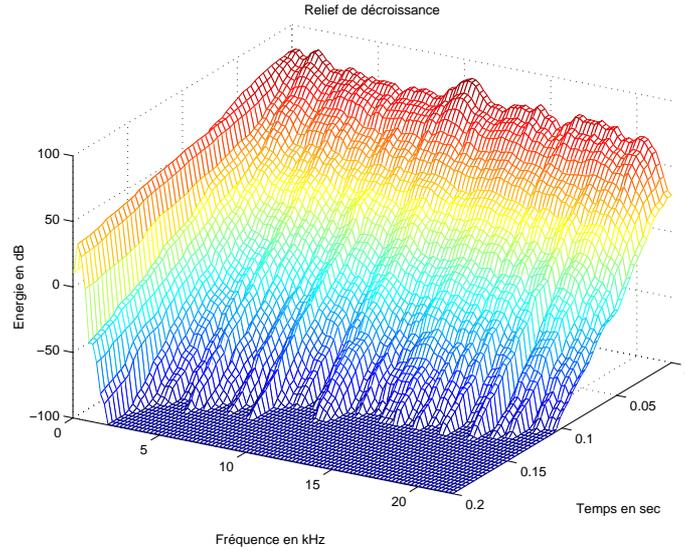


FIGURE I.32 – Relief de décroissance obtenu par cumul rétrograde sur le spectrogramme de la figure I.31

faire appel à la distribution de Wigner-Ville. Comme on le verra au § I.4.2.2 et au § I.4.2.3, la distribution temps-fréquence de Wigner-Ville a la bonne propriété de conserver les lois marginales temporelle et fréquentielle, alors que le spectre de Fourier à court terme ne les conserve pas. Il y a donc un problème important de *choix* de distribution temps-fréquence $\rho_h(t, f)$.

Jot [138] spécifie certaines propriétés que doit vérifier la représentation choisie. Les trois contraintes retenues pour le choix d'une distribution appartenant à la classe de Cohen [44, 86, 87] sont :

1. La conservation de la distribution marginale en fréquence, qui permet de retrouver la densité spectrale d'énergie $|H(f)|^2$ à partir du relief $\text{EDR}_{h,\rho}(t, f)$.
2. La préservation du support temporel des signaux causaux, qui garantit que la densité spectrale d'énergie $|H(f)|^2$ coïncide avec le temps 0 du relief, et non avec un temps négatif.
3. L'anticausalité de la distribution ρ_h , qui fournit une interprétation pour chaque temps τ du relief, comme densité spectrale de l'énergie résiduelle du signal après l'instant τ .

Ces trois contraintes se révèlent indispensables afin de pouvoir considérer chaque fréquence du relief de décroissance comme une courbe de décroissance intégrée au sens de Schroeder. Ainsi, chaque point du relief $\text{EDR}_{h,\rho}(t, f)$ représente l'énergie restant dans le signal à la fréquence f et à partir de l'instant t , sans contribution des valeurs passées. La courbe de décroissance intégrée telle que définie à l'équation (I.57) est donc fournie par la distribution marginale temporelle du relief, lui-même décrit par le *spectre courant futur* de Page-Levin [44], c.-à-d.

$$\text{EDR}_h(t, f) = \left| \int_{\tau=t}^{+\infty} h(\tau) e^{-2j\pi f\tau} d\tau \right|^2 ; \quad (\text{I.61})$$

il s'agit du carré du module de la transformée de Fourier du signal considéré à partir de l'instant t .

Calcul du relief de décroissance L'équation (I.61) permet de calculer directement le relief de décroissance *sans passer par le choix d'une représentation temps-fréquence*. On évalue simplement une transformée de Fourier discrète sur des fenêtres de longueur variable. La résolution temporelle d'un tel calcul est égale au pas d'avancement choisi pour t . La résolution fréquentielle est bien entendu inversement proportionnelle à la longueur de la fenêtre, et par conséquent maximale quand $t = 0$. Le fait que la résolution fréquentielle varie avec le temps n'est pas très pratique. On remédie à ce problème en prenant des fenêtres de tailles fixes et maximales (égales à la longueur de la réponse impulsionnelle), appliquées au signal préalablement tronqué [138]. La résolution fréquentielle est de ce fait toujours maximale.

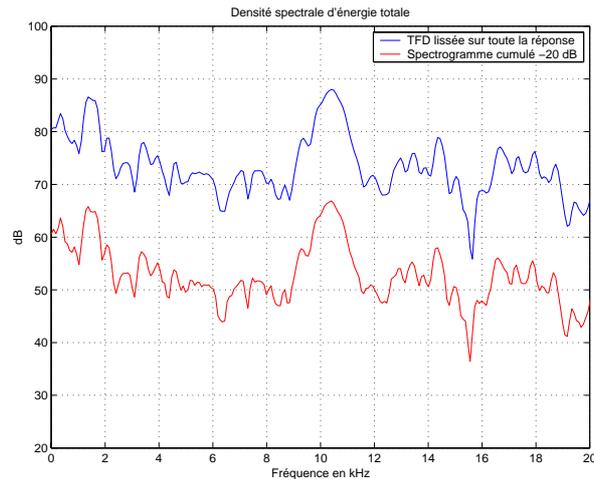


FIGURE I.33 – Densité spectrale d'énergie totale de la réponse impulsionnelle

En pratique, cette méthode de calcul du relief de décroissance peut être recommandée pour des signaux courts, mais devient beaucoup trop coûteuse pour des signaux longs (à cause des transformées de Fourier calculées sur une longueur égale à celle de la réponse impulsionnelle, soit plusieurs dizaines de milliers de points). Cependant, on peut réussir à exploiter les redondances dans le calcul, en cherchant à exprimer le relief à l'instant τ_2 en fonction de celui à l'instant τ_1 . D'après Jot [138], la récurrence exprimée rend l'algorithme sous-optimal si on ne cherche pas à atteindre des résolutions temporelles et fréquentielles maximales. Cette méthode est de toute façon inutilisable sur des signaux longs.

Le spectrogramme cumulé est bien plus avantageux en coût de calcul, mais il ne vérifie qu'approximativement les propriétés nécessaires évoquées au paragraphe précédent. Notamment, la non-conservation de la localité temporelle (anticausalité et support) entraîne une erreur sur les coupes à fréquence constante. Cependant, la non-conservation de la loi marginale en fréquence paraît plus acceptable [138] : la densité spectrale d'énergie totale calculée par cette méthode (c.-à-d. le premier temps du relief de décroissance) a la même allure que celle calculée avec une transformée de Fourier sur l'ensemble de la réponse, ramenée à la même résolution fréquentielle (cf. fig. I.33).

Le choix du spectre de Fourier à court terme en tant que distribution temps-fréquence est acceptable si on le calcule avec la « bonne » fenêtre de pondération (cf. § C.2). Comme on le remarque à la figure I.33, la fenêtre de Blackman-Harris d'ordre 4 [119, 239] semble bien adaptée. L'interprétation du relief de décroissance restera pertinente si on tient compte des résolutions temporelles et fréquentielles de la distribution choisie.

Une analyse approfondie de la qualité des différents estimateurs de la distribution temps-fréquence en énergie $\rho_h(t, f)$ est effectuée par Baskind [12].

On y apprend que l'écart type du relief de décroissance (tel que défini par le spectre courant futur) est de l'ordre de sa moyenne. Ceci explique les variations importantes observées lors de l'application directe de l'équation (I.61), et rend nécessaire l'utilisation d'un procédé de lissage, soit en utilisant une définition approchée comme celle du spectre cumulé, soit en effectuant une estimation de la moyenne plus robuste que l'observation d'une seule observation.

En effet, Baskind montre qu'en effectuant une moyenne des reliefs associés à des positions variables (mais suffisamment éloignées entre elles) du récepteur, pour une position fixe de la source, il est possible de réduire la variance de l'estimation d'un facteur égal à la racine carrée du nombre d'estimations ; ceci suppose bien entendu que le signal aléatoire associé à la réverbération tardive soit ergodique dans le domaine spatial, c.-à-d. que le signal capté à chaque position du récepteur puisse être vu comme une

nouvelle réalisation statistique du signal à une position donnée. Cette observation corrobore les études menées par Kawakami sur la courbe de décroissance énergétique globale [155].

D'autre part, si l'on suppose que l'énergie et le temps de réverbération sont à variation lente, il est légitime d'effectuer une moyenne de plusieurs isofréquences adjacentes ; en pratique, il faut augmenter la largeur de chaque bande au fur et à mesure que la fréquence augmente pour garder une précision constante de l'estimateur.

Il est même possible de cumuler les deux effets.

I.4.2.2 Analyses statistiques

Comme nous l'avons écrit au I.4.1.2, on peut modéliser la réverbération d'un canal acoustique par un signal aléatoire gaussien non stationnaire dont l'énergie décroît exponentiellement en fonction du temps. Il convient cependant de valider cette approximation par l'étude des propriétés statistiques de cette réponse impulsionnelle. Il y a deux approches possibles : l'une dans le domaine fréquentiel, et l'autre dans le domaine temporel. Historiquement, c'est l'approche fréquentielle qui a d'abord vu le jour, sous l'impulsion de Schroeder dans les années cinquante [287]. Puis Polack a complété cette approche dans le domaine temporel en proposant un modèle stochastique de la réponse impulsionnelle [257].

Statistique dans le domaine fréquentiel D'après Schroeder [287], les propriétés des fonctions de transfert de salle sont décrites *statistiquement* par la *densité fréquentielle* des modes propres de la salle. Kuttruff [172, p. 70] donne le nombre approximatif de modes propres inférieurs à une fréquence f dans une salle rectangulaire, ainsi que la densité de modes par Hertz, respectivement

$$N_m(f) \approx \frac{4\pi}{3} V \frac{f^3}{c^3} \quad \text{et} \quad D_m(f) = \frac{\partial N_f}{\partial f} \approx 4\pi V \frac{f^2}{c^3}, \quad (\text{I.62})$$

où V est le volume de la salle. Ces expressions approchées ne sont pas valables aux très basses fréquences. Un calcul rapide montre que le nombre de fréquences propres inférieures à 20 kHz dans une salle de 1000 m³ est d'environ 10⁹. Une conséquence de l'équation (I.62) est que l'écart moyen entre fréquences propres voisines est petit devant l'encombrement fréquentiel d'un mode. Celui-ci est mesuré par la largeur de bande d'un mode Δf , lui-même relié au coefficient d'amortissement δ du mode par la relation

$$\Delta f = \delta / \pi. \quad (\text{I.63})$$

Il est donc impossible de trouver un signal qui excite un mode en particulier. Mieux même, le signal capté par un micro, en régime sinusoïdal stationnaire, est la somme des contributions d'un grand nombre de modes propres. Dans ce cas, et en vertu du théorème de la limite centrale [35], la fonction de transfert complexe, recueillie lorsque la fréquence d'excitation varie, peut être assimilée à un processus aléatoire gaussien centré [172, 286, 287]. La partie réelle et la partie imaginaire sont alors des variables aléatoires gaussiennes centrées, non corrélées et de même variance. Le module de cette fonction de transfert suit une loi de Rayleigh [35], et la phase est uniformément répartie entre $-\pi$ et $+\pi$. L'écart moyen entre les maxima et les minima de la réponse en fréquence vaut environ 10 dB. Cette propriété est indépendante de la position de la source dans la salle, et aussi de la salle elle-même.

Le nombre moyen de pics par Hertz se révèle proportionnel au temps de réverbération [290], à savoir

$$D_f \approx \sqrt{3}/\delta \approx Tr/4, \quad (\text{I.64})$$

où δ est le coefficient d'absorption moyen de plusieurs modes de résonances ayant des fréquences voisines. D'autre part, on peut relier le temps de réverbération à ce coefficient d'absorption moyen par la relation

$$20 \log_{10}(e^{-\delta Tr}) = -60 \text{ dB} \quad \Rightarrow \quad Tr = 3 \ln 10 / \delta. \quad (\text{I.65})$$

Il demeure cependant capital de remarquer que cette théorie statistique ne se vérifie que lorsque le recouvrement modal est suffisant. Ceci n'est pas le cas en basses fréquences. Il existe une fréquence limite au-dessus de laquelle on doit se situer : c'est la *fréquence de Schroeder*. La première estimation de cette fréquence a été donnée en 1954 par Schroeder ; elle correspondait à la nécessité que l'espacement entre les modes soit inférieur au dixième de leur largeur de bande. Plus tard, de nombreuses mesures ont montré qu'au lieu de dix, un recouvrement (dans la limite de la largeur à -3 dB) de trois modes était suffisant pour que le modèle statistique puisse s'appliquer [67, 281] ; dans ce cas, la fréquence de Schroeder se calcule approximativement par

$$f_{\text{Schroeder}} \approx 2000 \sqrt{Tr/V} \quad \text{en Hz,} \quad (\text{I.66})$$

où Tr est le temps de réverbération en seconde et V le volume de la salle en m^3 . Pour une salle de 400 m^3 au temps de réverbération de 1 sec, la fréquence de Schroeder vaut environ 100 Hz. Elle est toujours très basse, sauf dans les petites salles relativement réverbérantes (salles de bains carrelées).

Statistique dans le domaine temporel De même que l'on peut calculer la densité de modes par Hertz, il est possible de calculer la densité de réflexions par unité de temps [172, p. 98] ; elle est donnée par

$$D_r(t) = \frac{\partial N_r}{\partial t} \approx 4\pi \frac{c^3 t^2}{V}. \quad (\text{I.67})$$

À l'inverse de l'équation (I.62), cette formule est valide pour une salle de forme quelconque.

La similitude auditive entre la réverbération tardive d'une salle et un bruit blanc exponentiellement décroissant est d'abord remarquée par Moorer [212]. Puis Polack décrit la réponse impulsionnelle comme une réalisation d'un processus aléatoire non stationnaire [257], c.-à-d.

$$h(t) = b(t) e^{-\delta t} \quad \text{pour } t \geq 0, \quad (\text{I.68})$$

où $b(t)$ est un bruit gaussien stationnaire centré, et δ est lié au temps de réverbération par la relation (I.65). Si on mesure plusieurs réponses impulsionnelles $h(t)$ à plusieurs endroits d'une même salle, on obtient plusieurs réalisations de ce même processus aléatoire. Il faut remarquer que le bruit $b(t)$ n'est pas nécessairement blanc ; on le caractérise par sa densité spectrale de puissance, définie comme la transformée de Fourier de sa fonction d'autocorrélation $R_{bb}(t)$ [35], c.-à-d.

$$P(f) = P_{bb}(f) = \int_{-\infty}^{+\infty} R_{bb}(t) e^{-2j\pi ft} dt \quad (\text{I.69})$$

Il se trouve que les variables t et f sont séparables pour le calcul des principales grandeurs caractéristiques [257]. Les grandeurs remarquables sont l'enveloppe énergétique de la réponse dans le domaine fréquentiel,

$$\mathcal{E}(f) = E[|H(f)|^2] \approx P(f) \int_0^{+\infty} e^{-2\delta t} dt = \frac{P(f)}{2\delta}, \quad (\text{I.70})$$

et l'enveloppe énergétique de la réponse dans le domaine temporel,

$$\mathcal{E}(t) = E[h^2(t)] = e^{-2\delta t} \int_{-\infty}^{+\infty} P(f) df. \quad (\text{I.71})$$

À ce stade, il faut remarquer que le temps de réverbération est toujours indépendant de la fréquence, ce qui n'est pas très réaliste.

MODÈLE AVEC TR DÉPENDANT DE LA FRÉQUENCE

Afin d'améliorer ce modèle, Polack fait appel au spectre moyen de Wigner-Ville, défini comme l'espérance de la distribution de Wigner-Ville appliquée à la réponse impulsionnelle $h(t)$. La distribution de Wigner-Ville, qui est une méthode de représentation temps-fréquence énergétique des signaux non stationnaires [39–41, 44, 86, 87, 126, 193], est définie par

$$W_h(t, f) = \int_{-\infty}^{+\infty} h(t - \tau/2)h(t + \tau/2) e^{-2j\pi f\tau} d\tau. \quad (\text{I.72})$$

C'est une fonction réelle répartissant l'énergie d'un signal dans le plan temps-fréquence, c.-à-d.

$$\int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} W_h(t, f) dt df = \int_{-\infty}^{+\infty} |h(t)|^2 dt. \quad (\text{I.73})$$

Une propriété remarquable de cette représentation temps-fréquence est la conservation des distributions marginales, à savoir

$$h^2(t) = \int_{-\infty}^{+\infty} W_h(t, f) df \quad \text{et} \quad |H(f)|^2 = \int_{-\infty}^{+\infty} W_h(t, f) dt \quad (\text{I.74})$$

Polack généralise le modèle décrit par l'équation (I.68) en imposant que $h(t)$ soit, pour tout instant t , une variable aléatoire gaussienne telle que le spectre moyen de Wigner-Ville de h vérifie

$$E[W_h(t, f)] = P(f) e^{-2\delta(f)t}. \quad (\text{I.75})$$

Cette expression définit une enveloppe temps-fréquence énergétique de h dont on suppose qu'elle varie lentement avec t et f [257]. Elle permet de généraliser, d'après (I.74), l'expression (I.70) de l'enveloppe énergétique de la réponse en fréquence $E[|H(f)|^2]$, même si δ devient une fonction de f . Mais il est impossible de généraliser celle de l'enveloppe énergétique de la réponse dans le domaine temporel (I.71), sauf à l'instant initial [138]. La densité spectrale de puissance du bruit stationnaire $b(t)$, notée $P(f)$, peut être appelée *spectre de puissance initial* [138] : c'est la densité spectrale de puissance de la réponse impulsionnelle en l'absence d'une décroissance exponentielle. Cette quantité jouera un rôle important dans la suite de l'étude.

DOMAINE DE VALIDITÉ

Le modèle statistique développé ci-dessus implique, par linéarité de la transformée de Fourier, que la réponse en fréquence de la salle est aussi un processus aléatoire gaussien complexe. Il ne s'applique donc que pour des fréquences supérieures à la *fréquence de Schroeder* définie par l'équation (I.66). De plus, elle ne peut être considérée comme aléatoire qu'au bout d'un certain temps, appelé *temps de mélange*. D'après Cremer et Müller, l'oreille ne peut distinguer de différences entre deux signaux aléatoires comportant plus de 2000 impulsions par seconde en dessous de 2 kHz [49, p. 485 et 422] ; d'après l'équation (I.67), ils en déduisent une valeur du temps de mélange $t_{\text{mélange}} = 2\sqrt{V}$. Mais Polack [257] en donne une définition plus subtile : il considère que la réponse impulsionnelle peut être considérée comme aléatoire au bout du temps à partir duquel au moins dix échos aléatoires se superposent à tout instant ; ceci permet d'approcher une distribution gaussienne en vertu du théorème de la limite centrale [277]. D'après l'équation (I.67), ce temps est donné par

$$t_{\text{mélange}} = \sqrt{\frac{10V}{4\pi c^3 \tau}}. \quad (\text{I.76})$$

Ce temps de mélange dépend de la largeur de bande de l'impulsion émise, ou de son étalement temporel, et de la durée d'intégration τ de l'observation. Une idée intuitive serait de considérer la durée d'intégration de l'oreille ; mais l'estimation de cette durée pose problème, car sa valeur dépend du protocole

expérimental, de la bande de fréquence et du niveau sonore [255, 319, 329]. Généralement, il est admis que le temps de mélange est approximé par

$$t_{\text{mélange}} \approx \sqrt{V} \quad \text{en ms,} \quad (\text{I.77})$$

où V est le volume de la salle en m^3 . Ce temps de mélange vaut environ 20 ms dans une salle de 400 m^3 . Il peut être considéré comme le temps nécessaire à l'établissement d'un champ « perceptivement diffus » dans la salle, c'est-à-dire au-delà duquel aucune direction prépondérante ne peut être perçue. En pratique, l'expression (I.77) correspond à une légère surestimation, si les parois de la salle ont des propriétés diffusantes [138].

I.4.2.3 Modèle stochastique temps-fréquence de la réverbération tardive

Nous avons décrit au § I.4.2.2 un modèle stochastique de la réverbération tardive où l'on considère que $h(t)$ est une réalisation d'un processus aléatoire non stationnaire. Pour Polack [257], ce processus est caractérisé par son *enveloppe* temps-fréquence $E[W_h(t, f)]$, comme indiqué à l'équation (I.75), de sorte que la distribution choisie $\rho_h(t, f)$ soit une réalisation de la représentation temps-fréquence $W_h(t, f)$. Le relief de décroissance est alors donné par l'équation (I.59) avec $\rho_h = W_h$.

Si on prend maintenant pour $\rho_h(t, f)$ la moyenne des représentations temps-fréquence $E[W_h(t, f)]$, on obtient comme expression du relief, d'après l'équation (I.75),

$$\text{EDR}_{h,\rho}(t, f) = \int_t^{+\infty} \rho_h(\tau, f) d\tau = \frac{P(f)}{2\delta(f)} e^{-2\delta(f)t}, \quad (\text{I.78})$$

avec $P(f)$ le spectre de puissance initial (équation (I.69)). C'est donc l'espérance du relief de décroissance (on a effectué une moyenne). On peut dire qu'une réalisation particulière l'approche. Et l'*enveloppe* temps-fréquence $\text{Env}_h(t, f)$ qui caractérise le phénomène de réverbération tardive est obtenue (à une constante près) par normalisation du relief par le temps de réverbération à toute fréquence, c.-à-d.

$$\text{Env}_h(t, f) = P(f) e^{-2\delta(f)t} = 2\delta(f) \text{EDR}_{h,\rho}(t, f) = \frac{6 \ln 10}{Tr(f)} \text{EDR}_{h,\rho}(t, f), \quad (\text{I.79})$$

avec l'expression du Tr défini à l'équation (I.65). Pour représenter l'allure de cette enveloppe temps-fréquence, on a donc besoin :

1. Du spectre de puissance initial $P(f)$
2. De la variation fréquentielle du temps de réverbération $Tr(f)$ qui fournit la pente $\delta(f)$ par la relation $\delta(f) = 3 \ln 10 / Tr(f)$ (cf. éq. (I.65)).

La figure I.34 a été tracée uniquement avec ces deux paramètres ; les lissages temporels et fréquentiels sont toujours identiques aux figures I.31 et I.32.

D'après l'équation précédente, le *spectre de puissance initial* est défini par

$$P(f) = 6 \ln 10 \frac{\text{EDR}_{h,\rho}(0, f)}{Tr(f)}, \quad (\text{I.80})$$

ce qui se vérifie aisément en reprenant la définition du temps de réverbération (indépendant de la fréquence), pour lequel on fait l'hypothèse que la décroissance énergétique est exponentielle, à savoir

$$10 \log_{10} \left(\frac{P(t)}{P_0} \right) = -60 \frac{t}{Tr} \quad \text{et} \quad \text{EDR}(0) = \int_0^{+\infty} h^2(t) dt = \int_0^{+\infty} P(t) dt. \quad (\text{I.81})$$

La valeur de $\text{EDR}_h(0, f)$ à prendre en compte dans l'équation (I.80) est l'ordonnée à l'origine de la droite de régression calculée sur une fréquence particulière du relief de décroissance de la figure I.35. La

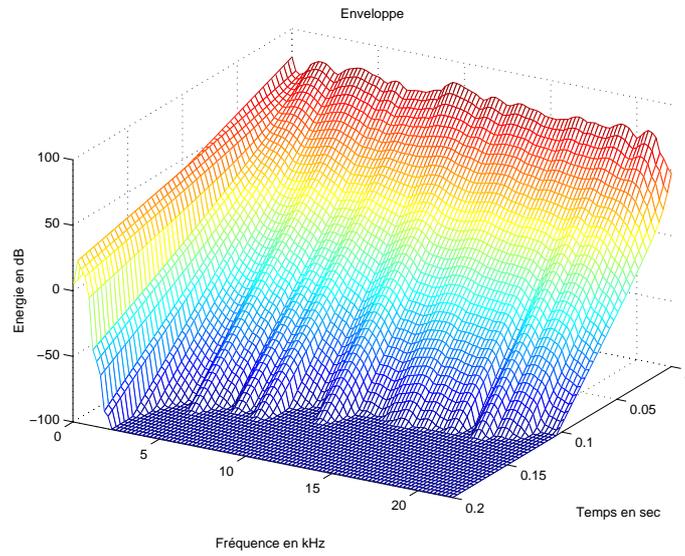


FIGURE I.34 – Enveloppe temps-fréquence caractérisant la réverbération tardive. Au temps 0 (instant d'émission de l'impulsion) on lit le spectre de puissance initial $P(f)$

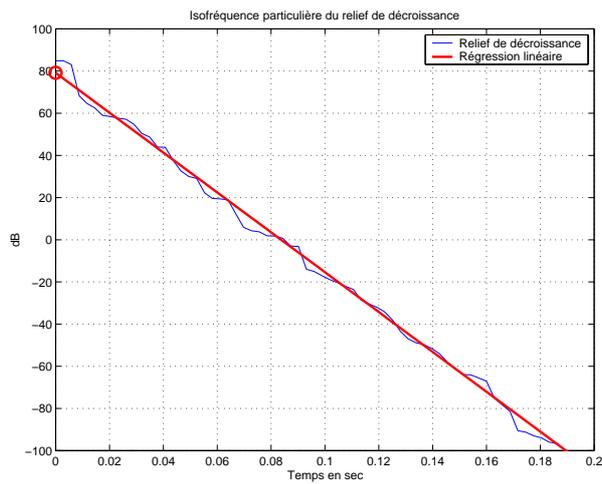


FIGURE I.35 – Énergie d'un canal particulier du relief de décroissance et régression linéaire pour trouver $EDR(0, f_i)$

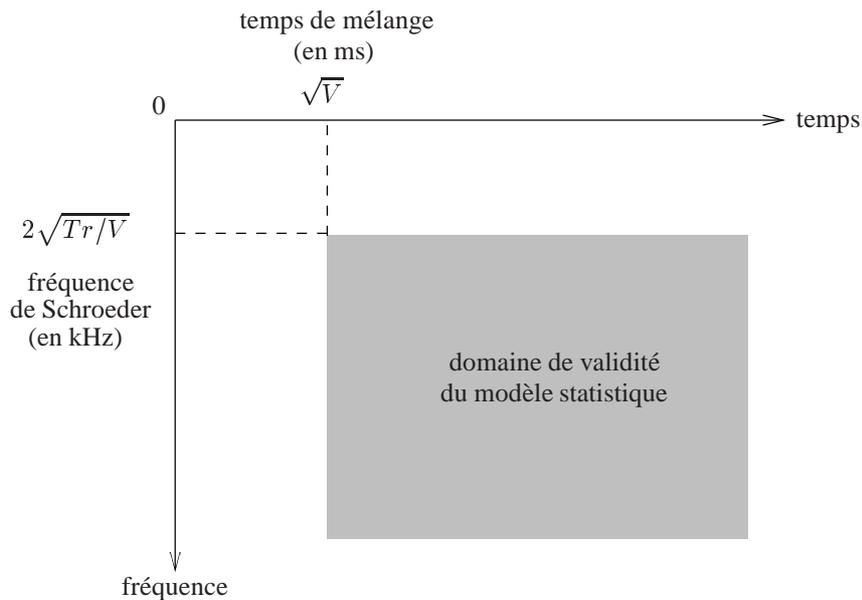


FIGURE I.36 – *Domaine de validité du modèle statistique dans le plan temps-fréquence : il faut se situer au-delà du temps de mélange et de la fréquence de Schroeder. V représente le volume de la salle, et Tr le temps de réverbération*

valeur de $P(f)$ donnée par l'équation (I.80) est donc le *spectre de puissance initial* de la réverbération tardive, à l'instant initial, c'est-à-dire au moment de l'émission de l'impulsion. Il s'agit du produit de deux fonctions de la fréquence : la directivité de la source moyennée suivant les directions d'émission, et celle du récepteur moyennée suivant les directions d'incidence.

L'enveloppe temps-fréquence peut donc être considérée comme un lissage d'une représentation temps-fréquence d'une réponse impulsionnelle. On peut également la considérer comme le gabarit d'un filtre variant dans le temps à utiliser pour générer une réverbération tardive. Il reste seulement à déterminer l'instant à partir duquel on appliquera ce gabarit, c'est-à-dire le début de la réverbération tardive. Celui-ci peut être aisément approximé dans le cas d'une modélisation géométrique de la salle, mais beaucoup moins dans le cas d'une réponse impulsionnelle mesurée. Cependant, on a dit que la réverbération tardive pouvait être assimilée à une réalisation d'un processus aléatoire gaussien non stationnaire (§ I.4.2.2) ; et on sait que le *facteur de crête* d'un tel processus vaut environ 10 dB. Il est défini par [35]

$$f_c(t) = \frac{h_{max}^2(t)}{E[h^2(t)]}. \quad (\text{I.82})$$

D'après [138], la puissance crête $h_{max}^2(t)$ décroît plus rapidement que la puissance moyenne $E[h^2(t)]$ au tout début de la réponse impulsionnelle, puis décroît de manière identique lorsque le facteur de crête atteint la valeur 10 dB correspondant à un signal gaussien. On sait que l'on a alors atteint la réverbération tardive.

CONCLUSION

Pour résumer, il existe deux quantités à connaître pour décrire statistiquement une salle : *la fréquence de Schroeder* et *le temps de mélange*. Dans une certaine région du plan temps-fréquence, délimitée par ces deux grandeurs, la réponse impulsionnelle d'un canal acoustique admet une représentation stochastique indépendante des positions de la source et du récepteur (voir figure I.36). D'une part, la densité d'échos doit être suffisante, ce qui explique la distinction entre réflexions précoces et réverbération tardive, et justifie la notion de temps de mélange. D'autre part, la densité de modes doit aussi être suffisante, ce qui

explique l'absence de référence à une fréquence propre en particulier, et justifie la notion de fréquence de Schroeder ; remarquons que, dans les salles de concert, cette fréquence peut être considérée comme suffisamment basse pour que l'on déclare valide l'approximation stochastique de la réverbération tardive sur tout le domaine des fréquences audibles.

Enfin, la réverbération tardive d'une salle est complètement définie par deux quantités qui varient lentement en fonction de la fréquence : *le temps de réverbération* et *le spectre de puissance initial*. Le temps de réverbération $Tr(f)$ caractérise la salle en elle-même ; il est relié à la forme et au volume de la salle, à l'absorption des parois et à celle de l'air. Le spectre de puissance initial $P(f)$ est communiqué aux modes propres de la salle ; il dépend de la directivité de la source et du récepteur, mais pas de leur position dans la salle.

L'extraction de ces paramètres n'est pas détaillé dans le présent document ; le lecteur se reportera à [139, 310]. L'estimateur choisi de la distribution temps-fréquence en énergie est le spectrogramme cumulé ; chaque isofréquence tracée en dB fait apparaître une décroissance linéaire suivie d'un palier correspondant au bruit de mesure. Le principe consiste à détecter, pour chaque fréquence, l'instant d'arrivée du bruit de fond, puis à effectuer une régression linéaire sur la partie valide. On en déduit directement le temps de réverbération, et, par prolongement de la régression à l'instant d'émission de l'impulsion, le spectre de puissance initial.

I.4.3 Comportement en basses fréquences

Il faut cependant souligner qu'en dessous de 500 Hz, l'analyse temps-fréquence est délicate car on ne rentre plus dans le cadre du formalisme sous-entendu [172] : la densité fréquentielle de modes est faible (à cause du volume réduit) et par conséquent les isofréquences du relief de décroissance sont parfois peu assimilables à des exponentielles décroissantes.

Tout espace clos peut alors être étudié grâce à l'analyse modale ; ce principe consiste à décomposer la fonction de Green sur les modes propres, ceux-ci correspondant aux solutions de l'équation de Helmholtz homogène associée. Pour une salle rectangulaire de dimensions $l_x \times l_y \times l_z$ comportant des parois rigides, les modes propres ont pour expression

$$\phi_{m,n,p}(x,y,z,f,t) = \cos\left(\frac{m\pi x}{l_x}\right) \cos\left(\frac{n\pi y}{l_y}\right) \cos\left(\frac{p\pi z}{l_z}\right) e^{2j\pi ft} \quad m,n,p \in \mathbb{N}, \quad (\text{I.83})$$

et les fréquences propres associées

$$f_{m,n,p} = \frac{c}{2} \sqrt{\left(\frac{m}{l_x}\right)^2 + \left(\frac{n}{l_y}\right)^2 + \left(\frac{p}{l_z}\right)^2}. \quad (\text{I.84})$$

La fonction de Green

$$G(\vec{r}, \vec{r}_0) = \sum_{m,n,p} \frac{\phi_{m,n,p}(\vec{r}_0)}{k_{m,n,p}^2 - k^2} \phi_{m,n,p}(\vec{r}), \quad (\text{I.85})$$

avec \vec{r}_0 la position de la source, \vec{r} la position courante et k le nombre d'onde, fournit donc une expression analytique qui, lorsque les fonctions propres sont connues, permet de calculer le champ sonore à tout instant et en tout point de l'espace.

En réalité, l'espace clos n'est jamais de forme simple, et les propriétés acoustiques des matériaux tapissant la surface intérieure ne sont pas uniformes et parfois même inconnues. Le calcul des fonctions propres devient impossible. Tout au plus, peut-on les exprimer sur une base de fonctions propres connues correspondant à une géométrie voisine mais plus simple : il s'agit du phénomène de couplage modal [30].

Le problème est rendu encore plus complexe dans le cas d'un habitacle de véhicule, puisqu'il existe des phénomènes de couplages acoustiques entre différents volumes [49, 172], principalement avec le

coffre et l'intérieur des portières ; en posant la main sur la paroi externe des portières tout en diffusant à l'intérieur du véhicule un signal comportant des basses fréquences, on ressent des vibrations.

Dans les petites salles, le domaine de fréquences inférieur à la fréquence de Schroeder est beaucoup plus étendu. Dans cette région, le champ sonore est gouverné par les modes propres ; même si il existe un recouvrement plus ou moins important, les modes sont bien séparés les uns des autres, et les propriétés spécifiques d'un habitacle transparaissent beaucoup plus facilement. La réponse de la salle dépend précisément des positions de la source et du récepteur, ainsi que du mode excité. La direction de propagation est fixée par ce mode, et les concepts de réflexions non spéculaires ou de diffraction n'ont pas de sens. Lorsque la fréquence varie, la distribution de probabilité de la réponse en fréquence varie fortement : c'est la différence fondamentale avec le comportement en hautes fréquences.

La distribution des fréquences propres est renseignée par l'indice d'espacement fréquentiel : Kuttruff [171] définit cet indice par

$$\psi = \frac{\frac{1}{N} \sum f_i^2}{\left(\frac{1}{N} \sum f_i\right)^2}, \quad (\text{I.86})$$

où f_i est la i -ème fréquence propre. Lorsque les fréquences propres sont régulièrement espacées, alors $\psi = 1$; si elles sont distribuées de manière aléatoire suivant une loi de Poisson, alors $\psi = 2$. Dans ce domaine de fréquences, on pourra voir apparaître de fortes variations dans la réponse en fréquence ; ces variations peuvent être mesurées par l'écart type en fonction de la fréquence. Comme le taux d'amortissement de chaque mode est différent, la décroissance sonore peut présenter de fortes variations fréquentielles dans les basses fréquences, puisque le recouvrement modal est faible et que la largeur fréquentielle d'un mode est proportionnelle à son amortissement. Seuls quelques modes peuvent être excités par un signal à bande étroite. Si ces modes ont tous à peu près le même taux d'amortissement, la décroissance logarithmique de l'énergie sera presque uniforme ; par contre, si certains modes ont un facteur d'amortissement faible et continuent à décroître alors que d'autres ont un facteur d'amortissement élevé et sont déjà éteints, la décroissance logarithmique de l'énergie présentera plusieurs pentes. Dans ce cas, on ne peut pas parler d'un unique temps de réverbération. Ce paramètre est donc peut pertinent dans le cas des petites salles en dessous de la fréquence de Schroeder.

D'après l'équation (I.84), on peut remarquer que le nombre de modes propres augmente rapidement avec la fréquence. Ceci a été confirmé par Kuttruff [172] dans un espace clos de volume quelconque. Or les modes propres sont à support fréquentiel fini : dû aux propriétés absorbantes des parois, leur largeur de bande est non nulle. Lorsque la fréquence augmente et que l'on s'approche de la fréquence de Schroeder, le recouvrement modal augmente. Mais il n'est pas encore suffisant pour que le champ sonore n'obéisse uniquement qu'à des considérations géométriques. La réponse en fréquence converge vers la réponse mesurée en hautes fréquences, mais il peut subsister des endroits où la densité de probabilité de la pression est très disparate ; dans cette région, la caractéristique la plus importante est d'avoir une distribution modale régulière afin d'optimiser le recouvrement modal.

À partir d'une certaine fréquence, il devient impossible de discerner les modes les uns des autres. La réponse en fréquence résulte alors d'une superposition inextricable de modes, indépendante de la position d'écoute (sauf au voisinage des parois où le son est amplifié). À ce stade, l'analyse modale ne permet plus d'extraire des informations sur la répartition du son dans la salle. Il faut avoir recours au modèle probabiliste précédemment décrit.

Lorsque l'on peut distinguer les différents modes, c.-à-d. en dessous de la fréquence de Schroeder, le caractère ondulatoire du champ sonore doit prévaloir. L'influence de l'impédance interne d'une source sonore sur sa diffusion dans l'espace doit être prise en compte, de même que l'influence de la charge acoustique apportée par l'environnement sur la puissance rayonnée par la source. Ces deux quantités (impédance de transfert acoustique et impédance de charge) sont utilisées par Salava [275] pour décrire le comportement des sources sonores en basses fréquences.

De manière générale, deux méthodes sont utilisées en basses fréquences : l'analyse par éléments finis, et l'analyse modale. On peut montrer que les valeurs des fréquences propres sont plus influencées par la forme de la cavité que par le revêtement acoustique des surfaces ; donc l'analyse modale permet de calculer la réponse acoustique de la cavité. D'autre part, on peut montrer qu'il est possible de calculer la réponse acoustique de la cavité sans calculer explicitement les modes de cavités, ce qui est du ressort de la modélisation par éléments finis. Les deux méthodes sont donc concurrentes [302]. Dans une berline de classe moyenne, le premier mode est aux alentours de 80 Hz, et le second aux environs de 150 Hz.

Des solutions passives ou actives ont été proposées pour atténuer l'influence de ces modes [122]. Il n'est pas nécessaire de connaître leur forme exacte, mais seulement leur fréquence de résonance et une estimation de leur facteur de qualité. Le contrôle passif est réalisé par le biais d'atténuateurs accordés et placés de manière appropriée ; ces atténuateurs sont en général de grandes membranes ou des résonateurs de Helmholtz comportant un orifice à perte. Il est nécessaire d'utiliser plusieurs résonateurs pour éviter un couplage avec les modes d'ordre supérieurs, ce qui produirait une répartition non homogène de la pression. Leur position doit être proche des maxima des modes à atténuer. Une autre méthode pour contrôler le champ acoustique en basses fréquences est l'utilisation d'un dispositif actif, dans lequel les haut-parleurs sont alimentés par des signaux mesurés par des microphones et préalablement filtrés ; le but recherché est d'avoir une distribution du champ sonore uniforme, c.-à-d. de limiter les différences relatives entre les modes. Le dispositif doit agir uniquement autour des résonances des modes à atténuer, et modifier aussi peu que possible les autres fréquences ; le but n'étant pas de minimiser l'amplitude des modes, mais de leur donner un taux d'amortissement identique en réduisant leur amplitude. Ces deux approches ont été testées sur des modèles réduits parallélépipédique, et ont permis d'atteindre une atténuation de 10 dB pour le premier mode.

Une égalisation en fréquence peut également être menée spécifiquement en basses fréquences [8, 69, 246], pour une position précise de l'auditeur. La principale difficulté provient des régions fréquentielles où la pression générée par les modes est faible : dans ce cas, il faut augmenter fortement le niveau du signal injecté dans les haut-parleurs. D'une part, l'effet produit peut être perceptivement néfaste, d'autre part des phénomènes de distorsion non linéaire peuvent apparaître au niveau des transducteurs.

Notons enfin que la prise en compte des phénomènes particuliers en basses fréquences est nécessaire pour réaliser une simulation de la propagation acoustique dans l'habitacle en vue d'une auralisation [81, 102]. Des modèles à éléments finis ou à éléments de frontières doivent être utilisés en basses fréquences en conjonction avec les modèles géométriques classiques en hautes fréquences (méthodes des rayons ou des cônes) [52].

I.5 Analyse temps-fréquence de réponses impulsionnelles d'habitacle

Dans le cas d'un habitacle de véhicule automobile, la fréquence de Schroeder, c.-à-d. la fréquence au-dessus de laquelle le modèle stochastique de la réverbération tardive est valide, est de l'ordre de quelques centaines de Hertz. En dessous de cette fréquence, le champ acoustique doit être décrit par la théorie classique de l'analyse modale.

Le temps de mélange, c.-à-d. le temps à partir duquel on considère que la densité temporelle d'échos est suffisante pour établir un champ acoustique *perceptivement* diffus, est de l'ordre de quelques millisecondes ; et l'examen de la décroissance énergétique des isofréquences du relief de décroissance montre que la décroissance exponentielle de l'énergie débute entre 5 et 10 ms après le son direct (cf. I.35). Dans le cas de réponses impulsionnelles d'habitacle, l'examen du facteur de crête n'est cependant pas un bon indicateur de l'arrivée du champ diffus (cf. eq. I.82) ; en effet, si la valeur de ce coefficient diminue effectivement dans les 50 premières millisecondes de la réponse impulsionnelle, elle n'atteint pas la valeur constante de 10 dB au-delà (le facteur de crête varie entre 5 et 10 dB). La raison est probablement que,

dans un habitacle de véhicule, le champ diffus n'est atteint qu'approximativement.

En effet, un champ réellement diffus ne peut pas exister dans les espaces clos pour lesquels les parois ont tendance à concentrer l'énergie réfléchie dans une direction prépondérante ; cependant, dans la réalité, des parois totalement diffusantes (c.-à-d. vérifiant la loi de Lambert [172, p. 111]) sont plutôt rares. De plus, lorsque les coefficients d'absorption des parois ne sont pas uniformes, certains chemins acoustiques seront absents, et cela empêche une bonne diffusion. C'est particulièrement le cas dans un habitacle automobile, puisque coexistent des parois à faible coefficient d'absorption comme les surfaces vitrées, et des parois à coefficient d'absorption élevé comme les sièges.

On préfère donc parler de champ diffus « au sens large », c.-à-d. que l'on suppose que le signal aléatoire associé à la réverbération tardive est ergodique dans le domaine spatial (ce qui veut dire que l'on peut substituer une moyenne spatiale à une moyenne statistique).

En deçà du temps de mélange, il faut théoriquement tenir compte des particularités des réflexions (date, direction de provenance et amplitude). Ces premières réflexions sont malheureusement difficilement identifiables sur des réponses impulsionnelles mesurées dans un habitacle automobile, tant l'échelle de temps est ramassée, et il arrive même que l'onde directe ne soit pas d'amplitude maximale ! Ceci peut être dû à l'orientation et à la directivité des haut-parleurs, ainsi qu'à des phénomènes d'interférences localisées en fréquence. Dans la suite de l'étude, nous ferons toujours la distinction entre :

1. Les 5 premières millisecondes de la réponse impulsionnelle, considérées dans leur globalité, que nous appellerons « son direct équivalent ».
2. Le reste de la réponse impulsionnelle, qui contient la réverbération, ou plutôt l'« effet d'habitacle » tant celle-ci décroît rapidement.

De plus, un simple examen de l'échogramme (cf. fig. I.30(a)) permet de constater que toute l'information utile se trouve présente dans les 50 premières millisecondes, c.-à-d. avant que le bruit de fond ne perturbe la mesure.

Bien que les explorations perceptives ont montré que plusieurs facteurs perceptifs font intervenir des limites temporelles allant jusqu'à 50 voire 80 ms [50, 150, 181], les indices objectifs classiques de l'acoustique des salles ne peuvent s'appliquer dans l'habitacle d'un véhicule. En effet, la décroissance exponentielle de l'énergie débute environ 5 ms après le son direct, et toute l'information a disparu au bout de 50 ms ! C'est pourquoi nous concentrons nos efforts sur l'analyse temps-fréquence des réponses impulsionnelles.

I.5.1 Protocole de mesure

Nous avons mené une campagne de mesure de réponses impulsionnelles dans un grand nombre de véhicules, avec différents capteurs.

Les véhicules utilisés font partie du parc de véhicules d'essai de la Direction de la Recherche (pour les véhicules de marque RENAULT) ou du Centre d'Analyse de la Concurrence (pour les véhicules d'autres marques). Le véhicule dont proviennent les courbes présentées dans la suite de ce chapitre est un modèle Safrane de série, équipé de haut-parleurs de marque SOFAITAL, à savoir

- à l'avant, de deux haut-parleurs de type *boomer* de diamètre 130 mm situés en portière, couplés à deux haut-parleurs de type *tweeter* situés en planche de bord ;
- à l'arrière, de deux haut-parleurs coaxiaux de diamètre 160 mm situés en portière.

Les capteurs utilisés sont

- un microphone omnidirectionnel (modèle 4165 1/2 pouce) de marque BRÜEL & KJÆR, relié à son préamplificateur (modèle 2639), puis à un amplificateur de mesure (modèle 2609) de la même marque ;
- un mannequin acoustique (modèle HMS II) de marque HEAD ACOUSTICS ;
- un microphone à directivité variable (modèle ST 250) de marque SOUNDFIELD (cf. § A.1).

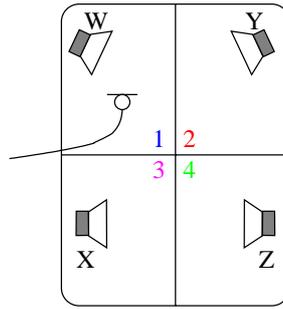


FIGURE I.37 – Repérage des haut-parleurs et des positions de mesure

Le microphone omnidirectionnel et le microphone à directivité variable ont toujours été placés à la position du centre de la tête du mannequin. La position du mannequin est référencée à l'aide de trois mesures de distances, ce qui permet une reproductibilité de cette position à l'identique d'une séance de mesure à une autre.

Le système de mesure est le logiciel MLSSA. Les réponses impulsionnelles sont mesurées à une fréquence d'échantillonnage de 44100 Hz ; on ne stocke que les 16384 premiers points, soit 371,5 ms, ce qui est largement suffisant, puisque le bruit de fond perturbe la mesure à partir de 100 ms environ.

La nomenclature est conforme à la figure I.37, à savoir que

- les haut-parleurs sont repérés par les lettres W, X, Y et Z ;
- la position du capteur est repérée par les chiffres 1, 2, 3 et 4 ;
- le type de capteur est repéré par les lettres
 - O pour le microphone omnidirectionnel ;
 - G et D pour les oreilles gauche et droite du mannequin acoustique ;
 - w, x, y, et z pour les différents canaux de sortie du microphone à directivité variable, correspondant à une directivité monopolaire pour w et une directivité dipolaire dans les trois directions principales de l'espace pour x, y, et z.

Les analyses temps-fréquence sont effectuées à l'aide du spectrogramme, avec

- une fenêtre de Blackman-Harris d'ordre 4 de longueur 512 échantillons, soit 11,6 ms, ce qui permet d'atteindre une résolution fréquentielle de 86 Hz ;
- un pas d'avancement des fenêtres de 128 échantillons, soit 2,9 ms.

I.5.2 Temps de réverbération

Le temps de réverbération correspond au temps qu'il faut pour que l'énergie atteigne le millionième de sa valeur initiale, soit une décroissance de 60 dB.

On constate rapidement que le temps de réverbération montre un comportement identique, que l'on soit dans l'habitacle d'un petit véhicule (tel une Clio) ou dans l'habitacle d'un grand véhicule (tel une Espace) (cf. fig. I.38, sur laquelle le trait épais correspond à la moyenne, et la zone grisée correspond à la plage de variation). Chaque graphique de la figure I.39 correspond à la mesure d'un haut-parleur donné, les quatre courbes représentant une position différente du microphone, selon le code de couleur établi sur la figure I.37. En basses fréquences, les valeurs sont de l'ordre de plusieurs centaines de millisecondes, mais elles varient dans des proportions importantes avec la fréquence, ce qui est certainement lié à la faible densité modale : tant que l'on se situe en dessous de la fréquence de Schroeder, ces valeurs ne correspondent pas véritablement à un temps de réverbération, mais plutôt à un taux d'amortissement modal. Au fur et à mesure que la fréquence augmente, le temps de réverbération diminue, pour atteindre une valeur constante voisine de 50 ms à partir de 1 kHz environ³. Cette valeur est extrêmement faible,

3. La valeur du temps de réverbération est bien réelle, et n'est en aucun cas une limitation de l'algorithme d'analyse temps-fréquence, qui est capable de calculer des temps de réverbération plus faibles.

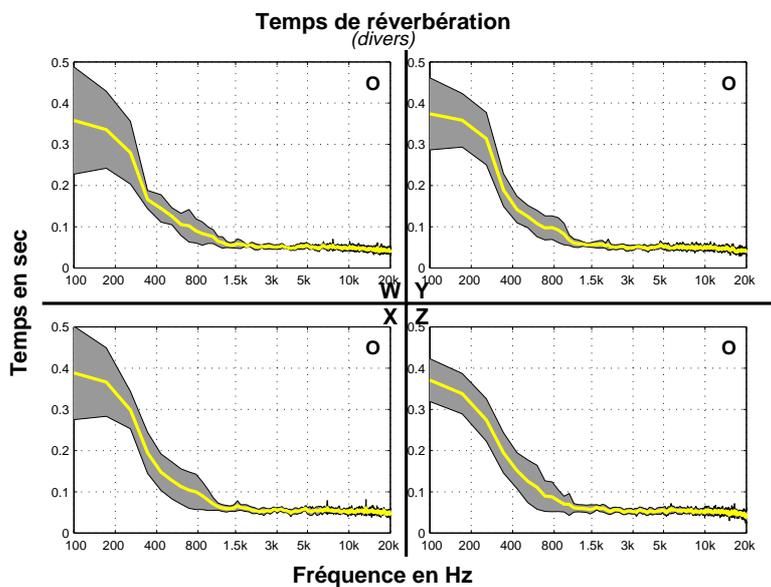


FIGURE I.38 – Temps de réverbération calculé à partir de réponses impulsionnelles mesurées avec un microphone omnidirectionnel placé en position conducteur dans dix véhicules (un graphique par haut-parleur)

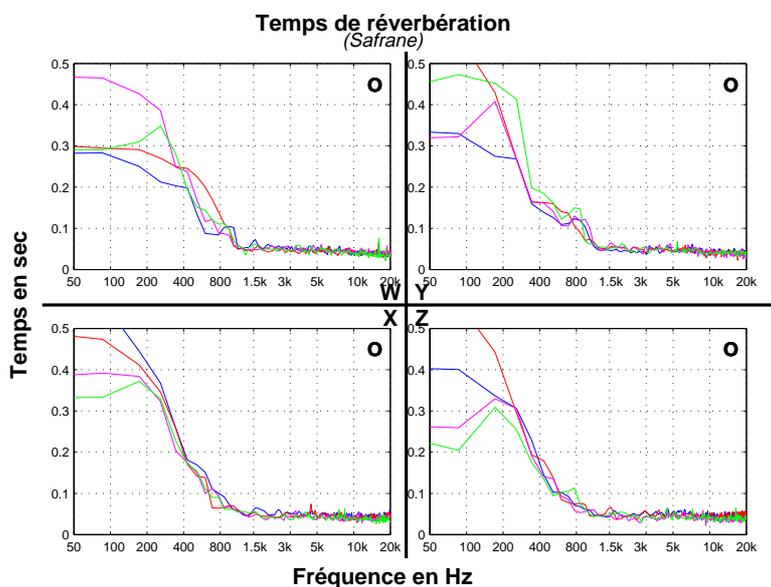


FIGURE I.39 – Temps de réverbération calculé à partir de réponses impulsionnelles mesurées avec un microphone omnidirectionnel placé à quatre positions différentes dans un véhicule de type Safrane (un graphique par haut-parleur)

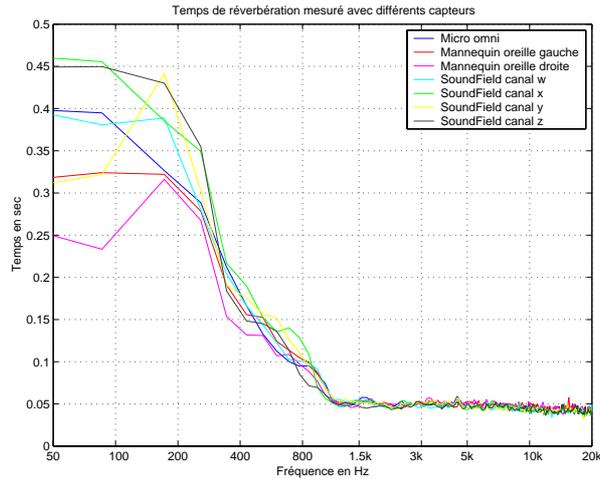


FIGURE I.40 – Temps de réverbération calculé à partir de mesures réalisées avec le microphone omnidirectionnel, le mannequin acoustique et le microphone SOUNDFIELD placés en position conducteur ; pour chaque capteur, nous avons effectué une moyenne des résultats obtenus avec les quatre haut-parleurs

au point que l'on puisse se demander si la réverbération possède une influence perceptive. Nous y reviendrons dans la première partie du chapitre 3, lorsque nous nous interrogerons sur la correction spectrale.

Nous vérifions que, à partir de 400 Hz environ, le temps de réverbération est indépendant de la position du récepteur dans l'habitacle. Nous décidons alors d'effectuer une moyenne spatiale du temps de réverbération, pour un haut-parleur donné et pour une position variable du microphone. Cette moyenne est licite car la réverbération tardive peut être décrite par un processus statistique ergodique dans le domaine temporel et spatial. De plus, nous constatons que le temps de réverbération est indépendant de la position du haut-parleur, toujours à partir de 400 Hz environ ; en-dessous, les variations peuvent être liées aux conditions différentes de montage des haut-parleurs, et à l'excitation des différentes parois de la portière. Nous effectuons également une moyenne spatiale entre les temps de réverbération calculés à partir des mesures effectuées avec les différents haut-parleurs. À l'aide de ce temps de réverbération moyen, nous effectuons une nouvelle analyse temps-fréquence, avec un temps de réverbération *imposé* pour chaque réponse impulsionnelle correspondante ; en effet, le calcul du temps de réverbération et de la densité spectrale de puissance initiale sont interdépendants (cf. éq. I.80). La pente de l'enveloppe de décroissance est donc imposée, il ne reste plus qu'à la caler sur la décroissance énergétique de chaque isofréquence du spectrogramme, et à calculer l'ordonnée à l'origine pour trouver la densité spectrale de puissance initiale. Le but de cette technique est de réduire les variations sur la densité spectrale de puissance initiale (principalement celles qui sont dues à un problème d'estimation).

On peut s'attendre à ce que le temps de réverbération soit identique quel que soit le microphone de mesure, à condition que celui-ci ne soit pas trop directif. Ceci est vérifié sur la figure I.40, qui représente le temps de réverbération calculé à partir de mesures effectuées avec les trois capteurs dont nous disposons (microphone omnidirectionnel, mannequin avec deux oreilles, et microphone SOUNDFIELD avec quatre directivités) ; les capteurs ont été positionnés de manière aussi proche que possible. Au-dessus de 400 Hz, on peut considérer que la concordance entre les courbes est satisfaisante. Le temps de réverbération calculé à l'aide des mesures réalisées avec les directivités dipolaires du microphone SOUNDFIELD n'est d'ailleurs pas si différent de celui calculé à l'aide de la mesure réalisée avec la directivité monopolaire, alors que leur diagramme de directivité est plutôt réduit.

En dessous de 400 Hz, les valeurs diffèrent ; on peut expliquer ceci par le fait que les capteurs ne sont pas tous strictement à la même position, et il est probable que nous subissions l'influence de deux

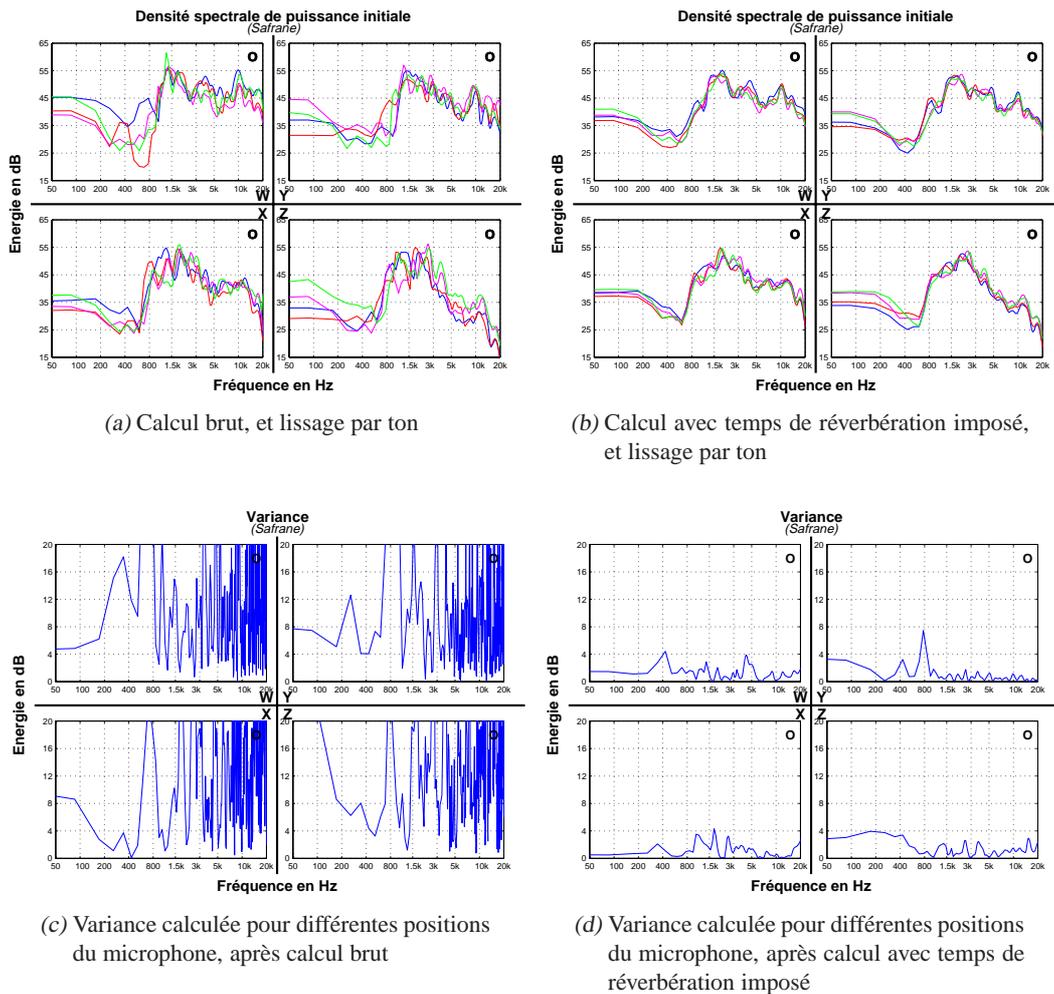


FIGURE I.41 – Densité spectrale de puissance initiale, et variance pour différentes positions du microphone, calculées à partir de réponses impulsionnelles mesurées avec le microphone omnidirectionnel dans un véhicule de type Safrane (un graphique par haut-parleur)

modes différents, donc de deux taux d’amortissement modal différents.

I.5.3 Densité spectrale de puissance initiale

Le spectre de puissance initial est la deuxième quantité qui caractérise la réverbération tardive. Il s’agit de la densité spectrale de puissance associée à la décroissance exponentielle, mais calculée pour l’instant d’émission de l’impulsion ; elle dépend de la directivité de la source et du récepteur, mais pas de leur position dans la salle.

Comme nous l’avons évoqué précédemment, nous décidons d’imposer le temps de réverbération dans l’habitacle d’après la moyenne spatiale des courbes obtenues avec les différentes positions des haut-parleurs et du microphone de mesure. Cela permet de réduire la variance de l’estimation de la densité spectrale de puissance dans des proportions importantes (cf. fig. I.41). On remarque sur la figure I.41(b) que les densités spectrales de puissance initiale correspondant aux deux haut-parleurs avant sont identiques, tout comme celles correspondant aux deux haut-parleurs arrières ; en revanche, les deux couples sont différents. Ceci est en accord avec la réalité physique, puisque les haut-parleurs avant et les haut-parleurs arrières sont différents, et la densité spectrale de puissance initiale dépend de la source.

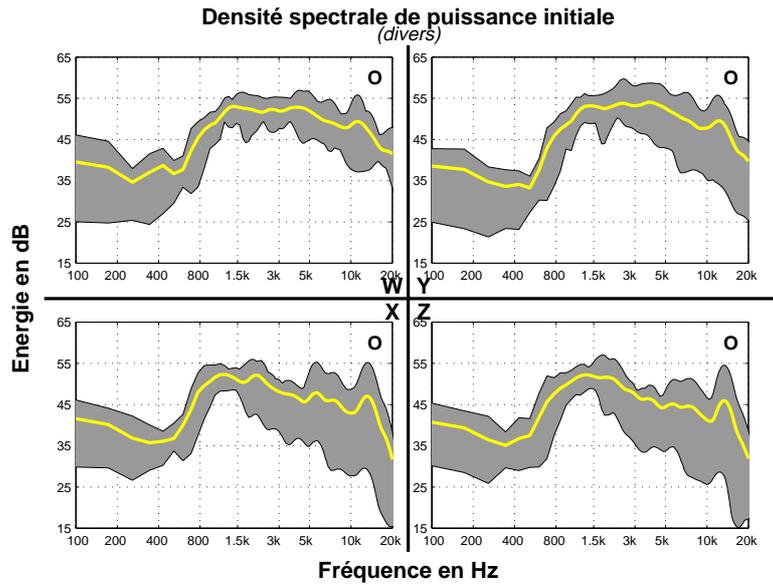


FIGURE I.42 – Densité spectrale de puissance initiale calculée (avec temps de réverbération imposé) à partir de réponses impulsionnelles mesurées entre le haut-parleur W et le microphone omnidirectionnel placé en position conducteur, dans dix véhicules (un graphique par haut-parleur)

Si le temps de réverbération ne présente pas de variations importantes d'un véhicule à l'autre, il n'en est pas de même pour la densité spectrale de puissance initiale (cf. fig. I.42). Les résultats obtenus à l'arrière présentent parfois un caractère passe-bas plus accentué, et cela est dû à la bande passante plus étroite des haut-parleurs installés à l'arrière des véhicules correspondant.

Nous avons également calculé la densité spectrale de puissance initiale à partir de mesures de réponses impulsionnelles effectuées avec les trois capteurs dont nous disposons, positionnés de manière aussi proche que possible (cf. fig. I.43). Pour que les résultats soient comparables entre eux, il faut les normaliser par la fonction de transfert en champ diffus du capteur considéré. Or,

- pour le mannequin acoustique, nous ne disposons que de 24 mesures de réponses impulsionnelles binaurales dans le plan horizontal, et de 24 mesures à 30° de site ;
- pour le microphone SOUNDFIELD, nous ne disposons que de 72 mesures dans le plan horizontal.

On constate que les densités spectrales de puissance ont globalement la même allure, mais des différences importantes existent en basses fréquences, entre 2 et 5 kHz, et au-delà de 10 kHz. Ceci peut s'expliquer par la mauvaise caractérisation de la fonction de transfert du capteur en champ diffus.

En l'état actuel des choses, la densité spectrale de puissance initiale ne peut donc être obtenue de manière fiable qu'avec le microphone omnidirectionnel, puisque c'est celui qui a le diagramme de directivité le plus homogène.

I.5.4 Décroissance temporelle de l'énergie

Remarquons d'abord que les variations de l'énergie totale sont importantes d'une place à l'autre dans le même véhicule, pour un haut-parleur donné. D'autre part, les variations sont importantes d'un véhicule à l'autre (cf. fig. I.44). La tendance passe-bas est fortement accentuée dans certains véhicules à l'arrière, et cela est dû à la bande passante plus faible des haut-parleurs installés.

Le temps de réverbération étant extrêmement faible, la « vitesse de décroissance » de l'énergie dans le plan temps-fréquence est très rapide, comme on peut le constater sur la figure I.45. À peine 5 ms après l'arrivée du son direct, le contenu fréquentiel de l'énergie est déjà modifié, et la décroissance temporelle

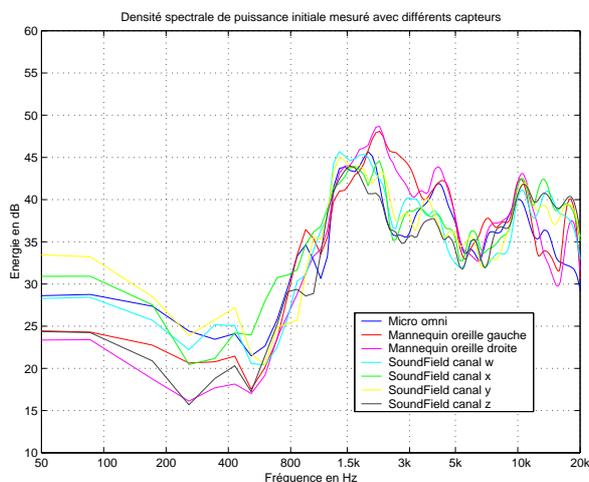


FIGURE I.43 – Densité spectrale de puissance initiale calculée (avec temps de réverbération imposé) à partir de mesures réalisées entre le haut-parleur W et le microphone omnidirectionnel, le mannequin acoustique et le microphone SOUNDFIELD placés en position conducteur

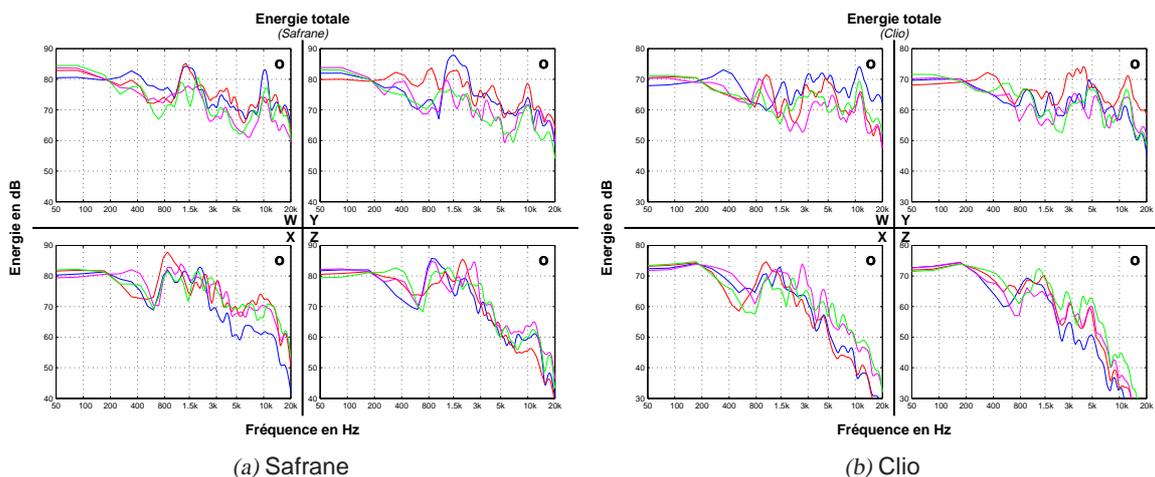


FIGURE I.44 – Énergie totale délivrée par les différents haut-parleurs, calculées à partir de mesures réalisées avec le microphone omnidirectionnel ; chaque courbe représentant l'énergie captée à une position différente dans le véhicule (un graphique par haut-parleur)

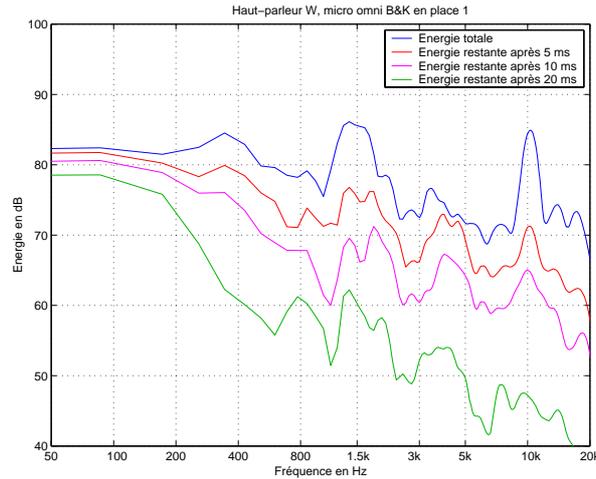


FIGURE I.45 – Évolution temporelle de la densité d'énergie restant à l'instant t dans une réponse impulsionnelle mesurée dans un véhicule de type Safrane avec le microphone omnidirectionnel placé en position conducteur

s'amplifie avec la fréquence, puis se stabilise (insignifiante à 100 Hz, elle est de 6 dB environ à partir de 800 Hz) ; ceci corrobore les observations effectuées sur le temps de réverbération, qui diminue avec la fréquence jusqu'à une valeur constante de 50 ms à partir de 800 Hz environ. Après 5 ms (soit 10 ms après l'arrivée du son direct), la chute est de 5 dB supplémentaire, et elle s'accroît encore 10 ms plus tard (soit 20 ms après l'arrivée du son direct). Cependant, l'allure générale de la courbe tend à se stabiliser (jusqu'à 8 kHz environ) ; ceci confirme que nous sommes rentrés dans le mode de décroissance exponentielle de l'énergie réverbérée.

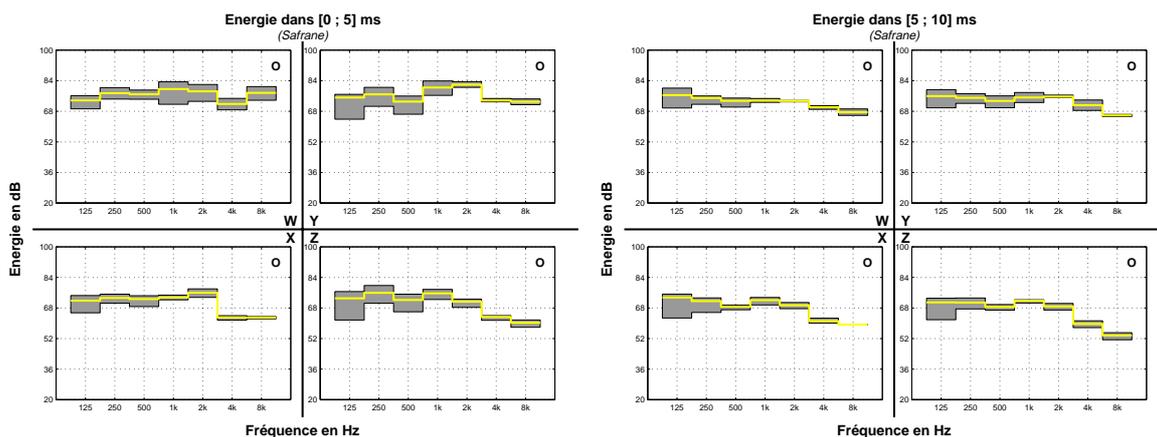
La décroissance énergétique étant aussi rapide, le contenu fréquentiel de l'énergie du son direct équivalent (c.-à-d. entre 0 et 5 ms) diffère assez peu du contenu fréquentiel de l'énergie totale. Cette observation est importante, et il serait intéressant de savoir si nous sommes perceptivement en mesure de distinguer indépendamment le contenu fréquentiel du son direct et celui de la réverbération, ou en d'autres termes la différence entre le contenu fréquentiel du son direct et celui de l'énergie totale. Pour effectuer une égalisation de la réponse en fréquence du système électroacoustique de l'habitacle, il est légitime de se demander si l'on doit égaliser indépendamment le son direct et la réverbération. Nous répondrons à cette question dans la première partie du chapitre 3.

I.5.5 Répartition spatiale de l'énergie

Grâce aux mesures effectuées avec le microphone SOUNDFIELD, nous pouvons explorer la répartition spatiale de l'énergie à l'intérieur de l'habitacle.

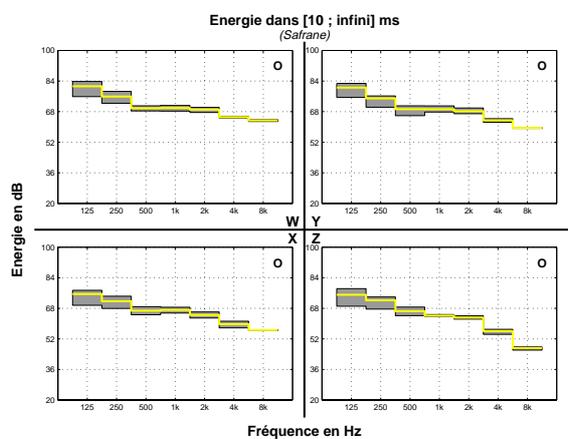
Intéressons-nous d'abord à l'évolution temporelle de la répartition spatiale, qui pourra nous renseigner sur le temps nécessaire à l'établissement du champ diffus. Examinons pour cela les différences entre les énergies captées par chacun des trois canaux directifs x , y et z . La figure I.46 représente l'énergie contenue dans une tranche temporelle donnée, mesurée par chacun des trois capteurs, et moyennée dans sept bandes d'octave ; la zone grisée schématise la plage de variation sur chacun des trois capteurs, et le trait épais représente la moyenne. On s'aperçoit que les différences d'énergie captées par les trois directivités sont de plus en plus faibles à partir de 500 Hz ; ceci veut dire que le champ sonore devient de plus en plus diffus. Les différences ont presque disparues à partir de 5 ms après le son direct.

Intéressons-nous ensuite à la distribution spatiale de l'énergie totale, qui nous permettra de cartographier le champ acoustique à l'intérieur du véhicule. Par combinaison linéaire des différents canaux



(a) Énergie contenue dans la tranche [0 ; 5] ms

(b) Énergie contenue dans la tranche [5 ; 10] ms



(c) Énergie restante après 10 ms

FIGURE I.46 – Énergie captée par les trois figures de directivité en huit ; le microphone est placé en position conducteur dans un véhicule de type Safrane (un graphique par haut-parleur)

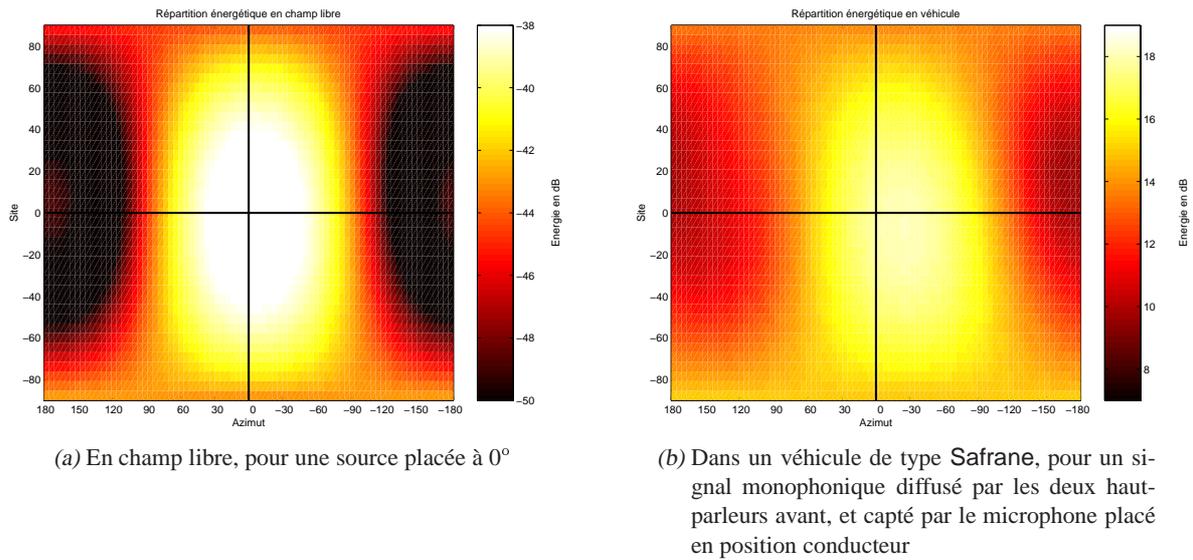


FIGURE I.47 – Répartition spatiale de l'énergie calculée à partir de mesures effectuées avec le microphone SOUNDFIELD

du microphone SOUNDFIELD, nous pouvons construire une autre figure de directivité. Par exemple, en sommant le canal w avec un des autres canaux, on obtient une cardioïde dirigée selon l'axe correspondant ; de plus, en pondérant les trois canaux directifs à l'aide de lois en sinus et cosinus, on peut orienter la figure en huit dans n'importe quelle direction de l'espace, selon la formule

$$w(t) + [(x(t) \times \cos \theta + y(t) \times \sin \theta) \times \cos \phi + z(t) \times \sin \phi], \quad (\text{I.87})$$

avec θ l'angle d'azimut et ϕ l'angle de site. Il est donc possible d'orienter la cardioïde, et d'explorer l'énergie dans toutes les directions de l'espace. On peut alors tracer une carte représentant la distribution spatiale de l'énergie. La figure I.47 montre une telle distribution, calculée avec la totalité de l'information temporelle présente dans chaque réponse impulsionnelle.

Cependant, nous devons remarquer que les figures de rayonnement de la combinaison des deux premières harmoniques sphériques diffèrent en champ proche et en champ lointain : alors que les figures de rayonnement d'un monopôle et d'un dipôle sont identiques quelle que soit la distance, il n'en est pas de même pour la figure de rayonnement d'une cardioïde, qui varie avec la fréquence et la distance [64]. En particulier, la figure de rayonnement associée à l'équation (I.87) n'est de forme cardioïde qu'en champ lointain ; en champ proche et en basses fréquences, elle tend à devenir omnidirectionnelle, ceci étant dû aux propriétés réactives du champ sonore [280] (la pression n'est plus en phase avec la vitesse particulière). C'est pourquoi nous avons effectué un filtrage passe-haut sur les mesures effectuées avec le microphone SOUNDFIELD en véhicule avant de tracer la figure I.47(b).

L'énergie maximale est située à l'azimut 25° et au site -8°. L'écart en azimut est important, et nécessite sans aucun doute l'utilisation d'un dispositif de correction de l'image sonore ; l'écart en site est moins important, ceci étant probablement dû à la position des tweeters en planche de bord.

Nous aimerions savoir si il est possible d'utiliser cette technique pour valider des traitements permettant de modifier l'image sonore. Nous avons donc refait le même genre de calcul après implantation d'une correction transaurale. Cette méthode, décrite au § III.3.3.1, permet de créer une ou plusieurs sources virtuelles à l'intérieur du véhicule tout en compensant l'effet d'habitable sur le début de la réponse ; elle est optimale en dessous de 6 kHz. Nous avons donc injecté à l'entrée du système un signal binaural correspondant à la HRIR à 0°, qui, après passage dans la matrice de transfert inverse et dans la

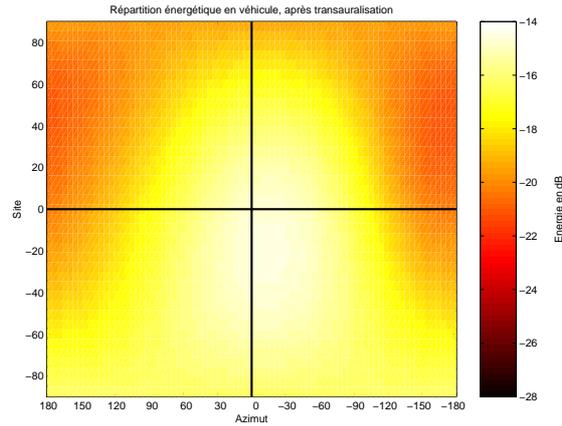


FIGURE I.48 – Répartition énergétique dans un véhicule de type Safrane, pour un signal binaural constitué de la HRIR à 0° injecté à l'entrée d'un système transaural

matrice de transfert directe, se trouve capté par le microphone SOUNDFIELD placé approximativement au centre de la tête du mannequin ; la figure I.48 présente le résultat obtenu, en limitant les calculs à 6 kHz. L'énergie maximale est située l'azimut 11° et au site -20° , alors qu'on s'attendrait à trouver 0° en azimut et en site.

Ce résultat décevant ne doit pas faire oublier qu'à la différence de l'holophonie, un système d'annulation des trajets croisés n'a pas pour ambition d'effectuer une reconstruction exacte d'un front d'onde, mais uniquement de restituer une certaine information sur les deux oreilles d'un individu situé à une position déterminée.

De plus, ce microphone semble fortement influencé par les effets de champ proche, qui ne sont pas négligeables lors d'une utilisation en véhicule, tant les distances entre les sources primaires (haut-parleurs) mais aussi les sources secondaires (sources images) sont faibles. D'après les investigations effectuées par Farina, et relatées à de nombreuses reprises sur une mailing liste dédiée au son 3D⁴, ce microphone est inutilisable comme capteur de prise de son dans les petits volumes, à l'intérieur desquels le champ sonore est très réactif : les relations mises en œuvre pour reconstituer les directivités dipolaires à partir des signaux des capsules ne sont valables qu'en champ lointain. L'information de localisation est donc difficilement exploitable ; de même son utilisation en vue d'une prise de son et d'une restitution à partir du format B (cf. annexe A.1) est inadaptée dans ce cas.

Compte-tenu de ces remarques, nous proposons d'abandonner ce microphone comme moyen de caractérisation spatiale du champ acoustique régnant dans un habitacle de véhicule automobile.

I.5.6 Corrélation interaurale

Le coefficient de corrélation interaurale, ou IACC (Inter Aural Cross Correlation) mesure l'inter-corrélation des signaux parvenant aux deux oreilles (cf. § III.4.1). Cet indice est basé sur la fonction de corrélation interaurale normalisée $\phi_{gd}(t)$ mesurée avec une tête artificielle, entre les temps d'intégration t_1 et t_2 , c.-à-d.

$$\phi_{gd}(\tau) = \frac{\int_{t_1}^{t_2} h_g(t)h_d(t + \tau) dt}{\sqrt{\int_{t_1}^{t_2} h_g^2(t) dt \int_{t_1}^{t_2} h_d^2(t) dt}}, \quad (\text{I.88})$$

4. <http://mail.music.vt.edu/mailman/listinfo/sursound>

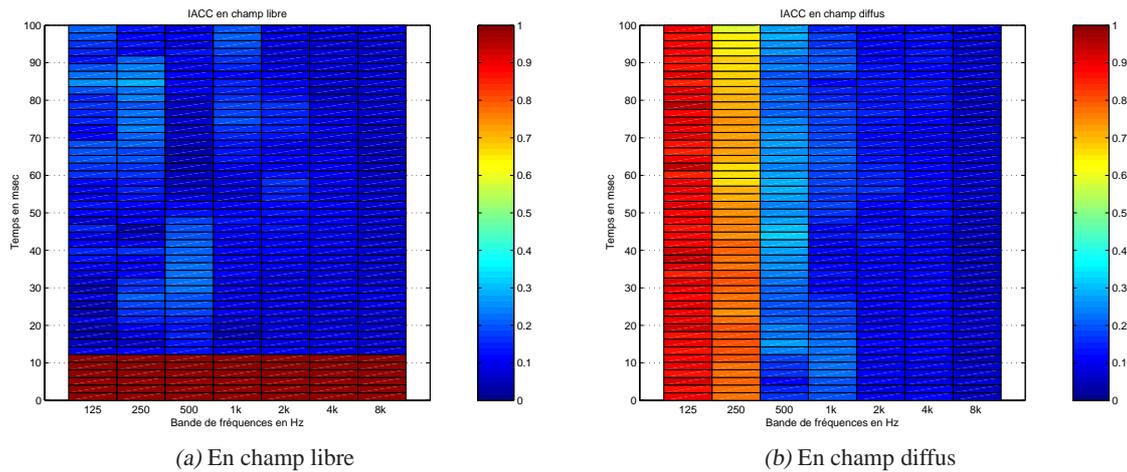


FIGURE I.49 – IACC à court terme calculé à partir de mesures réalisées en champ libre et en champ diffus

avec $h_g(t)$ et $h_d(t)$ les signaux de pression arrivant aux oreilles gauche et droite. La valeur de IACC correspond au maximum de cette fonction de corrélation, c.-à-d.

$$\text{IACC} = \max_{|\tau| \leq 1 \text{ ms}} |\phi_{gd}(\tau)|, \quad (\text{I.89})$$

avec τ le retard interaural ou ITD (Interaural Time Delay).

Nous avons calculé ce coefficient à « court terme », c.-à-d. sur des fenêtres de signal de 50 ms qui se recouvrent partiellement ; de cette façon, nous avons accès à l'évolution temporelle de ce coefficient, par ailleurs calculé dans sept bandes d'octaves de 125 Hz à 8000 Hz.

Intéressons-nous d'abord aux valeurs de ce coefficient mesuré avec le mannequin HMS II en champ libre et en champ diffus, avec une seule source. En champ libre (cf. fig. I.49(a)), la corrélation est maximale sur toutes les bandes de fréquence, sur un intervalle de temps correspondant à la durée de la HRIR. En champ diffus (cf. fig. I.49(b)), la corrélation est maximale en basses fréquences, puis diminue pour converger vers 0 à partir de 500 Hz.

On peut supposer une situation de reproduction idéale qui procurerait une sensation précise de localisation (respect du signal source diffusé), et aussi une certaine sensation d'espace ; l'IACC calculé dans ce cas correspondrait à une superposition des figures I.49(a) et I.49(b).

Regardons les valeurs de ce coefficient mesuré avec le même mannequin dans un véhicule de type Safranè. Pour un signal monophonique diffusé par les deux haut-parleurs avant, la figure I.50(a) ne fait pas apparaître une zone où l'IACC est élevé pour le son direct (d'où un manque de précision de l'image sonore), et cependant le champ tardif n'atteint pas tout à fait les conditions du champ diffus (d'où une faible sensation d'espace). Les résultats sont globalement identiques lorsque les quatre haut-parleurs sont alimentés en mono (cf. fig. I.50(b)).

La figure I.51 montre l'IACC et l'ITD calculés sur les 50 premières ms, pour un signal monophonique diffusé sur les deux haut-parleurs avant. Bien que l'IACC garde une valeur élevée pour les bandes d'octaves allant de 125 à 1000 Hz, le retard correspondant au maximum de la corrélation interaurale varie dans des proportions importantes avec la fréquence ; la précision de la localisation de la source se trouve gravement affectée, alors même que nous analysons les 50 premières ms des signaux arrivant sur les oreilles du mannequin.

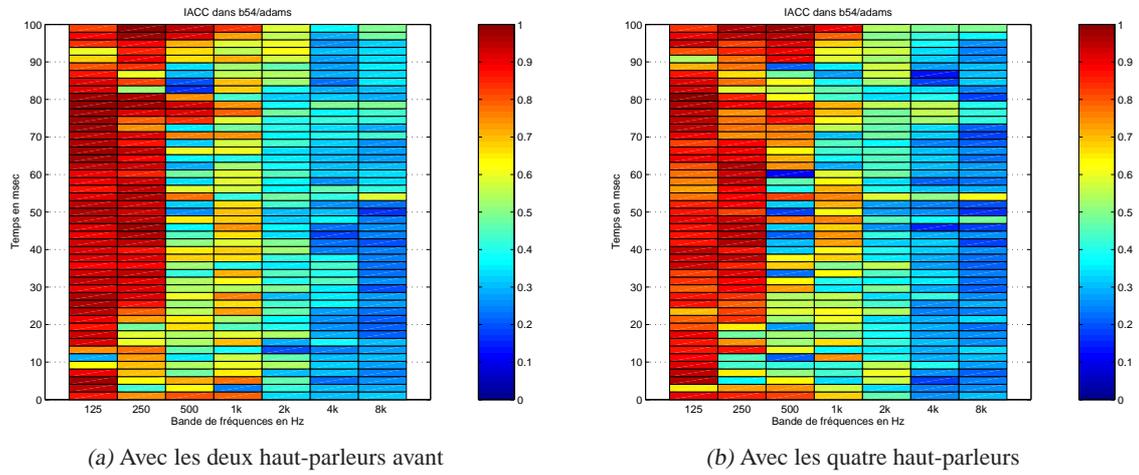


FIGURE I.50 – IACC à court terme calculé à partir de mesures réalisées dans un véhicule de type Safrane pour un signal monophonique, le mannequin étant placé à la position du conducteur

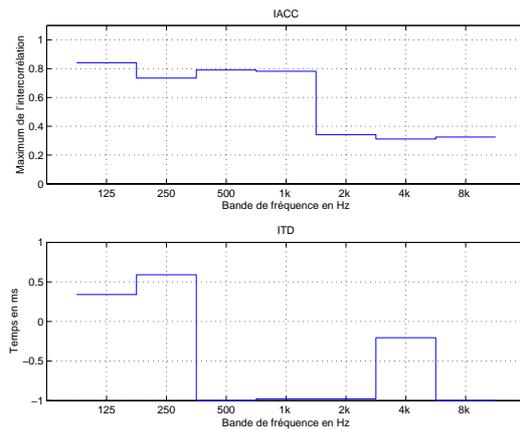


FIGURE I.51 – Corrélation interaurale et retard interaural calculés sur les 50 premières ms à partir de mesures réalisées dans un véhicule de type Safrane pour un signal monophonique diffusé sur les deux haut-parleurs avant, le mannequin étant placé à la position du conducteur

L'idéal étant de disposer d'un champ sonore permettant une localisation précise de la source sonore (comme en champ libre), et restituant une sensation importante d'espace (comme en champ diffus), il est nécessaire :

- d'une part, de stabiliser l'image sonore provenant des deux haut-parleurs avant ;
- d'autre part, de diminuer la corrélation provenant des haut-parleurs arrières.

I.6 Conclusion

La première partie de cette thèse était consacrée à la caractérisation objective de la chaîne de reproduction audio d'un véhicule.

Nous avons démontré la possibilité de réaliser une mesure asynchrone de réponse impulsionnelle, c.-à-d. avec un dispositif de lecture découplé du dispositif d'enregistrement. Pour cela, il est nécessaire de détecter avec précision la différence de fréquence d'échantillonnage entre les deux dispositifs, et de rééchantillonner le signal enregistré.

Ce système permet d'utiliser le lecteur de disque compact de l'autoradio. Il n'est donc pas nécessaire de démonter celui-ci pour accéder aux connexions électriques des haut-parleurs ; en outre, il devient possible d'inclure la fonction de transfert de l'autoradio dans la chaîne de mesure.

Nous avons réalisé une procédure de mesure entièrement automatique, qui permet de récupérer l'ensemble des réponses impulsionnelles de l'habitacle, ainsi que les retards de propagation relatifs entre les différents haut-parleurs.

Nous avons ensuite calculé un ensemble d'indices objectifs à partir des réponses impulsionnelles mesurées dans un grand nombre de véhicules avec trois capteurs différents : microphone omnidirectionnel, mannequin acoustique et microphone à directivité variable.

Nous avons mis en évidence une faible disparité du temps de réverbération entre les différents véhicules mesurés (environ 50 ms à partir de quelques centaines de Hz). En revanche, il n'en est pas de même pour la densité spectrale de puissance initiale qui varie à la fois avec le véhicule et la position du récepteur dans le véhicule.

La décroissance énergétique est très rapide, et son contenu spectral varie rapidement avec le temps : les différences de contenus spectraux entre son direct et le champ réverbéré sont importantes (de l'ordre de 6 à 10 dB), et pourront induire des effets de coloration à l'écoute ; de plus, le résultat varie avec la position du récepteur.

On peut également remarquer, grâce à l'analyse des réponses impulsionnelles effectuées avec le microphone à directivité variable, que la distribution directionnelle de l'énergie varie avec le temps, et devient rapidement homogène.

D'autre part, nous avons tenté de définir un protocole d'évaluation fiable de la répartition spatiale de l'énergie. La synthèse d'une directivité de type cardioïde à orientation variable à partir des différents canaux du microphone SOUNDFIELD ne répond pas à nos attentes, car ce microphone est fortement affecté par les effets de champ proche. L'utilisation du mannequin acoustique par contre est plus adaptée, car l'examen de la corrélation interaurale et du retard associé à son maximum nous renseignent à la fois sur la précision de la localisation d'une source sonore et sur la sensation d'espace. L'analyse des mesures brutes effectuées dans un véhicule montre, d'une part, que nous ne sommes pas en mesure de définir avec précision une direction de provenance d'une source sonore, et d'autre part, que le champ tardif ne devient jamais véritablement diffus.

Ces travaux de caractérisation permettent également de dégager des pistes concernant les traitements à mettre en place pour améliorer l'écoute dans les habitacles de voiture, à savoir :

1. Une correction spectrale.

2. Une amélioration de la précision de l'image sonore.
3. Une augmentation de la sensation d'espace.

CHAPITRE II

Restitution binaurale

II.1 Introduction

Ce deuxième chapitre décrit la recherche d'une méthode spécifique de restitution sur écouteurs du champ sonore régnant à l'intérieur d'un habitacle de véhicule automobile. Cette méthode se doit d'être optimale au sens où elle doit être la plus fidèle possible, et universelle au sens où elle doit pouvoir être utilisée de manière transparente par un grand nombre de sujets.

Elle repose sur le principe de la simulation binaurale, qui permet à un auditeur écoutant au casque de s'immerger virtuellement dans un champ acoustique donné [20]. Sous certaines conditions restrictives, le rendu acoustique peut être exact : le signal de pression arrivant aux tympans de l'individu et émis par le casque est identique à celui qui lui serait parvenu s'il s'était trouvé dans le champ acoustique réel.

Ce principe a connu une première application pour l'évaluation de la qualité perceptive des salles de concert [288]. Dans le secteur automobile, il permettrait de comparer instantanément plusieurs installations audio à l'aide d'un système de convolution temps-réel, et de se soustraire à la présence physique de l'habitacle.

Malheureusement, les conditions nécessaires à son exactitude sont réductrices, et empêchent toute universalité, puisqu'il faut mesurer certaines quantités propres à chaque sujet. Ceci n'est pas acceptable, car pour définir la qualité de la restitution HiFi à l'intérieur d'un grand nombre d'habitacles, il faut effectuer des tests d'écoute subjectifs en aveugle avec un grand nombre de sujets. Nous ne pouvons donc envisager d'utiliser tel quel le principe de la simulation binaurale pour effectuer ces tests. Il est important de généraliser ce principe, et de l'adapter afin qu'il reste valide pour le plus grand nombre possible d'individus.

Après un retour sur le principe général de la simulation binaurale, nous nous demanderons comment nous pourrions le généraliser. Pour cela, nous partirons de mesures réalisées sur une tête artificielle, et nous essaierons d'adapter ces mesures aux caractéristiques individuelles de chaque individu.

Comme nous voulons trouver le système d'écoute binaural qui soit également le plus proche possible de la sensation d'écoute réelle en habitacle, nous avons réalisé un test d'écoute subjectif en double aveugle qui nous a permis de comparer deux écoutes :

- celle du champ sonore régnant dans un habitacle automobile, et diffusé par les haut-parleurs de la voiture ;
- celle du même champ sonore, enregistré à l'intérieur des conduits auditifs, traité de différentes manières, et diffusé par un casque d'écoute (restitution binaurale).

Les résultats de ce test ont été analysés par des méthodes statistiques classiques, mais également par des méthodes multidimensionnelles. Ces analyses ont permis de dégager le meilleur traitement pour généra-

liser la restitution binaurale, tout en la gardant la plus proche possible de l'écoute réelle en habitacle.

Ce choix a en outre été confirmé par une analyse objective des différents traitements.

II.2 Simulation binaurale

II.2.1 Cadre de l'étude

L'évaluation de la chaîne audio d'un véhicule est traditionnellement réalisée par le biais de mesures objectives, telles que la réponse en fréquence ou le taux de distorsion, ainsi que par des tests d'écoute psychoacoustiques réalisées *in situ*. Le déroulement de ces tests peut être long et fastidieux si chaque sujet doit s'asseoir dans chaque voiture à tester ; cette manière de procéder exige des sujets qu'ils gardent la mémoire des différents systèmes [304]. De plus, il est difficile de prendre en compte le bruit de roulement du véhicule, car il faut disposer d'une piste d'essai ou d'un laboratoire spécialement équipé.

En employant une technique d'auralisation, ces difficultés disparaissent : les stimuli sonores utilisés pour les tests subjectifs ne sont pas enregistrés à l'intérieur de la voiture, mais reconstruits par convolution entre un signal source et des filtres spécifiques mesurés au préalable entre chaque source et chaque oreille. Ces stimuli reconstruits sont diffusés par l'intermédiaire d'un casque d'écoute. Le bruit de roulement peut également être enregistré par avance, et ajouté aux signaux convolués.

II.2.2 Formulation générale

La propagation des ondes sonores dans un espace clos peut être modélisée par un filtre linéaire invariant dans le temps, de sorte que le signal reçu en un point est régi par une équation de convolution entre le signal source et la réponse impulsionnelle du filtre.

Notre présence dans un espace clos modifie la propagation des ondes sonores à l'intérieur de cet espace. Plus précisément, le signal parvenant à nos tympans est profondément modifié par les multiples réflexions sur la tête, le pavillon et le conduit de l'oreille ; mais ce sont ces transformations qui nous permettent de localiser un son dans l'espace. En faisant l'hypothèse que ces transformations peuvent être modélisées par un filtre linéaire invariant dans le temps, on introduit la notion de réponse impulsionnelle de tête, ou HRIR (Head Related Impulse Response). La transformée de Fourier d'une HRIR est la fonction de transfert de tête ou HRTF (Head Related Transfer Function). En toute rigueur, il y a autant de filtres que de directions de provenance dans l'espace, et ils sont différents pour chaque individu ! Les réponses impulsionnelles sont mesurées par l'intermédiaire d'un microphone placé à l'intérieur du conduit auditif. Le sujet doit impérativement être placé en chambre anéchoïque afin de s'affranchir des réflexions parasites sur les parois du local. Au contraire, si on effectue ces mesures dans une salle, les filtres contiennent, outre la composante introduite par le corps, celle introduite par la salle ; on parle alors de réponse impulsionnelle binaurale ou BIR (Binaural Impulse Response).

Le but de la *simulation binaurale* [175, 176] est précisément de reproduire aux tympans d'un individu le signal de pression qui y serait capté si l'individu s'était trouvé dans un environnement acoustique donné. Le protocole se déroule en deux étapes (cf. fig. II.1) :

- la mesure des BIR de chaque source sonore ;
- la restitution sur casque d'écoute d'un signal source convolué par les BIR enregistrées.

À l'enregistrement, le signal Y_g capté par le microphone peut s'écrire (cf. fig. II.1(a))

$$Y_g = \text{BIR}_g \times X, \quad (\text{II.1})$$

avec

- | | |
|----------------|--|
| BIR_g | la fonction de transfert binaurale mesurée à l'oreille gauche du sujet, incluant l'effet du haut-parleur et du micro, et |
| X | la transformée de Fourier du signal source. |

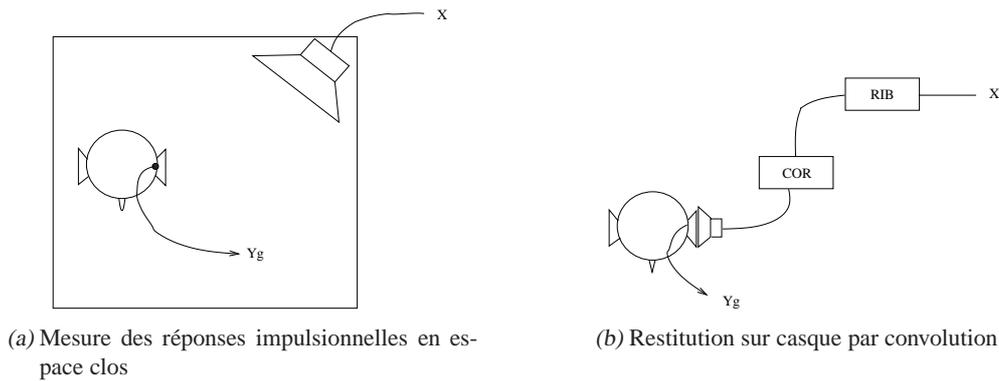


FIGURE II.1 – Principe de la simulation binaurale

À la restitution, le signal capté par le tympan se décompose en (cf. fig. II.1(b))

$$Z_g = \text{BIR}_g \times \text{COR}_g \times \text{CASQUE}_g \times X, \quad (\text{II.2})$$

avec

COR_g le filtre de post-traitement à définir, pour le canal gauche, et
 CASQUE_g la fonction de transfert de l'écouteur gauche du casque, mesurée dans l'oreille gauche du sujet pour la même position du microphone que lors de la mesure de BIR_g .

II.2.3 Différentes corrections possibles

Afin d'assurer une restitution aussi proche que possible de la réalité, nous devons chercher à éliminer l'influence de la chaîne de mesure (principalement amplificateur et microphone) et celle du casque d'écoute. Plusieurs solutions existent, avec différents degrés d'exactitude, suivant que les BIR ont été mesurées sur l'individu qui écoute, ou sur un autre individu.

II.2.3.1 À partir des BIR mesurées sur nos propres oreilles

Les BIR et les réponses du casque contiennent chacune l'influence de la chaîne de mesure : si les mesures ont été effectuées lors de la même séance, la contribution de l'amplificateur est identique dans les deux cas ; il en est de même pour la contribution des microphones, si toutefois on s'est assuré que leur position n'a pas été modifiée au cours de la séance. Afin d'éliminer globalement l'influence de la chaîne de mesure et du casque d'écoute, on peut déconvoluer les BIR par les réponses du casque. La simulation binaurale est donc **exacte** si l'on a $Y_g = Z_g$, condition qui définit le filtre d'égalisation que nous introduisons (cf. fig. II.2) :

$$\text{COR}_g = \frac{1}{\text{CASQUE}_g}. \quad (\text{II.3})$$

Malheureusement, cette normalisation n'est pas idéale dans le sens où elle lie irrémédiablement la banque de BIR mesurées à une série particulière de transducteurs électroacoustiques. Pour que la restitution soit fidèle, le casque à utiliser lors de la restitution est donc **imposé**.

Une autre procédure de normalisation permet d'éliminer la contribution de la chaîne de mesure. Elle a l'avantage de laisser ouvert le choix du casque d'écoute [175]. L'équation (II.2) devient, pour une restitution idéale,

$$Y_g = \text{BIR}_g \times \text{COR}_{\text{BIR}} \times \text{COR}_{\text{CASQUE}} \times \text{CASQUE}_g \times X. \quad (\text{II.4})$$

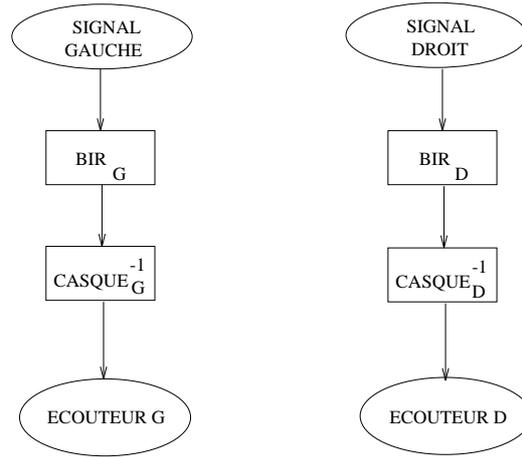


FIGURE II.2 – Schématisation des traitements à effectuer sur les réponses impulsionnelles binaurales en vue d'une restitution exacte au casque

Il s'agit de normaliser les BIR par un filtre de référence FLT, obtenu par combinaisons de mesures contenant l'effet de la chaîne de mesure, soit

$$\text{COR}_{\text{BIR}} = \frac{1}{\text{FLT}_g}. \quad (\text{II.5})$$

Une égalisation de même type est requise pour les BIR et les casques d'écoute, soit

$$\text{COR}_{\text{CASQUE}} = \frac{\text{FLT}_g}{\text{CASQUE}_g}. \quad (\text{II.6})$$

Certains casques sont « égalisés » par construction [141, 207, 305], de telle sorte que leur fonction de transfert vaut FLT. Si l'on suppose que ce filtre rassemble certaines caractéristiques individuelles importantes, la compensation des microphones de mesure est effectuée à l'aide du filtre COR_{BIR} . Aucun filtre $\text{COR}_{\text{CASQUE}}$ n'est alors nécessaire avec ces casques. Un post-traitement des BIR à l'aide du filtre COR_{BIR} est même superflu si l'on effectue les mesures avec une tête artificielle égalisée, elle aussi, par construction.

II.2.3.2 À partir des BIR mesurées sur un autre individu

Pour des raisons pratiques, il est difficilement envisageable de réaliser un enregistrement par auditeur, notamment si on veut réaliser un test d'écoute avec un grand nombre de sujets. Or, la reproduction binaurale n'est fidèle que si chaque auditeur utilise les BIR enregistrées sur ses propres oreilles [209, 317]. C'est pourquoi nous cherchons à effectuer une adaptation individuelle pour chaque auditeur. Le but de cette adaptation est de corriger, à l'aide d'un filtre propre à chaque individu, les BIR mesurées sur un individu de référence. Ce filtre, noté $\text{COR}_{\text{individu}}$, permet donc de faire écouter à un individu les BIR mesurées sur un autre individu (ici un mannequin acoustique) ; il est de la forme [175]

$$\text{COR}_{\text{ind}} = \frac{\text{FLT}_{\text{individu}}}{\text{FLT}_{\text{mannequin}}}. \quad (\text{II.7})$$

On espère faire en sorte que la reproduction soit la plus fidèle possible, c.-à-d. que le signal arrivant à l'entrée du conduit auditif de l'individu soit aussi proche que possible de celui qu'il aurait capté si il avait lui-même effectué la mesure, soit

$$\text{BIR}_{\text{individu}} \approx \text{BIR}_{\text{mannequin}} \times \frac{\text{FLT}_{\text{individu}}}{\text{FLT}_{\text{mannequin}}}. \quad (\text{II.8})$$

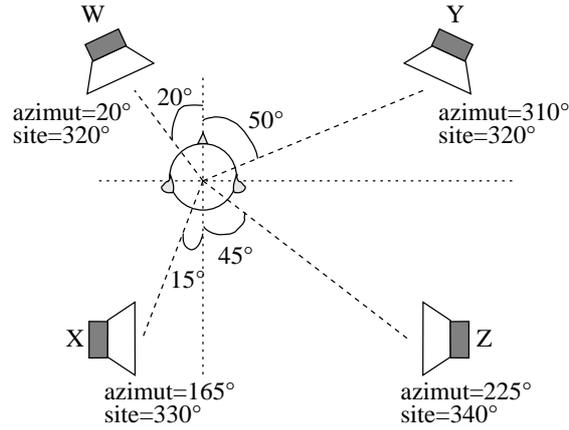


FIGURE II.3 – Disposition dans l'habitacle du système audio étudié

En remplaçant l'équation (II.8) dans l'équation (II.2), pour une restitution idéale, on obtient

$$Y_{\text{individu}} \approx \frac{BIR_{\text{mannequin}}}{FLT_{\text{mannequin}}} \times \frac{FLT_{\text{individu}}}{CASQUE_{\text{individu}}} \times CASQUE_{\text{individu}} \times X. \quad (\text{II.9})$$

On suppose que les filtres FLT s'éliminent, pour enlever au mieux les différences entre individus. Les filtres de correction COR_{BIR} et COR_{CASQUE} de l'équation (II.4) peuvent donc se récrire

$$\begin{aligned} COR_{BIR} &= \frac{1}{FLT_{\text{mannequin}}} \\ COR_{CASQUE} &= \frac{FLT_{\text{individu}}}{CASQUE_{\text{individu}}} \end{aligned} \quad (\text{II.10})$$

En général, on ne connaît pas la distribution directionnelle du champ acoustique régnant dans l'enceinte de mesure. On teste donc plusieurs hypothèses concernant les directions d'incidences des ondes planes arrivant aux oreilles de l'auditeur. Le filtre FLT se trouve ainsi défini comme une combinaison linéaire de HRTF mesurées pour un ensemble de directions d'incidence. Nous décidons d'explorer trois choix :

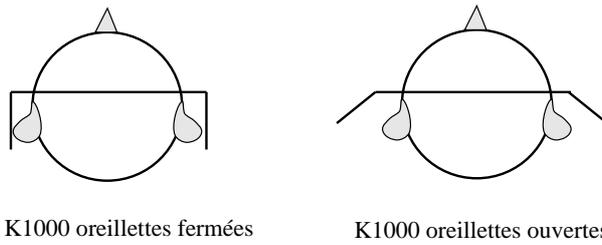
1. Une correction *champ libre*, obtenue pour une seule direction d'incidence (on fait ici l'hypothèse très restrictive que la direction d'incidence directe est prépondérante).
2. Une correction *champ pondéré*, obtenue pour un ensemble réduit de directions d'incidence (on privilégie certaines directions, par exemple un cône autour de l'incidence directe).
3. Une correction *champ diffus*, obtenue pour un grand nombre de directions d'incidence (on fait alors l'hypothèse que le champ acoustique est une somme d'ondes planes non corrélées provenant d'incidences uniformément réparties).

Dans l'habitacle automobile utilisé, le son est diffusé par quatre haut-parleurs placés dans les portières avant et arrière, comme schématisé figure II.3. La correction individuelle globale peut donc aussi être obtenue en superposant les contributions de chacun des haut-parleurs vers chaque oreille, soit huit corrections différentes [175],

$$\begin{pmatrix} Y_g \\ Y_d \end{pmatrix} = \left[\begin{pmatrix} BIR_{Wg} & BIR_{Yg} \\ BIR_{Wd} & BIR_{Yd} \end{pmatrix} + \begin{pmatrix} BIR_{Xg} & BIR_{Zg} \\ BIR_{Xd} & BIR_{Zd} \end{pmatrix} \right] \begin{pmatrix} \frac{1}{c_g} & 0 \\ 0 & \frac{1}{c_d} \end{pmatrix} \begin{pmatrix} X_g \\ X_d \end{pmatrix} \quad (\text{II.11})$$

avec

$$\begin{aligned} BIR &= BIR_{\text{mannequin}} \times \frac{FLT_{\text{individu}}}{FLT_{\text{mannequin}}} \\ C &= CASQUE_{\text{individu}} \end{aligned}$$

FIGURE II.4 – Position du casque K₁₀₀₀ sur la tête de l'individu

II.2.4 Matériel disponible

Pour effectuer les mesures de référence, nous avons choisi un mannequin acoustique. L'usage de cet « individu » particulier comporte plusieurs avantages : il est toujours disponible, manœuvrable, et parfaitement immobile (ce qui permet d'obtenir des mesures reproductibles). Nous utiliserons le mannequin HMS II, de marque HEAD ACOUSTICS, appartenant à la Direction de l'Ingénierie Véhicule de RENAULT. Il possède une tête, deux oreilles (pavillon + conduit), une bouche, et un buste ; il ne possède pas de jambes. Cependant, ces membres auraient été souhaitables, car les jambes d'un conducteur masquent plus ou moins partiellement le haut-parleur avant gauche placé en portière. Il associe un système d'enregistrement et un système de restitution sur casque. Deux méthodes de normalisation des BIR sont proposées :

1. La méthode FF (Free Field), qui permet d'égaliser les BIR par rapport au champ libre.
2. La méthode ID (Independent Direction), qui permet d'extraire des BIR l'information ne dépendant pas de la direction.

Il existe une troisième sortie, appelée LIN (Linear), qui n'effectue aucun traitement sur les BIR. À ces trois options sur le dispositif de mesure correspondent les mêmes options sur le dispositif de restitution. Les filtres du dispositif de restitution contiennent l'inverse du filtre d'enregistrement et l'inverse de la réponse du casque d'écoute. D'après les informations fournies par le constructeur [101, 311], ces filtres sont d'abord calculés de manière objective, puis optimisés de manière subjective par des tests d'écoute réalisés avec des individus. HEAD ACOUSTICS espère donc, avec les corrections FF ou ID, rapprocher les caractéristiques physiques du mannequin de celles d'un grand nombre d'individus.

Le casque utilisé est un casque *ouvert*, c.-à-d. se couplant avec l'oreille de la même manière qu'un haut-parleur en champ libre. D'après Møller [205], ce type de casque est nécessaire pour une simulation binaurale optimale. De plus, nous avons choisi un modèle dont les oreillettes sont en fait constituées de véritables petits haut-parleurs placés à distance des oreilles, afin d'améliorer le sentiment d'extériorisation des sons diffusés par le casque. Il s'agit du modèle K₁₀₀₀ de marque AKG. En outre, les oreillettes sont orientables, l'ouverture pouvant atteindre 50° (cf. fig. II.4).

Nous disposons également d'un casque égalisé par rapport au champ diffus, d'après les informations fournies par le constructeur. Il s'agit du modèle HD₄₁₄ de marque SENNHEISER.

II.3 Réalisation, traitements et exploitation des mesures

Comme nous l'avons expliqué précédemment, il est nécessaire de mettre en œuvre une méthode de correction individuelle des BIR mesurées sur un mannequin acoustique afin d'assurer une fidélité acceptable de la restitution binaurale. Ce mécanisme de correction est réalisé par des opérations de traitement numérique du signal effectuées dans le domaine fréquentiel ; il repose sur une décomposition des filtres en module et excès de phase.

II.3.1 Réalisation des mesures

Le filtrage inverse du casque, pour chacun des deux côtés, est l'opération indispensable à effectuer pour obtenir une restitution binaurale exacte, (cf. éq. (II.3)) :

$$Y_{individu} = \frac{BIR_{ind}}{CASQUE_{ind}} \times CASQUE_{ind} \times X.$$

Afin d'égaliser complètement le dispositif de mesure (position du microphone, filtrage du microphone et des amplificateurs), les mesures de BIR et la mesure de la réponse du casque doivent être effectuées lors de la même séance. Nous devons donc mesurer dans un habitacle, pour chacun des individus :

- les BIR pour chacun des quatre haut-parleurs ;
- les réponses d'un casque d'écoute.

Les mesures sont réalisées avec le logiciel MLSSA [269, 270] à une fréquence d'échantillonnage de 44.1 kHz.

La réalisation des filtres de correction individuelle et le test de leur validité repose sur la mise en œuvre de l'équation (II.9), pour le côté gauche et le côté droit, soit

$$Y_{individu} \approx \frac{BIR_{man}}{FLT_{man}} \times \frac{FLT_{ind}}{CASQUE_{ind}} \times CASQUE_{ind} \times X.$$

Les filtres FLT étant obtenus à partir de mesures de HRIR effectuées en chambre anéchoïque, la mesure du casque doit aussi être effectuée lors la même séance de mesure. Nous devons donc mesurer dans une chambre anéchoïque, pour chacun des mêmes individus, et pour le mannequin :

- les HRIR ;
- les réponses d'un casque d'écoute.

Nous devons aussi disposer d'une mesure de référence contenant, outre la réponse du haut-parleur de mesure, celle de la chaîne d'acquisition.

Les mesures sont réalisées avec le logiciel AMS développé par la société APIA et l'IRCAM, à une fréquence d'échantillonnage de 48 kHz.

La procédure de mesure des HRIR (ou des BIR) et des réponses impulsionnelles du casque est d'une importance capitale. Comme le préconise Møller [209], nous effectuons les mesures dans les conditions « blocked ear », c.-à-d. avec le conduit auditif bouché. En effet, toute l'information sur le timbre et la position de la source est présente à l'entrée du conduit auditif [305]. Les microphones utilisés pour les individus (modèle KE4-211-2 à électrets de marque SENNHEISER) sont donc placés à l'entrée du conduit à l'aide de bouchons d'oreille perforés en leur centre. Les microphones utilisés pour le mannequin HMS II sont insérés dans les conduits auditifs (modèle 4166 de marque BRÜEL & KJÆR). Le signal de mesure est récupéré en sortie du dispositif de mesure, filtré ou non par les filtres FF ou ID.

II.3.2 Traitements sur les mesures

Les mesures de HRIR effectuées en chambre anéchoïque contiennent l'effet de l'enceinte de mesure. Cette contribution est éliminée si on normalise les HRIR par la mesure de l'enceinte à l'aide d'un micro omnidirectionnel placé à la position du centre de la tête du sujet.

Nous devons également inverser la réponse impulsionnelle du casque d'écoute, afin d'obtenir une restitution binaurale exacte.

Enfin, nous devons établir le filtre de correction individuel, c.-à-d. le rapport de deux filtres de référence.

II.3.2.1 Décomposition des filtres

Tout filtre peut être vu comme la mise en série d'un filtre à phase minimale et d'un filtre passe-tout (cf. § C.1.3), de sorte que sa fonction de transfert peut s'écrire

$$H = H_{\min} H_{\text{exc}}, \quad (\text{II.12})$$

avec $H_{\min} = |H|e^{j\varphi_{\text{mph}}}$ le filtre à phase minimale, et $H_{\text{exc}} = e^{j\varphi_{\text{eph}}}$ le filtre passe-tout. On rappelle que tout filtre H est entièrement déterminé par son module $|H|$ et par l'excès de phase de sa composante passe-tout eph [174, 243], puisque

$$\begin{aligned} \text{mph} &= \Im(\text{Hilbert}(-\log |H|)) \quad , \text{ et} \\ \varphi &= \text{mph} + \text{eph}. \end{aligned} \quad (\text{II.13})$$

Mehrgardt et Mellert [196] ont montré expérimentalement que les HRTF peuvent être considérées comme des filtres à phase minimale ayant un retard indépendant de la fréquence jusque 10 kHz environ, retard fonction de l'angle d'indidence des sources sonores.

II.3.2.2 Compensation de l'effet de l'enceinte de mesure

Les traitements sont identiques dans le cas du calcul des filtres de corrections individuelles par rapport au champ libre, pondéré ou diffus. Cependant, la correction par rapport au champ libre fait intervenir à la fois un traitement sur le module et sur la phase, alors que les corrections par rapport au champ pondéré et au champ diffus sont des corrections à phase minimale, c.-à-d. faisant intervenir uniquement un traitement sur le module.

Nous décidons d'effectuer la compensation de l'effet de l'enceinte de mesure à la fois en module et en phase, car outre les déformations spectrales de la chaîne de mesure, nous désirons rattraper le retard de propagation entre le haut-parleur et le centre de la tête (ce deuxième traitement est néanmoins facultatif). L'opération à réaliser se résume donc sous la forme

$$h_{\text{compensée}} = \text{Fourier}^{-1} \left[\frac{|H_{\text{RTF}}|}{|HP|}, \text{Arg}_{H_{\text{RTF}}} - \text{Arg}_{HP} \right]. \quad (\text{II.14})$$

Egalisation en module Avant d'inverser la réponse de l'enceinte de mesure, il est nécessaire d'éliminer l'information qui n'en fait pas partie. Il est notamment important de s'affranchir d'anti-résonances éventuellement trop prononcées, dont l'inversion provoquerait à coup sûr l'amplification audible d'une plage de fréquence. C'est pourquoi un traitement en trois étapes a été décidé sur le module de la réponse de l'enceinte :

1. Un *lissage* est effectué avec une fenêtre non symétrique dont la largeur est proportionnelle à la fréquence. Cela revient à faire une moyenne avec une fenêtre symétrique de largeur constante sur une échelle logarithmique. Ce lissage désaccentue les anti-résonances profondes, qui apparaissent en haute fréquence, en altérant peu le module pour les plus basses fréquences. La largeur de la fenêtre est prise égale à un demi-ton.
2. Une *limitation de la dynamique* écrête les anti-résonances au-delà d'un certain seuil. Elle a été choisie égale à 20 dB.
3. Une *mise à plat* permet de conserver la même bande passante. En effet, les fréquences extrêmes présentent une amplitude très faibles, à cause de la bande passante des appareils de mesure utilisés ; si l'on inversait brutalement le module, l'amplitude inverse serait très élevée. On décide donc de prolonger le module par une constante, en deçà de 140 Hz et au-delà de 16 kHz.

Une fois ces traitements effectués, le module de la fonction de transfert de l'enceinte est inversé (cf. fig. II.5(a)), puis multiplié avec le module des HRTF. Les effets sur le module des HRTF maintenant égalisées sont (cf. fig. II.6) :

- une atténuation du pic entre 800 Hz et 1500 Hz ;

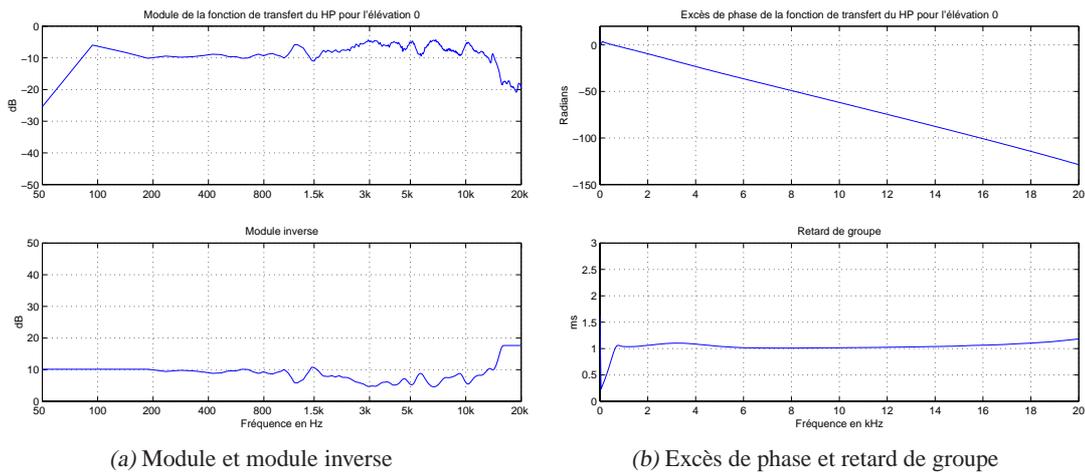


FIGURE II.5 – *Décomposition de la fonction de transfert du haut-parleur utilisé pour les mesures de HRIR en chambre anéchoïque. À gauche : module et module inverse ; le prolongement par une constante apparaît nettement dans les basses et hautes fréquences. À droite : excès de phase quasiment linéaire et retard de groupe associé, correspondant au retard de propagation entre le haut-parleur de mesure et le centre de la tête.*

- une atténuation de la bosse autour de 3000 Hz ;
- une remontée des hautes fréquences (au-dessus de 14 kHz, sans altération de la forme générale).

Egalisation en phase La principale compensation de la phase est en fait déjà effectuée par le biais de la correction du module, puisqu'on calcule la phase minimale à partir du module (cf. § II.3.2.1). Donc, il ne reste plus à effectuer qu'une correction de l'excès de phase.

Les HRIR ont la propriété d'avoir un excès de phase quasiment linéaire sur une large bande de fréquences [141] ; il en est de même pour les réponses impulsionnelles des bons haut-parleurs. Or un excès de phase linéaire est assimilable à un retard pur, donc on peut écrire

$$\text{phase} = \text{phase minimale due au module} + \text{retard de propagation} .$$

La linéarité de cet excès de phase est parfois mise à mal dans les basses et hautes fréquences : on observe des sauts, dont la position sur l'axe des fréquences coïncide avec celle d'anti-résonances prononcées sur le module. Ceci est dû à une mauvaise estimation de la phase minimale par la transformée de Hilbert. Ce n'est pas le cas ici, puisque l'excès de phase de la réponse du haut-parleur est quasiment linéaire (cf. fig. II.5(b)). La procédure d'égalisation consiste à approximer cet excès de phase par une droite ; pour cela, nous effectuons une régression linéaire sur la plus grande bande de fréquence possible, dénuée de décrochement ; on en déduit un retard, correspondant au temps de propagation de l'onde sonore depuis le haut-parleur jusqu'au micro de mesure omnidirectionnel, placé à la même position que le centre de la tête. Mais on ne retranche à l'excès de phase des HRTF qu'une fraction de ce retard. En effet, si on le retranchait intégralement, on pourrait avoir des réponses partiellement non causales : le délai de propagation entre le haut-parleur et le micro omnidirectionnel serait supérieur au retard entre le haut-parleur et le micro inséré dans l'oreille ipsilatérale, comme on peut le voir sur la figure II.7.

Comme on peut le constater sur la figure II.6, l'égalisation de l'excès de phase n'a finalement qu'une portée limitée, puisqu'on ne retranche qu'un retard pur.

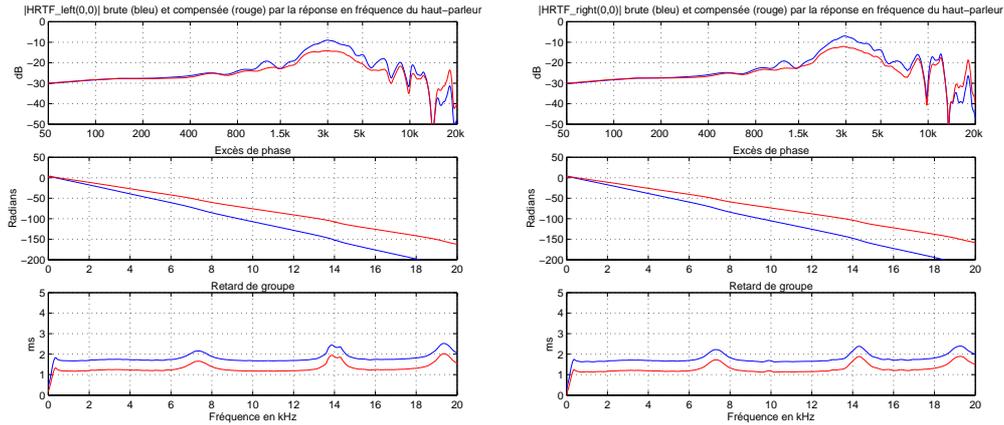


FIGURE II.6 – Compensation de l’influence du haut-parleur sur les HRTF. En bleu : les mesures de HRTF brutes ; en rouge : les mesures de HRTF corrigées par celle du haut-parleur.

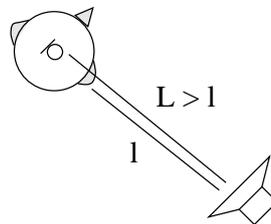


FIGURE II.7 – Représentation des chemins acoustiques lors de la mesure d’une HRIR pour un azimut quelconque

II.3.2.3 Inversion de la réponse impulsionnelle d’un casque d’écoute

L’opération à réaliser peut être symbolisée par

$$h_{inverse} = \text{Fourier}^{-1} \left[\frac{1}{|\text{CASQUE}|}, -\text{Arg}_{\text{CASQUE}} \right]. \quad (\text{II.15})$$

Nous remarquons sur la figure II.8 que l’excès de phase du casque K_{1000} est quasiment linéaire, et que le retard de groupe est constant jusque 10 kHz. De plus, la différence entre le retard gauche et droit est toujours inférieure à 0.04 ms, ce qui témoigne d’une imprécision de localisation inférieure à 4° dans le plan horizontal, d’après Blauert [20]. En d’autres termes, la différence de marche constatée entre les chemins acoustiques oreillette→micro, à gauche et à droite, est toujours inférieure à 1.3 cm. Nous décidons donc de ne pas tenir compte de l’excès de phase des casques d’écoute (nous lui imposons une valeur nulle). Nous réalisons donc uniquement une égalisation à phase minimale, c.-à.-d. une inversion du module. La démarche est la même que celle exposée au paragraphe précédent.

Si nous disposons de plusieurs mesures effectuées avec la même position des microphones, nous en effectuons la moyenne avant de commencer les traitements. L’écrêtage des résonances et anti-résonances trop prononcées permet de minimiser les phénomènes dûs à des artefacts de mesure (micro situé à un ventre ou à un noeud de pression d’ondes stationnaires créées par la mesure « blocked ear »), et la mise à plat en basse et haute fréquence permet de s’affranchir de la bande passante limitée du système de mesure (cf. fig. II.9). Les paramètres sont réglés à l’oreille pour que l’inversion des hautes fréquences ne génère pas de niveau beaucoup plus élevé que celui des basses fréquences.

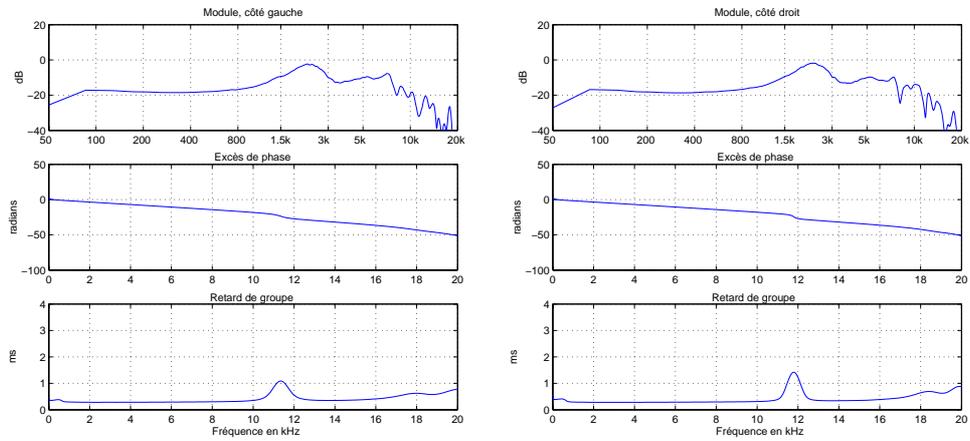


FIGURE II.8 – Module, excès de phase et retard de groupe d'une mesure de la réponse impulsionnelle du casque K₁₀₀₀ oreillettes ouvertes, mesuré sur le mannequin en chambre anéchoïque, côté gauche et côté droit

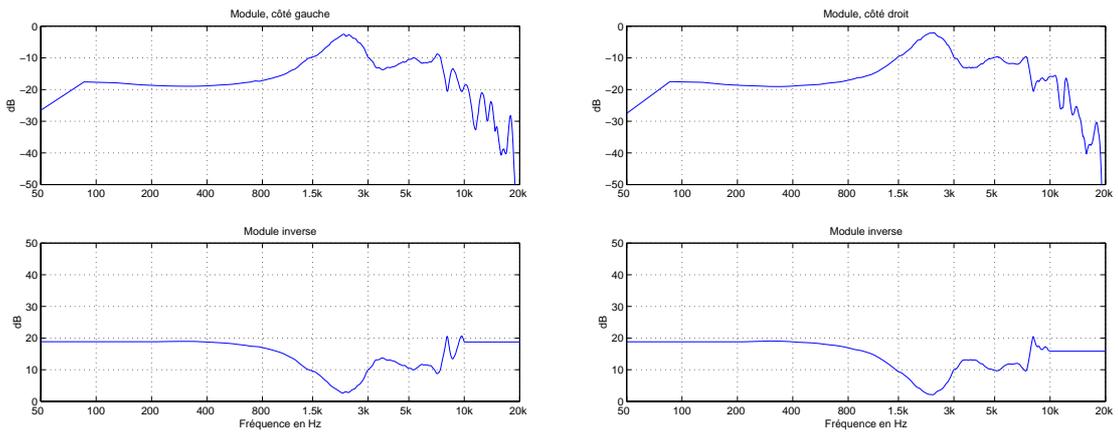


FIGURE II.9 – Inversion du module de la fonction de transfert du casque K₁₀₀₀ oreillettes ouvertes, mesuré sur le mannequin en chambre anéchoïque, côté gauche et côté droit

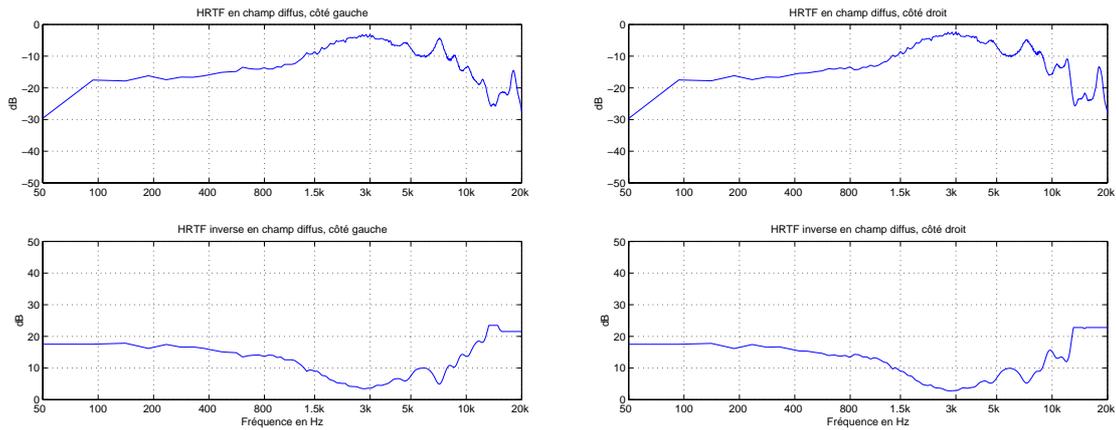


FIGURE II.10 – Inversion de la HRTF en champ diffus mesuré sur le mannequin en chambre anéchoïque, côté gauche et côté droit

II.3.2.4 Calcul et inversion de la HRTF en champ diffus ou pondéré

La fonction de transfert de la tête en champ diffus est obtenue par la racine carrée de la puissance moyenne de HRTF mesurées pour un grand nombre d'incidences régulièrement réparties [174]. Cette méthode permet d'introduire une pondération directionnelle pour l'établissement de la HRTF en champ pondéré : un poids est accordé aux HRTF suivant l'incidence considérée, correspondant aux positions géométriques des haut-parleurs référencés W, X, Y et Z dans l'habitacle [175].

Deux autres techniques existent pour le calcul strict de la HRTF en champ diffus [175] ; elles permettent de s'affranchir de mesures longues et pénibles en chambre anéchoïque. Il s'agit de la mesure en chambre réverbérante, en régime stationnaire, et du calcul par analyse temps-fréquence d'une BIR mesurée dans une salle courante.

Les HRTF en champ diffus ou pondéré étant délibérément définies uniquement par leur module, il s'agit de filtres à phase minimale (cf. § II.3.2.1). La procédure de calcul du filtre inverse est donc identique à celle mise en œuvre pour le casque (cf. fig. II.10). Les filtres de corrections individuelles sont obtenus en multipliant un filtre de référence (HRTF en champ diffus ou pondéré) par l'inverse d'un autre filtre de référence.

II.4 Optimisation du mode de reproduction

II.4.1 Méthodologie

Nous devons déterminer le meilleur système de reproduction au casque, c.-à-d. celui restituant les sensations auditives les plus proches de l'écoute réelle en habitacle.

Nous avons vu au § II.2.3.2 que la simulation binaurale est exacte seulement si chaque individu utilise ses propres HRTF, et si on peut éliminer l'influence du casque d'écoute. Comme nous l'avons déjà évoqué, cette situation n'est pas réaliste si on veut réaliser des tests d'évaluation subjective à grande échelle. Il a donc été choisi de réaliser les mesures sur un mannequin acoustique (modèle HMS II de marque HEAD ACOUSTICS). Cependant, nous avons suggéré de corriger individuellement ces mesures, afin de les rapprocher de celles qui auraient été obtenues avec les oreilles de chaque individu.

La nécessité et la validité de ces corrections sera évaluée par rapport à l'écoute réelle en habitacle. Nous comparerons d'abord les stimuli de manière objective, c.-à-d. tels qu'ils apparaissent au niveau des tympans d'un individu. D'autre part, un test d'écoute perceptif sera conduit sous forme de comparaisons par paires, et se déroulera *in situ*, à l'intérieur de l'habitacle du véhicule ayant servi à mesurer les BIR.

II.4.2 Test préliminaire

Un premier test, conduit par Larcher [175], a permis la comparaison entre un stimulus choisi comme référence, et un ensemble d'autre stimuli dont on voulait évaluer la proximité par rapport au premier. Les stimuli ont été convolués en temps réel avec une source sonore numérique à l'aide de cartes électroniques spécialement dédiées (système HURON CP4 de marque LAKE DSP¹ [188] mettant en œuvre un algorithme spécifique de convolution temps-réel sans retard [93, 267]). Le test s'est déroulé en laboratoire. La comparaison s'est effectuée à l'aide d'une notation sur une liste d'attributs représentant la sensation globale, la distribution spatiale et la distribution fréquentielle ressentie en écoutant les stimuli. Elle a été conduite en double aveugle. Quatre sujets ont passé le test.

Pour des raisons pratiques, les stimuli ont été écoutés à l'aide d'un casque de marque STAX, modèle LAMBDA PRO. Les stimuli initialement traités pour être écoutés avec le casque K₁₀₀₀, ont donc été corrigés par le filtre adéquat.

Définition des stimuli Le stimulus de référence était celui supposé restituer fidèlement ce que l'auditeur aurait entendu dans l'habitacle, c.-à-d.

$$\text{stimulus REF} = \frac{\text{BIR}_{\text{auditeur}}}{\text{CASQUE}_{\text{auditeur}}}.$$

Afin d'évaluer la qualité de l'adaptation individuelle de mesures réalisées sur une tête artificielle, le stimulus de référence a été comparé aux stimuli suivants :

- le signal traité par les filtres internes au dispositif HEAD ACOUSTICS (stimulus 1) : les BIR mesurées sur la tête artificielle sont filtrés à la restitution par le dispositif HEAD ACOUSTICS, pour un casque d'écoute STAX ;
- le stimulus de référence du mannequin :

$$\text{stimulus 2} = \frac{\text{BIR}_{\text{mannequin}}}{\text{CASQUE}_{\text{mannequin}}};$$

- les signaux adaptés à l'auditeur par une correction individuelle diffuse (3), pondérée (4) ou libre (5) :

$$\text{stimulus 3,4,5} = \frac{\text{BIR}_{\text{mannequin}}}{\text{CASQUE}_{\text{auditeur}}} \times \frac{\text{REF}_{\text{auditeur}}}{\text{REF}_{\text{mannequin}}}.$$

Les filtres normalisant (REF) étant obtenus par combinaison des HRIR mesurées en chambre anéchoïque, la mesure de casque indiquée a été effectuée lors de la même séance.

D'autre part, les études menées par Wenzel [317] et par Møller [208] montrent que le récepteur le plus proche d'un individu est un autre individu spécifiquement choisi pour minimiser un ou plusieurs critères (comme par exemple le nombre d'erreurs de localisation dans le plan médian, ou à l'intérieur ou encore à l'extérieur du cône de confusion). Si on choisit un individu typique, c.-à-d. un mannequin étudié pour que ses caractéristiques acoustiques correspondent à celles d'un auditeur moyen, Møller [206] montre que les performances de localisation sont globalement identiques à celles obtenues avec un individu quelconque (donc inférieures à celles obtenues avec un individu spécifiquement choisi). Ces observations sont confirmées par Minnaar [199], pour qui 60% des têtes humaines donnent de meilleurs résultats en localisation que la meilleure des têtes artificielles ! L'idée selon laquelle un mannequin contient l'essentiel de nos caractéristiques acoustiques serait donc fautive ; les BIR d'un individu de référence pourraient alors être utilisées par tous les autres, et les stimuli suivants ont également été testés :

- le signal de référence d'un autre individu :

$$\text{stimulus 6} = \frac{\text{BIR}_{\text{autre auditeur}}}{\text{CASQUE}_{\text{autre auditeur}}};$$

1. <http://www.lake.au>

TABLEAU II.1 – Résultats du test d'écoute préliminaire réalisé par Larcher [175]

Numéro du stimulus	6	2	3	7	4	5	1
Description	référence autre individu	référence mannequin	correction individuelle diffuse	correction non individuelle diffuse	correction individuelle libre	correction individuelle pondérée	filtrage HEAD ACOUSTICS
Rang par rapport au stimulus de référence	1	2	3	4	5	6	7

– et enfin la correction individuelle diffuse pour un autre individu :

$$\text{stimulus 7} = \frac{\text{BIR}_{\text{mannequin}}}{\text{CASQUE}_{\text{autre auditeur}}} \times \frac{\text{DIFF}_{\text{autre auditeur}}}{\text{DIFF}_{\text{mannequin}}}.$$

Résultats Les résultats de ce test d'écoute sont rassemblés dans le tableau II.1. Il apparaît que le stimulus 1 est jugé éloigné de la référence (les filtres proposés par HEAD ACOUSTICS dans le dispositif de mesure et dans celui de reproduction introduisent une forte coloration dans le grave). Le stimulus 5 est également jugé éloigné de la référence, tout comme le stimulus 4 spécialement pour la reproduction spatiale. Les stimuli 2, 3 et 7 forment un groupe de réponses assez proches du stimulus de référence, spécialement pour la reproduction du timbre. Enfin le stimulus 6 est jugé le plus proche de la référence.

Il ressort de ce test préliminaire que les filtres internes proposés par HEAD ACOUSTICS sont loin d'être optimaux, et que l'inversion numérique de la fonction de transfert du casque d'écoute, proposée par Larcher, apporte une amélioration substantielle de la fidélité de reproduction au casque. En outre, les corrections individuelles par rapport au champ libre ou pondéré doivent être abandonnées. Néanmoins, la validité du stimulus choisi comme référence n'a pas été évaluée : il n'est pas certain qu'il soit effectivement le plus proche de l'écoute réelle en habitacle. Notons enfin que la correction appliquée sur certains stimuli pour les rendre compatibles avec le dispositif d'écoute HEAD ACOUSTICS est délicate.

Ce test a été reconduit par nos soins, avec trois individus. Cette fois, l'écoute a été entièrement réalisée avec le casque K₁₀₀₀. Ce casque, de part sa structure, nous paraît mieux à même d'assurer une bonne extériorisation des stimuli. Le dispositif d'écoute HEAD ACOUSTICS, que l'on veut reproduire avec le stimulus 1, a été simplement simulé par filtrage, sa fonction de transfert ayant été préalablement mesurée.

Nous sommes arrivés sensiblement aux mêmes conclusions. Le stimulus le plus proche de la réalité, en ce qui concerne le timbre et la localisation, est l'écoute de nos propres stimuli de référence. Concernant l'extériorisation, les performances sont assez modestes, puisqu'il est assez difficile d'imaginer que l'on se trouve assis au volant d'une voiture. De bonnes performances en localisation existent pour le stimulus de référence du mannequin, ainsi que pour la correction individuelle en champ diffus. Concernant le timbre, tous les stimuli présentent des colorations plus ou moins violentes, le plus proche de notre stimulus de référence étant celui du mannequin, corrigé ou non par le champ diffus de l'auditeur.

II.4.3 Choix des méthodes d'égalisation

Le test préliminaire a permis d'éliminer les corrections individuelles par rapport au champ libre et au champ pondéré.

Il s'agit maintenant à déterminer la méthode d'égalisation permettant la restitution au casque la plus proche de l'écoute réelle en véhicule. Pour cela, nous allons comparer plusieurs stimuli par rapport à l'écoute naturelle des haut-parleurs de la voiture, à savoir :

- le stimulus de référence de l'individu qui écoute ;
- le stimulus de référence d'un autre individu ;

TABLEAU II.2 – Liste récapitulative des différents stimuli utilisés pour le test de validation concernant la restitution binaurale

Type	Nom	N°	Fonction de transfert	Source
Référence individu	Ref_ind	1	$\frac{BIR_{\text{auditeur}}}{CASQUE_{\text{auditeur}}}$	Casque
Référence autre individu	Ref_autre	2	$\frac{BIR_{\text{autre auditeur}}}{CASQUE_{\text{autre auditeur}}}$	Casque
Référence tête artificielle	Ref_tête	3	$\frac{BIR_{\text{mannequin}}}{CASQUE_{\text{mannequin}}}$	Casque
Correction individuelle diffuse	Diff_ind	4	$\frac{BIR_{\text{mannequin}}}{CASQUE_{\text{auditeur}}} \times \frac{DIFF_{\text{auditeur}}}{DIFF_{\text{mannequin}}}$	Casque
Correction non individuelle diffuse	Diff_autre	5	$\frac{BIR_{\text{mannequin}}}{CASQUE_{\text{autre auditeur}}} \times \frac{DIFF_{\text{autre auditeur}}}{DIFF_{\text{mannequin}}}$	Casque
Filtre ID proposé par HEAD ACOUSTICS	ID_tête	6	$BIR_{\text{mannequin}} \times ID$	Casque
Ecoute naturelle	Voiture	7	1	Haut-parleurs

- le stimulus de référence du mannequin ;
- une égalisation des BIR du mannequin par rapport à la fonction de transfert d'oreille (HRTF) en champ diffus du mannequin, suivie d'une correction individuelle du casque d'écoute égalisé par rapport à la HRTF en champ diffus de l'individu qui écoute ;
- une égalisation des BIR du mannequin par rapport à la HRTF en champ diffus du mannequin, suivie d'une correction individuelle du casque d'écoute égalisé par rapport à la HRTF en champ diffus d'un autre individu ;
- le signal traité par le dispositif spécifique HEAD ACOUSTICS.

Pour évaluer la validité de ces corrections, nous nous proposons de réaliser un test d'écoute subjectif. Il sera effectué sur une dizaine de sujets, dans l'habitacle de la voiture ayant servi à mesurer les BIR. Le but est double :

1. Savoir si l'introduction des HRTF en champ diffus dans les corrections individuelles est susceptible d'apporter une amélioration systématique de fidélité.
2. Caractériser les éventuels biais que l'évaluation psycho-expérimentale au moyen de la convolution est susceptible d'introduire systématiquement par rapport à l'écoute naturelle en habitacle.

Le test servira donc à fixer la technique de correction individuelle (si elle est nécessaire) effectuée sur les BIR du mannequin, et qui sera employée dans toute la suite de l'étude concernant la restitution Hi-Fi en habitacle automobile.

Les stimuli sont construits par convolution entre le signal source et une des réponses impulsionnelles dont la fonction de transfert est décrite dans le tableau II.2. Le filtre ID nécessite normalement l'écoute à l'aide du dispositif fourni par HEAD ACOUSTICS, c.-à-d. avec un casque spécifique ; pour qu'il puisse être écouté avec le K_{1000} , nous lui avons appliqué le filtre de correction adéquat.

II.4.4 Comparaison objective

Les mesures ont toutes été analysées avec une transformée de Fourier discrète sur 1024 points, fournissant une résolution fréquentielle de 47 Hz. Sur les figures II.11 à II.19, le trait gras représente la moyenne sur les individus sans le mannequin, le trait fin représente le mannequin, et la zone grisée correspond à un intervalle d'un écart type de part et d'autre de la moyenne (si on suppose une distribution

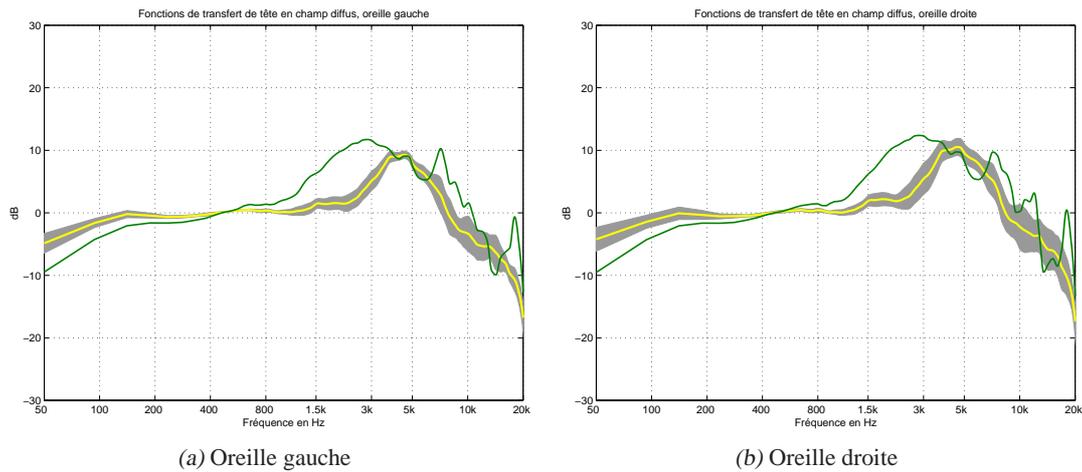


FIGURE II.11 – Fonctions de transfert en champ diffus de plusieurs têtes. Nous représentons la moyenne et l'écart-type ; la fonction de transfert en champ diffus du mannequin est tracée à part.

normale, cet intervalle contient 68% des observations). Comme on désire se concentrer sur les variations interindividuelles dans les hautes fréquences, un facteur de cadrage a été appliqué sur les fonctions de transfert pour leur donner un niveau commun dans la bande [200-800] Hz.

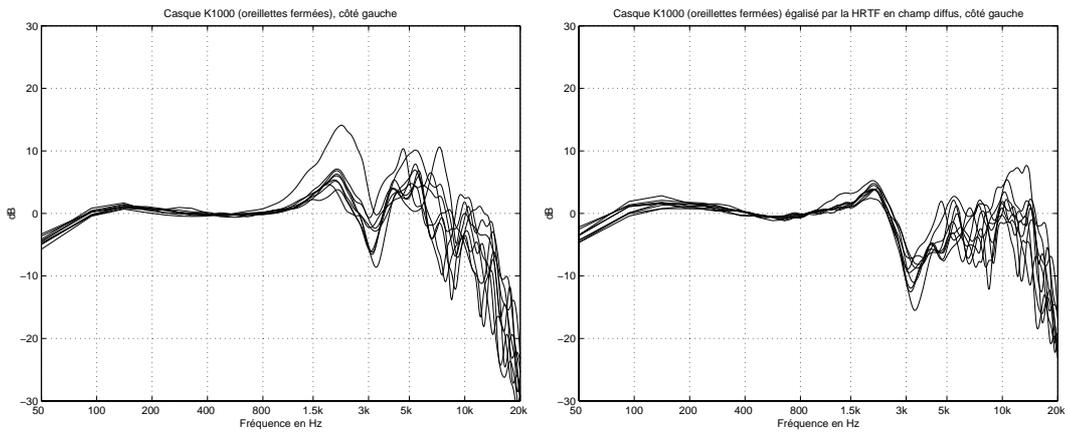
II.4.4.1 Dans une chambre anéchoïque

Nous décidons d'effectuer une « simulation » du test en chambre anéchoïque, avec une source unique placée en face de l'auditeur. Nous allons donc comparer, pour une dizaine individus :

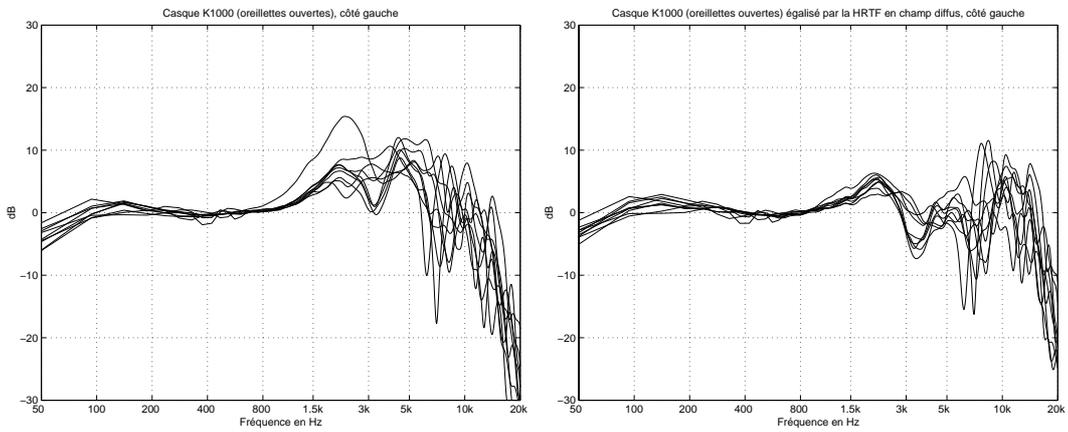
- leur HRTF en face ;
- leur HRTF en champ diffus ;
- leur HRTF en face normalisée par leur HRTF diffuse ;
- leur fonction de transfert de casque ;
- leur fonction de transfert de casque normalisée par leur HRTF diffuse ;
- leur HRTF en face normalisée par deux fonctions de transfert de casque (stimuli de référence) ;
- les différentes corrections individuelles de la HRTF en face du mannequin par rapport à leur HRTF en champ diffus, normalisées par les fonctions de transfert du casque lui-même égalisé par rapport à la HRTF en champ diffus (stimuli de corrections individuelles par rapport au champ diffus).

La figure II.11 représente les HRTF en champ diffus calculées à partir d'une distribution de mesures en chambre anéchoïque. Les différences interindividuelles sont faibles sur toute la bande de fréquences audibles (à peine quelques dB). En revanche, celle du mannequin s'en détache nettement, surtout dans la bande [800-3000] Hz.

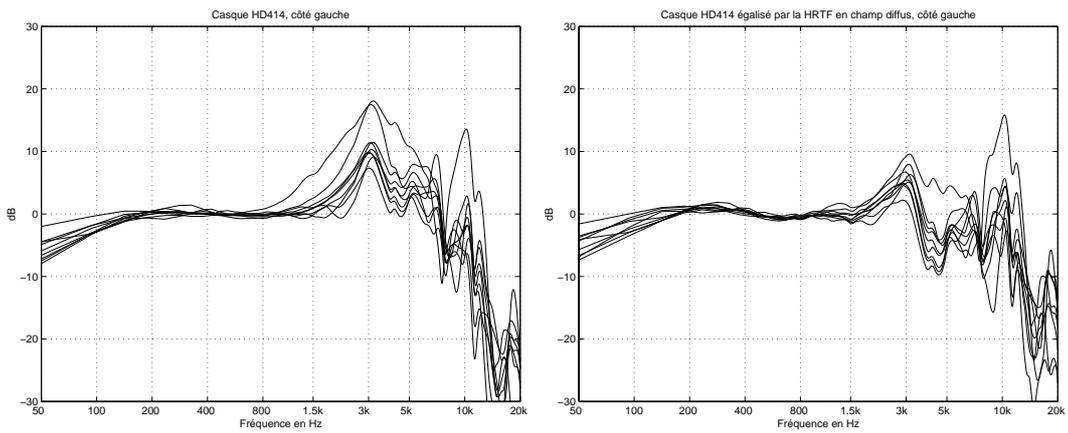
La figure II.12 présente les mesures de fonctions de transfert de casques effectuées sur différents individus. Deux casques ont été utilisés : le casque K_{1000} , oreillettes fermées et oreillettes ouvertes à 40° , et le casque HD_{414} . Sur les fonctions de transfert du K_{1000} oreillettes fermées, non égalisées par la HRTF en champ diffus, un individu se détache nettement, surtout dans la bande [800-3000] Hz. Ce n'est plus le cas sur les fonctions de transfert égalisées par la HRTF en champ diffus. La normalisation par la HRTF en champ diffus diminue donc les différences interindividuelles. Elle a également pour effet de réduire les différences entre les fonctions de transfert mesurées sur les humains et la fonction de transfert mesurée sur le mannequin.



(a) Casque K₁₀₀₀ oreillettes fermées

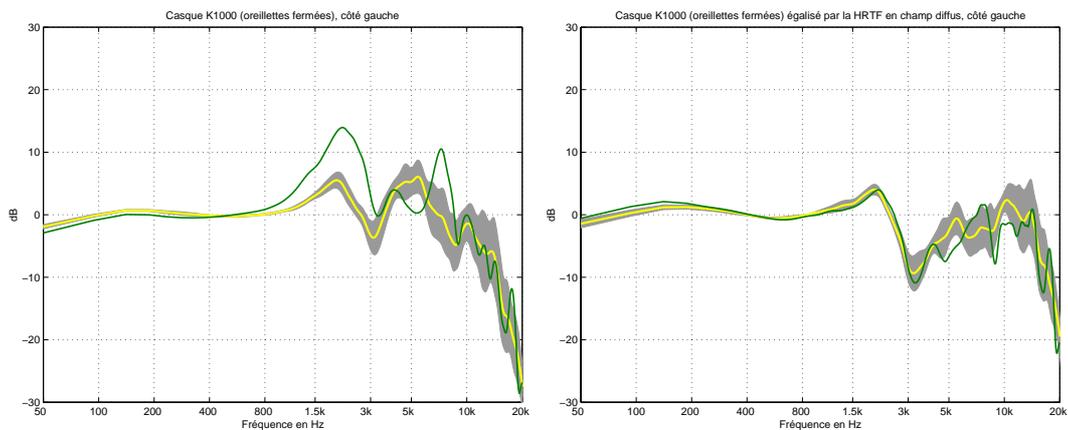


(b) Casque K₁₀₀₀ oreillettes ouvertes

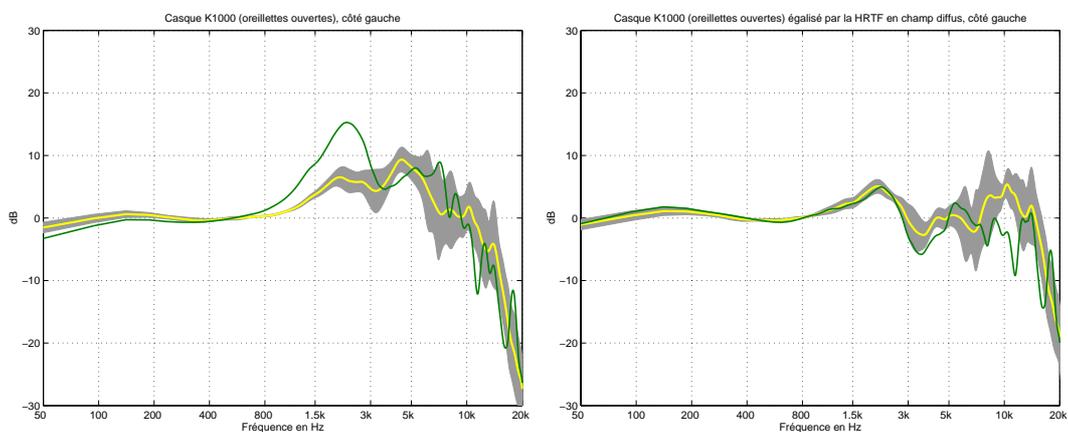


(c) Casque HD₄₁₄

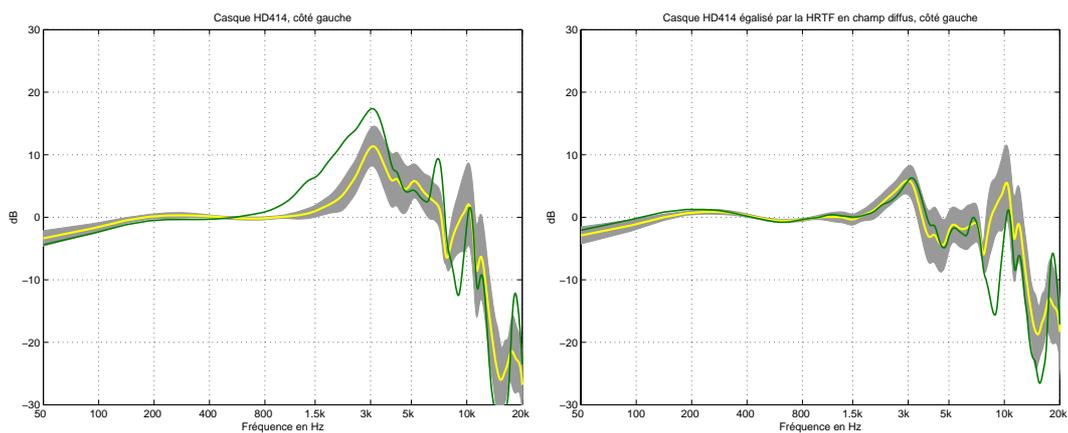
FIGURE II.12 – Fonctions de transfert des casques considérés. À gauche : mesures brutes ; à droite : mesures normalisées par la HRTF individuelle en champ diffus



(a) Casque K₁₀₀₀ oreillettes fermées



(b) Casque K₁₀₀₀ oreillettes ouvertes



(c) Casque HD₄₁₄

FIGURE II.13 – Fonctions de transfert des casques considérés. Nous représentons la moyenne et l'écart-type ; la fonction de transfert du casque mesurée sur le mannequin est tracée à part. À gauche : mesures brutes ; à droite : mesures normalisées par la HRTF individuelle en champ diffus

La représentation de la figure II.13 améliore considérablement la compréhension du phénomène. Si on excepte le mannequin, les écarts entre les différentes fonctions de transfert sont assimilables à :

- un gain pur (à 3 dB près) jusqu'à 3 kHz, sur la version non égalisée ;
- un gain pur (à 3 dB près) jusqu'à 5 kHz, sur la version égalisée.

En ce qui concerne le K_{1000} , les différences interindividuelles sont plus importantes lorsque la mesure est effectuée oreillettes ouvertes ; il est en effet difficile de repérer une position précise sur la tête des individus. En ce qui concerne le HD_{414} , les différences interindividuelles des mesures brutes sont plus importantes qu'avec le K_{1000} ; ceci permet de mettre en doute les informations fournies par le constructeur selon lesquelles ce casque serait déjà égalisé par rapport au champ diffus

La normalisation par la HRTF en champ diffus permet également de réduire les écarts droite-gauche, comme on peut le voir sur la figure II.14 ; l'émergence de différences interindividuelles significatives est repoussée d'une octave environ. En limitant l'étendue fréquentielle de notre étude à 5 kHz, on pourrait envisager de prendre le même filtre d'égalisation du casque K_{1000} sur les deux oreilles (COR_{CASQUE} dans l'équation (II.10)).

Comme les différences gauche-droite sont faibles pour toutes les fonctions de transfert considérées, nous choisissons de représenter la moyenne des deux oreilles.

La figure II.15 montre les HRTF frontales (azimut nul, site nul). Ici encore les mêmes conclusions s'imposent : la normalisation par la HRTF en champ diffus permet de réduire les différences interindividuelles, et de rapprocher la mesure effectuée sur le mannequin de celles effectuées sur les humains.

La figure II.16 montre les HRTF frontales égalisées par les fonctions de transfert de casque. Il s'agit des stimuli de référence, c.-à-d. restituant au casque, le plus fidèlement possible, un enregistrement effectué dans les oreilles des individus, soit

$$\frac{BIR_{\text{auditeur}}}{CASQUE_{\text{auditeur}}}, \text{ avec ici } BIR = HRIR.$$

On constate que cette égalisation fait presque jeu égal avec la normalisation par la HRTF en champ diffus jusque 5 kHz, surtout dans le cas du K_{1000} utilisé oreillettes ouvertes et dans le cas du HD_{414} . La HRTF mesurée sur le mannequin devient très proche de celles mesurées sur les humains. Tout se passe comme si les fonctions de transfert du K_{1000} contenaient certaines informations propres à chaque individu, et que la normalisation des HRTF par ces fonctions de transfert les enlevait. Cependant des différences importantes subsistent à partir de 7 kHz.

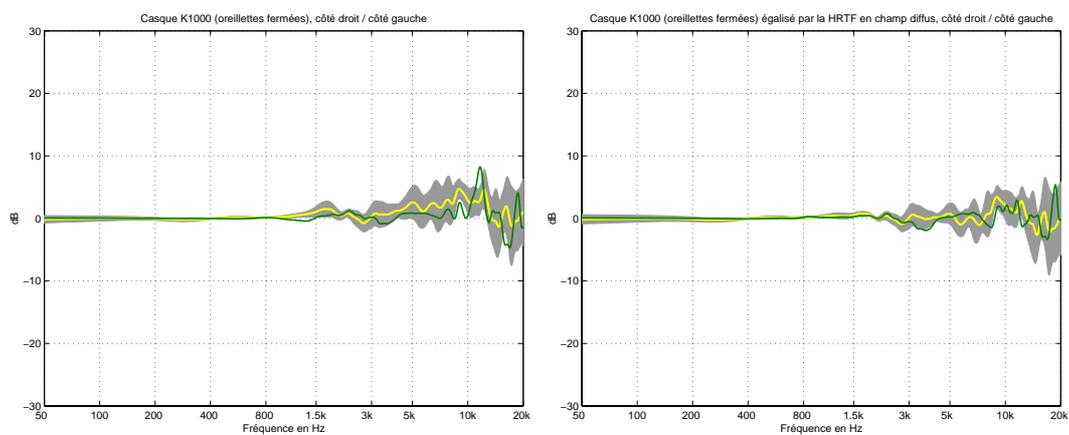
La figure II.17 montre les stimuli intégrant la correction individuelle par rapport à la HRTF en champ diffus, c.-à-d. le produit

$$\frac{BIR_{\text{mannequin}}}{DIFF_{\text{mannequin}}} \times \frac{DIFF_{\text{auditeur}}}{CASQUE_{\text{auditeur}}}, \text{ toujours avec } BIR = HRIR.$$

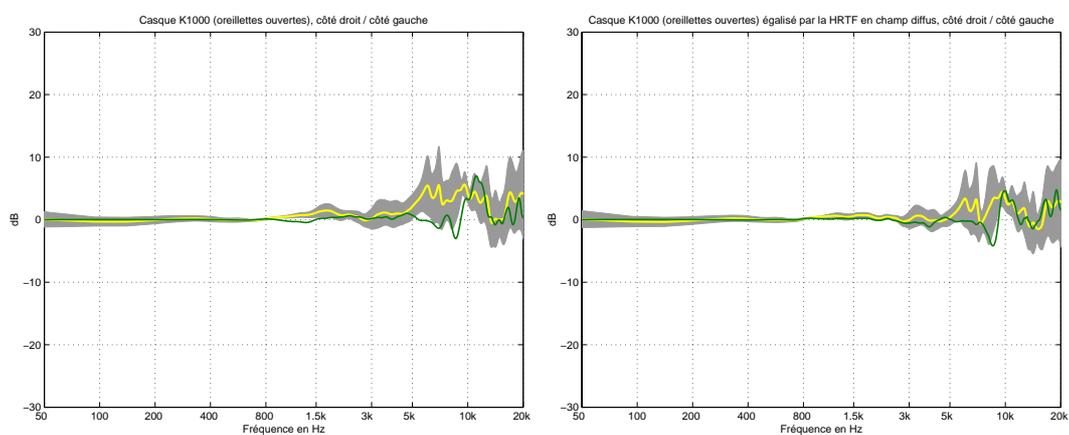
Aucun traitement n'est appliqué sur les mesures avant inversion. Le trait fin représente le stimulus de référence du mannequin, par simplification de l'équation ci-dessus. Il est très proche des corrections individuelles diffuses des autres individus, sur toute la largeur de la bande audible. La comparaison des figures II.16 et II.17 montre que, par rapport au stimulus de référence de chaque individu, et jusque 6 kHz environ, le stimulus de référence de la tête est presque aussi proche que chaque correction individuelle diffuse.

II.4.4.2 Dans l'habitacle

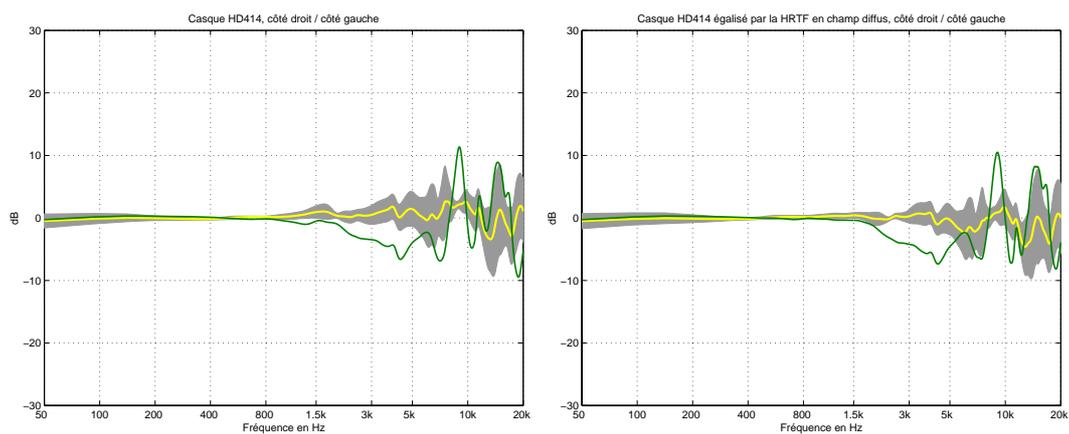
Nous choisissons dorénavant comme casque le K_{1000} oreillettes ouvertes. Bien que certaines différences persistent entre les individus même après l'égalisation par la HRTF en champ diffus, nous avons privilégié ce modèle car il permet, de par sa structure, une bonne extériorisation des sons.



(a) Casque K₁₀₀₀ oreillettes fermées



(b) Casque K₁₀₀₀ oreillettes ouvertes



(c) Casque HD₄₁₄

FIGURE II.14 – Différences droite-gauche des fonctions de transfert des casques considérés. À gauche : mesures brutes ; à droite : mesures normalisées par la HRTF individuelle en champ diffus

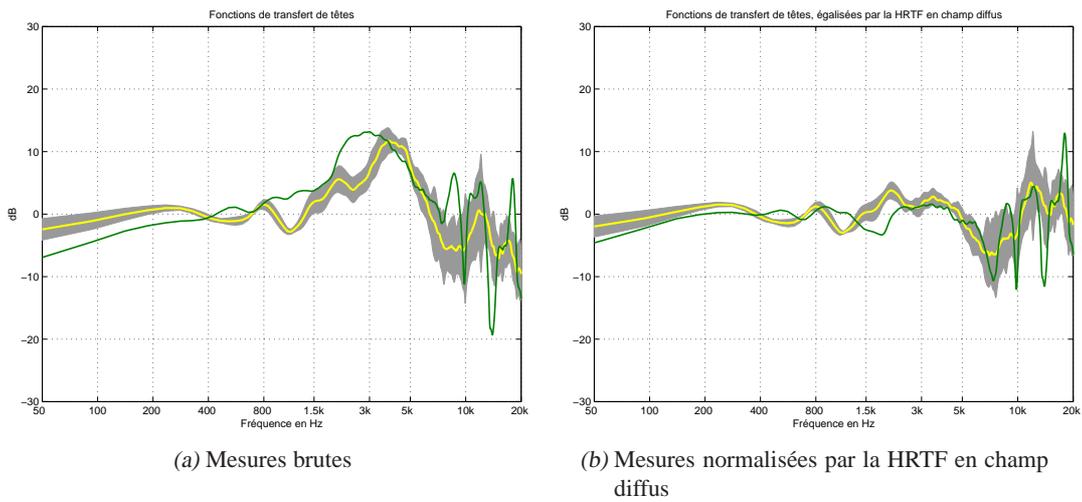


FIGURE II.15 – Fonctions de transfert de têtes $HRTF(0,0)$, moyennées entre les deux oreilles

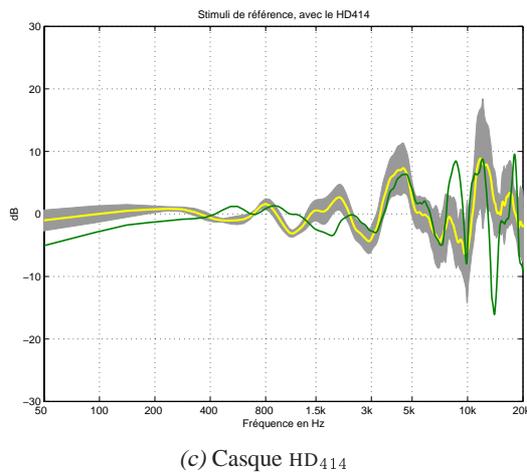
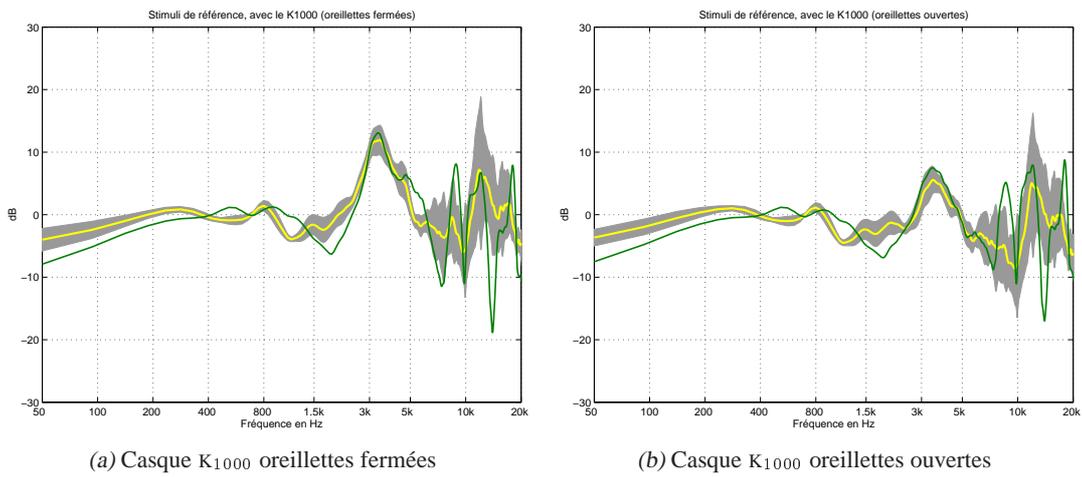


FIGURE II.16 – Fonctions de transfert de tête $HRTF(0,0)$ normalisées par trois fonctions de transfert de casque, moyennées entre les deux oreilles

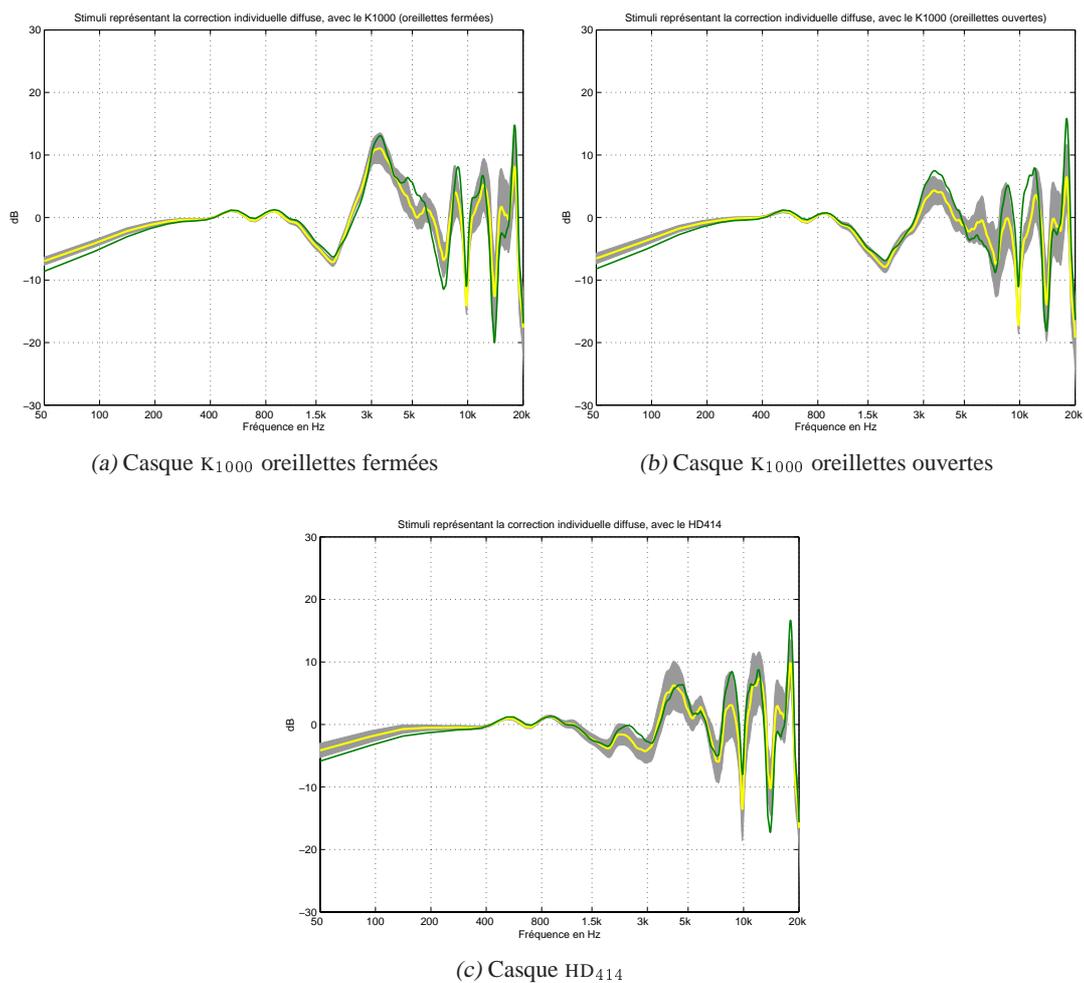


FIGURE II.17 – $HRTF(0,0)$ du mannequin adaptée aux auditeurs par une correction individuelle diffuse, suivie d'une inversion de la fonction de transfert individuelle du casque. Les fonctions de transfert finales sont moyennées entre les deux oreilles

Nous voulons effectuer la comparaison objective des stimuli, tels qu'ils apparaîtront au niveau des tympans de l'individu effectuant le test d'écoute dans l'habitacle. Nous remplaçons donc, dans les figures II.16 et II.17, les HRTF frontales en champ libre par quatre fonctions de transfert binaurales, correspondant aux quatre fonctions de transfert haut-parleur \rightarrow oreille. La figure II.18 présente les quatre stimuli de référence (calculés avec la mesure du casque effectuée en habitacle), c.-à-d.

$$\frac{\text{BIR}_{\text{auditeur}}}{\text{CASQUE}_{\text{auditeur}}},$$

et la figure II.19 présente les quatre stimuli intégrant la correction individuelle diffuse (calculés avec la mesure du casque effectuée en chambre anéchoïque), c.-à-d.

$$\frac{\text{BIR}_{\text{mannequin}}}{\text{DIF}_{\text{mannequin}}} \times \frac{\text{DIF}_{\text{auditeur}}}{\text{CASQUE}_{\text{auditeur}}}.$$

Le mode de représentation est identique pour les deux séries de figures. On notera la dissymétrie des fonctions de transfert, liée au placement particulier du conducteur par rapport au système électroacoustique.

Jusqu'à 5 kHz, le stimulus de référence du mannequin ne s'écarte pas fondamentalement de ceux mesurés sur les humains. Il reste compris dans la zone ayant une amplitude d'une valeur de l'écart type, soit quelque dB. Cependant, les différences seront certainement audibles, car on peut constater localement des résonances ou anti-résonances assez marquées.

L'examen des corrections individuelles par rapport au champ diffus révèle que les différences entre le mannequin et les individus sont inexistantes jusqu'à 3 kHz, et restent comprises dans la zone de tolérance jusqu'à 10 kHz. Rappelons que le trait fin correspond toujours au stimulus de référence du mannequin. Ces différences seront probablement difficiles à mettre en évidence à l'aide d'un test d'écoute.

Il ressort de cette étape que le stimulus de référence du mannequin est très proche des différentes corrections individuelles par rapport au champ diffus. Même si celui-ci est parfois éloigné de notre propre stimulus de référence, il ne semble pas justifié de mettre en œuvre un processus complexe de corrections individuelles. Nous allons tenter de confirmer ces observations par un test d'écoute subjectif.

II.4.5 Modalités du test d'écoute

Il existe plusieurs formes de tests d'évaluation subjective, mais tous ne sont pas aussi efficaces [247]. Il convient donc de choisir précisément le type de test, en fonction des hypothèses à vérifier.

La méthode d'attribution des rangs consiste à demander un classement des stimuli suivant un critère d'évaluation. Bien que très intuitive, cette méthode doit cependant être abandonnée car elle ne fournit aucun facteur d'échelle : on détermine simplement le fait qu'un stimulus soit préféré à un autre, mais on ne sait pas de combien il a été préféré. De plus, la complexité de cette méthode croît factoriellement avec le nombre de stimuli.

La méthode d'attribution des notes consiste à demander une valeur chiffrée et bornée pour chaque stimulus. Mais toujours d'après [247], cette méthode ne doit être utilisée qu'avec des auditeurs expérimentés et entraînés. Les principales limitations sont :

- la mise en correspondance, pour chaque stimulus, de l'impression perceptive avec un chiffre relié à une échelle ;
- l'utilisation différente de l'échelle par les différents sujets ;
- l'attribution très rare des notes extrêmes ;
- l'absence de corrélation fiable avec des indices objectifs mesurables.

La méthode des comparaisons par paires s'affranchit de ces difficultés, puisqu'elle consiste simplement à comparer deux stimuli entre eux selon un unique attribut sémantique. La présence d'une échelle de notation n'est pas indispensable. Cette méthode ne requiert que des comparaisons relatives, et les sujets n'ont donc pas à s'inquiéter de leurs jugements passés ou futurs. Elle est parfaitement adaptée aux

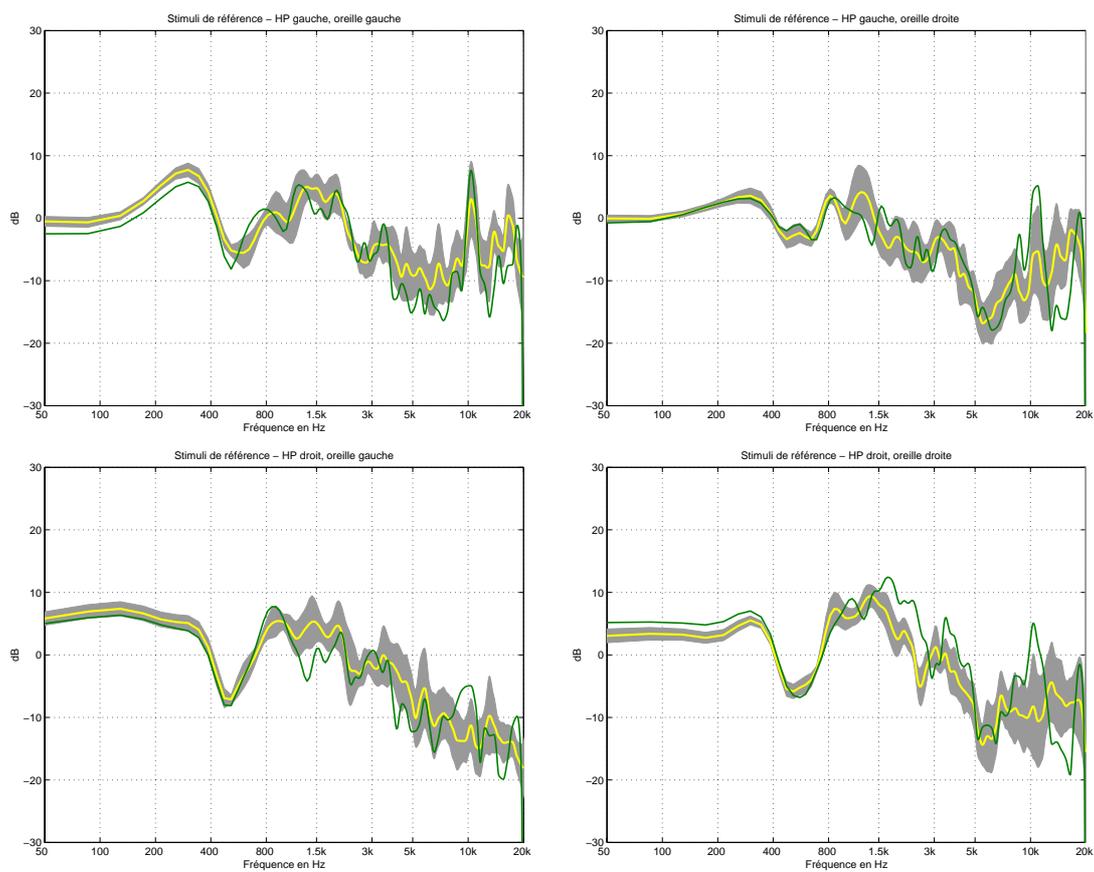


FIGURE II.18 – Stimuli de référence, mesurés dans l'habitacle d'un véhicule

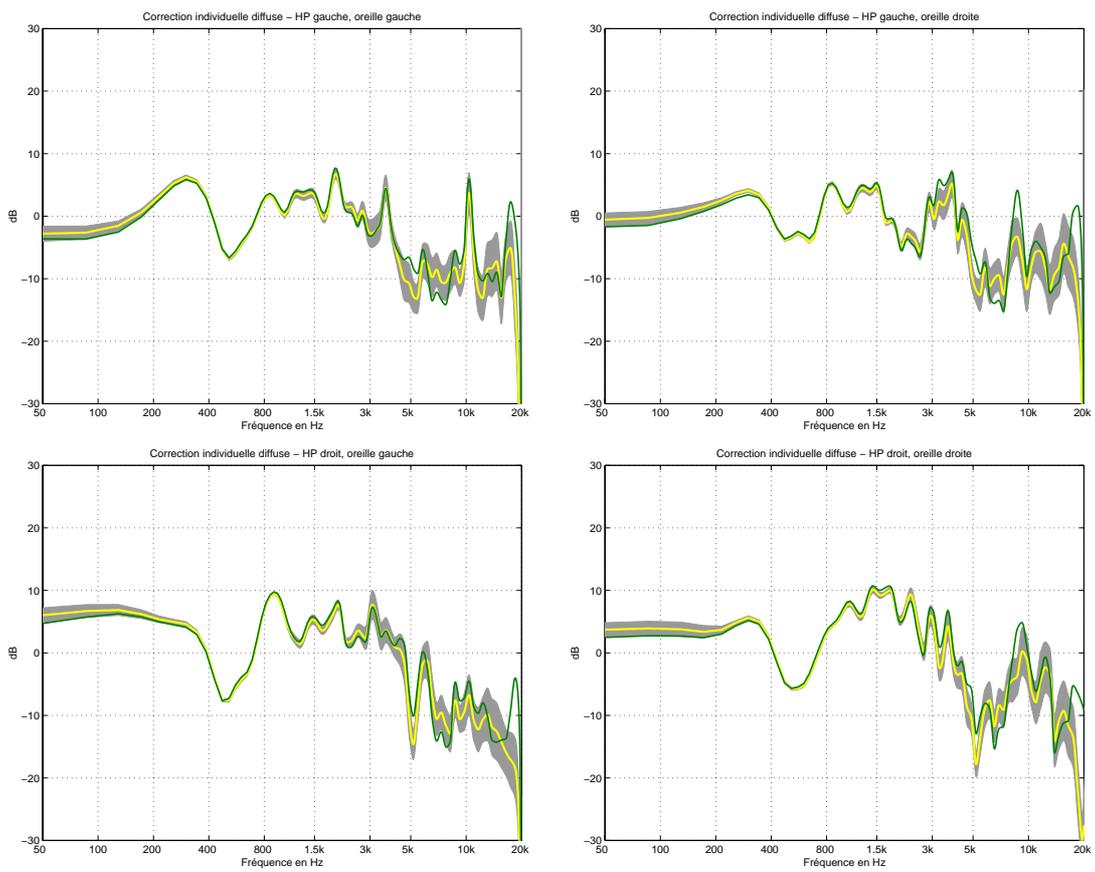


FIGURE II.19 – Stimuli intégrant la correction individuelle diffuse, mesurés dans l'habitacle d'un véhicule

individus non-experts. Son seul inconvénient est la croissance du nombre de paires à explorer, qui devient vite important ($n \cdot (n - 1) / 2$ paires avec n stimuli). En revanche, son déroulement peut être complètement automatisé : l'ordre de présentation des paires ainsi que l'ordre de présentation des stimuli à l'intérieur des paires peuvent être rendus totalement aléatoires, et inconnus à la fois du sujet et de l'examineur. On parle alors de test en *double aveugle*.

Ce type de test permet d'extraire des distances entre stimuli. Dans notre cas, nous en déduisons le stimulus le plus proche de l'écoute réelle par les haut-parleurs.

II.4.6 Protocole expérimental

Le test s'est déroulé à l'intérieur d'un véhicule à l'arrêt, et isolé des bruits environnants. Les sujets étaient assis à l'intérieur du véhicule à la position du conducteur. Tout le dispositif électronique et informatique était placé hors du véhicule. Seul un écran plat, comportant l'interface du test, était fixé sur le volant ; un clavier informatique et une souris de type « trackball » étaient posés sur les jambes du sujet.

Le casque utilisé pour le test est le modèle K₁₀₀₀ de marque AKG ; nous avons ouvert les oreillettes orientables, leur donnant un angle de 40° par rapport à leur position au repos. Rappelons que ce casque a été choisi pour privilégier le sentiment d'extériorisation.

Les sujets ont toujours porté le casque sur la tête, même dans le cas de la diffusion du signal par les haut-parleurs de la voiture, ceci afin d'effectuer le test en aveugle pour l'auditeur. En effet, le but de ce test d'écoute est de comparer l'écoute naturelle et l'écoute sur casque. L'écoute comparative doit être quasi simultanée, afin de bénéficier de notre courte mémoire auditive. Comme il n'est pas envisageable de demander aux sujets de retirer le casque pour écouter le système électroacoustique réel (ce qui reviendrait à les renseigner sur le stimulus écouté), et comme le K₁₀₀₀ n'est pas, malgré tout, acoustiquement transparent, il est nécessaire d'enregistrer les BIR avec le casque d'écoute placé sur la tête des individus, le signal test provenant bien des haut-parleurs.

En toute rigueur, la présence du casque sur la tête a effectivement une influence sur la fonction de transfert mesurée, comme on peut le constater sur la figure II.20, même si celui-ci est totalement ouvert. En dessous de 1 kHz, la perturbation est inférieure à 2 dB ; au-dessus de 1 kHz, elle est en moyenne inférieure à 3 dB. L'influence du casque n'est pas la même sur chaque fonction de transfert (le signal capté par l'oreille gauche est nettement perturbé autour de 2 kHz, alors que celui capté par l'oreille droite est largement perturbé entre 1 et 3 kHz) ; cette remarque nous conforte dans notre décision de laisser le casque sur la tête des individus pendant la mesure des BIR.

Le signal source convolué a été choisi de manière à couvrir le plus largement possible le spectre audible. Nous avons opté pour un morceau de musique Pop², dont un fragment de deux minutes était diffusé en boucle.

Ce signal est convolué en temps-réel avec les différentes fonctions de transfert, à l'aide de cartes électroniques spécialement dédiées. Le signal devant être diffusé par les haut-parleurs de la voiture, ne devant subir aucun traitement spécifique, est cependant également véhiculé par ces cartes, dans un souci de transparence totale du système entre l'écoute naturelle et l'écoute au casque.

Afin de réaliser un test en double aveugle, on se servira d'un dispositif de commutation automatique (matrice de commutation AKAI) pour aiguiller le signal source soit vers les haut-parleurs de la voiture, soit vers le casque d'écoute. Ce dispositif est commandé par des signaux MIDI, et peut donc être programmé pour que l'ordre des stimuli soit totalement aléatoire, et inconnu de l'opérateur. La commutation est totalement transparente pour le sujet.

2. extrait de l'album « Love Over Gold », du groupe Dire Straits

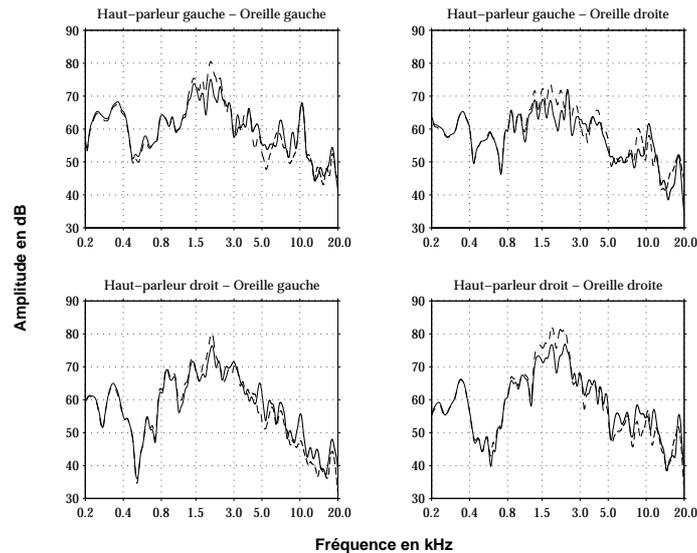


FIGURE II.20 – Influence de la présence du casque sur la tête du mannequin, pendant les mesures (trait plein : mesures sans casque ; trait pointillé : mesures avec casque, dans la même position du mannequin)

Nous devons également faire en sorte que tous les stimuli arrivent aux oreilles de l'auditeur avec un niveau sonore identique. En effet, ce paramètre est crucial dans une comparaison psychoacoustique : il peut générer des jugements de timbre différents. Une procédure d'égalisation en niveau des stimuli convolués a donc été appliquée : elle opère dans la bande [0-500] Hz, là où tous les stimuli convolués sont globalement identiques, à un gain près (puisque l'oreille humaine est globalement omnidirectionnelle en basse fréquence [20]).

Dix sujets ont passé le test. Ils ont été choisis parmi le personnel de l'IRCAM et de RENAULT, et ont chacun une certaine culture musicale, sans être de véritables experts en tests d'écoute.

Les tests de comparaison n'ont pas été effectués sous la forme A/B/X (choix de X entre A ou B), car, dans ce type de test, la notion de dissemblance subjective est déduite de la statistique ; or celle-ci est inefficace avec un nombre aussi restreint de sujets. Nous avons donc conduit les tests sous la forme A/B (comparaison par paire), où la notion de dissemblance est explicitement demandée : les sujets écoutent les stimuli A et B, et doivent juger de la différence du stimulus B par rapport au stimulus A. Il y a 7 stimuli à comparer, ce qui fait 21 paires différentes (voir tableau II.3).

Comme nous voulons également savoir de quelle manière les différences ont été jugées, nous avons choisi de simplifier l'extraction de dimensions perceptives, en les fixant à l'avance. Nous décidons donc d'explorer les axes suivants :

- **l'extériorisation**, ou la sensation de percevoir le son plus ou moins à l'intérieur de la tête ;
- **la localisation**, ou la précision de discrimination spatiale de différentes sources ;
- **le timbre**, ou la présence de colorations fréquentielles.

La réponse à plusieurs questions simultanées portant sur des axes perceptifs différents étant délicate, et à la longue fatigante, pour le sujet nous avons choisi d'explorer de manière séquentielle les trois axes perceptifs. Chaque sujet a procédé à l'écoute de toutes les paires, a testé une composante perceptive, puis a réécouté toutes les paires dans un ordre différent pour tester la composante suivante.

L'interface du test est présentée figure II.21. Le positionnement du curseur est traduit par une note entre 0 (très similaire) et 10 (très dissemblable). En ce qui concerne l'extériorisation, le traitement est un

TABLEAU II.3 – Ensemble des 21 paires retenues pour le test de validation de la plate-forme de restitution en laboratoire ; la description des noms est donnée dans le tableau II.2

Paire	Stimulus A	Stimulus B
1	Ref_autre	Ref_ind
2	Diff_ind	Ref_ind
3	Diff_ind	Ref_autre
4	Diff_autre	Ref_ind
5	Diff_autre	Ref_autre
6	Diff_autre	Diff_ind
7	Ref_tête	Ref_ind
8	Ref_tête	Ref_autre
9	Ref_tête	Diff_ind
10	Ref_tête	Diff_autre
11	ID_tête	Ref_ind
12	ID_tête	Ref_autre
13	ID_tête	Diff_ind
14	ID_tête	Diff_autre
15	ID_tête	Ref_tête
16	Voiture	Ref_ind
17	Voiture	Ref_autre
18	Voiture	Diff_ind
19	Voiture	Diff_autre
20	Voiture	Ref_tête
21	Voiture	ID_tête

peu moins direct : l'excursion du curseur est traduite de la même façon dans les deux sens, de sorte que l'on juge une différence d'extériorisation. Ce traitement n'est pas effectué dans le cas de l'analyse avec l'algorithme de Bradley-Terry (cf. § II.4.7.4).

Pour chaque sujet, le programme informatique de test fournit trois matrices de dissemblance, une par axe perceptif. On suppose qu'il n'y a pas de différence entre le stimulus A et lui-même, et que la dissemblance entre les stimuli A et B est identique à celle entre les stimuli B et A. Chaque matrice de dissemblance possède donc des 0 sur la diagonale, et sa forme est triangulaire (l'autre moitié étant reconstruite si besoin).

Remarquons que le jugement des sujets fournit une information de *dissemblance* et non de *distance*, puisque l'inégalité triangulaire ne sera pas forcément respectée ; nous sommes, pour l'instant, dans un espace non euclidien.

La durée moyenne du test a été de 60 minutes environ. Certains sujets ont effectué une pause entre les comparaisons concernant la deuxième et la troisième composante perceptive.

II.4.7 Analyse statistique classique

II.4.7.1 Impressions générales

Le test a été jugé difficile et délicat, notamment par le recouvrement des axes perceptifs. Tous les sujets ont constaté une différence globale d'extériorisation entre la diffusion par les haut-parleurs et la diffusion par le casque, et ont déclaré reconnaître à chaque fois la source de diffusion (haut-parleur ou casque).

À l'issue du test, il a été proposé à chaque sujet de réécouter un à un les différents stimuli, en aveugle, et de les classer par ordre de proximité à la diffusion par les haut-parleurs de la voiture. Tous ont placé leur référence individuelle en premier. Puis vient un groupe, mené par le stimulus de référence d'un autre individu, au sein duquel il est difficile de faire la différence entre la référence de la tête artificielle et les corrections diffuses (individuelles ou non). La correction ID proposée par HEAD ACOUSTICS arrive en dernière place.

Un deuxième classement a été effectué avec un signal de parole. Il ne diffère pas du précédent.

II.4.7.2 Analyse des distributions des jugements de dissemblance

Les trois distributions de l'ensemble des dix sujets sur l'ensemble des vingt et une paires sont présentées figure II.22. Aucune des distributions n'est gaussienne, ce qui montre que l'ensemble des sujets n'a pas répondu au hasard. Pour chaque attribut perceptif, les notes comprises entre 0 et 1, reflétant une dissemblance quasi nulle, forment le plus grand ensemble, spécialement pour la localisation ; ceci confirme la difficulté du test.

Pour l'extériorisation, la distribution est à peu près régulière, mis à part une forte proportion de notes entre 3 et 4. Pour la localisation, la distribution est également à peu près régulière : les différences d'extériorisation et de localisation sont donc à peu près homogènes. Il n'en est pas de même pour le timbre, où le nombre de notes inférieures à 5 est élevé et décroît régulièrement, soulignant les faibles différences de timbre pour un nombre important de paires.

Regardons ensuite les distributions des réponses de chaque individu. Elles ne sont pas uniformes, traduisant le fait que les sujets ont bel et bien remarqué des différences entre les paires, et n'ont pas répondu au hasard. Les moyennes entre les paires varient assez fortement d'un individu à l'autre, surtout pour l'extériorisation et la localisation ; il en est de même pour les écarts types. Le tableau II.4 rassemble les résultats.

Nous mettons en évidence des différences importantes dans les comportements des différents sujets : certains sont plus sensibles que d'autres à tel ou tel aspect, ce qui se traduit par une moyenne des notes

Indiquer de quelle maniere B est different de A

Moins exteriorise Plus exteriorise

Commentaire

Ecoute comparative

Cliquer sur A, B, puis enregistrer la reponse

Commandes

Interface pour le jugement de l'exteriorisation

Indiquer de quelle maniere B est different de A

Localisation identique Localisation differente

Commentaire

Ecoute comparative

Cliquer sur A, B, puis enregistrer la reponse

Commandes

Interface pour le jugement de la localisation

Indiquer de quelle maniere B est different de A

Timbre identique Timbre different

Commentaire

Ecoute comparative

Cliquer sur A, B, puis enregistrer la reponse

Commandes

Interface pour le jugement du timbre

FIGURE II.21 – Interface du test d'ecoute realise pour valider la plate-forme de restitution binaurale

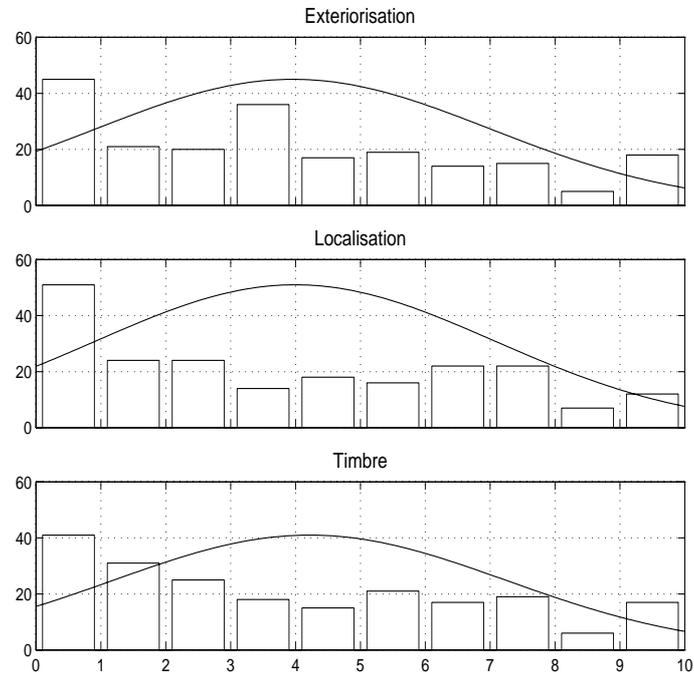


FIGURE II.22 – Distributions de l'ensemble complet des jugements de dissemblance, sur toutes les paires et tous les sujets ; superposition d'une gaussienne de même moyenne et de même écart type.

TABLEAU II.4 – Moyennes et écarts types de l'ensemble des individus sur toutes les paires

Extériorisation	1,3 < moyennes < 5,5 1,6 < écarts types < 4,0
Localisation	1,7 < moyennes < 5,9 1,6 < écarts types < 3,7
Timbre	2,0 < moyennes < 4,2 1,9 < écarts types < 3,6

TABLEAU II.5 – Moyennes et écarts types de l'ensemble des paires sur tous les individus

Extériorisation	1,3 < moyennes < 7,1 1,6 < écarts types < 4,3
Localisation	1,3 < moyennes < 7,6 1,8 < écarts types < 3,6
Timbre	2,4 < moyennes < 7,5 1,5 < écarts types < 3,6

plus ou moins importante. D'autre part, l'échelle n'est pas utilisée de manière homogène par tous. Il est donc nécessaire de *normaliser* les notes de chaque individu par leur variance, ce qui revient à donner à chaque individu la même importance quelle que soit sa dispersion : on rétablit l'équilibre entre les individus en leur donnant à tous une variance unité.

Nous constatons également que les corrélations entre individus sont faibles : les individus n'ont donc pas tous traduit les différences de la même manière. Remarquons que nous travaillons dans un espace perceptif non euclidien, et que le calcul des corrélations entre individus ne nous permet pas de déduire des stratégies de jugement, mais seulement une manière d'utiliser la notation. De plus, le coefficient de corrélation est sensible aux valeurs extrêmes [18].

L'examen des distributions de chaque paire sur tous les individus ne fait pas apparaître de forme gaussienne. C'est un point important : le nombre de sujets est trop faible pour appliquer le théorème de la limite centrale, et supposer la distribution gaussienne. Nous confirmons notre observation par un test de comparaison avec une distribution gaussienne idéale, de même moyenne et de même écart type ; ce test fait intervenir une loi du χ^2 à 9 degrés de liberté. En choisissant comme hypothèse nulle l'identité entre les deux lois, et un facteur de risque de 10 %, nous sommes obligés de considérer toutes les distributions testées comme non gaussiennes.

Sans surprise, on constate que les moyennes interindividuelles des notes varient fortement entre les paires ; il en est de même pour les écarts types. Le tableau II.5 rassemble les résultats. Remarquons que les écarts types sont importants, en regard de la moyenne, ce qui peut laisser présager une certaine difficulté dans l'extraction de résultats fiables.

L'examen détaillé des moyennes et écarts types interindividuels lève cependant un coin du voile sur certains résultats. Les paires contenant le stimulus 'Voiture' ou le stimulus 'ID_tête' ont toutes une moyenne élevée, ce qui traduit de fortes différences entre les stimuli, et un écart type élevé concernant l'extériorisation, ce qui traduit de fortes variations. Par contre, les paires contenant le stimulus 'Ref_tête' présentent une moyenne et un écart type nettement moins importants ; plus particulièrement, les paires ('Ref_tête' / 'Diff_autre') et ('Voiture' / 'ID_tête') possèdent une moyenne et un écart type très faibles, ce qui laisse présager une très faible différence perceptive entre les stimuli correspondant.

Concernant l'extériorisation, les individus ont relevé de faibles différences entre les stimuli 'Diff_ind' et 'Diff_autre'. Pour la localisation, les différences sont faibles entre les stimuli 'Diff_ind' et 'Ref_tête'. Enfin pour le timbre, il en est de même pour les stimuli 'Diff_ind' et 'Ref_tête', ainsi que 'Diff_ind' et 'Diff_autre'.

L'examen des notes attribuées aux paires constituées du stimulus 'Voiture' et d'un autre stimulus est détaillé figure II.23. On constate une mauvaise concordance des jugements entre les individus.

Terminons l'analyse des distributions par la recherche de l'intervalle contenant les moyennes et écarts types *théoriques* des distributions des notes des différentes paires, sur tous les individus. Ceci suppose que les distributions sont gaussiennes, mais on peut considérer les résultats fiables lorsque l'effectif de la population est au moins égal à une dizaine d'individus [18].

Soit m l'estimation de la moyenne des n échantillons considérés, μ la moyenne théorique inconnue,

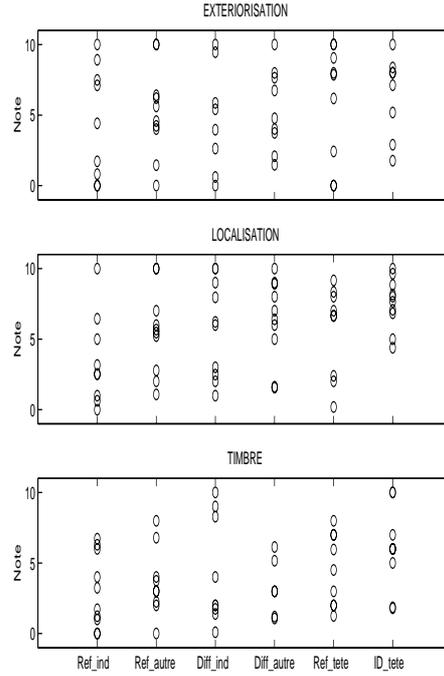


FIGURE II.23 – Notes des paires contenant le stimulus ‘Voiture’ et un autre stimulus, pour tous les individus

s l’estimation de l’écart type, et σ l’écart type théorique inconnu. Soit $T_{n-1;\alpha/2}$ le fractile au risque α d’une loi T de Student à $n - 1$ degrés de libertés, et $\chi_{n-1;\alpha/2}^2$ celui d’une loi du χ^2 à $n - 1$ degrés de libertés. D’après [18], la véritable moyenne μ de la distribution est telle que

$$m - \frac{sT_{n-1;\alpha/2}}{\sqrt{n-1}} \leq \mu \leq m + \frac{sT_{n-1;\alpha/2}}{\sqrt{n-1}}, \quad (\text{II.16})$$

et la véritable variance σ^2 est telle que

$$\frac{(n-1)s^2}{\chi_{n-1;\alpha/2}^2} \leq \sigma^2 \leq \frac{(n-1)s^2}{\chi_{n-1;1-\alpha/2}^2}. \quad (\text{II.17})$$

La figure II.24 présente les résultats concernant les moyennes et les écarts types des distributions des notes. Les quantités estimées sont représentées par des *, tandis que les intervalles contenant les quantités théoriques sont représentés par des barres verticales. Le risque est choisi égal à 5 %.

Compte tenu du faible nombre de sujets, et de la forte variabilité interindividuelle, les intervalles contenant les moyennes théoriques sont grands. Pour la paire n° 10 (‘Ref_tête’ / ‘Diff_autre’), concernant l’extériorisation et la localisation, l’intervalle contenant la moyenne théorique contient aussi la valeur nulle ; d’après ce test, il se peut que ces deux stimuli soient perçus de manière identique. Il en est de même pour la paire n° 9 (‘Ref_tête’ / ‘Diff_ind’) concernant la localisation. Les trois courbes représentant les moyennes présentent un saut à partir de la paire n° 11 jusqu’à la paire n° 21, ce qui traduit une dissemblance perceptive nettement plus importante pour les paires contenant le stimulus ‘ID_tête’ et le stimulus ‘Voiture’. Le saut est très nettement visible concernant le timbre.

Comme pour les moyennes théoriques, les intervalles contenant les écarts types théoriques sont importants. L’absence de symétrie est conséquence de la forme de la distribution du χ^2 utilisée pour calculer les intervalles. Comme déjà mentionné, de telles valeurs d’écarts types, en regard de celles des moyennes, ne peuvent que compliquer l’extraction de résultats fiables, surtout en ce qui concerne le timbre.

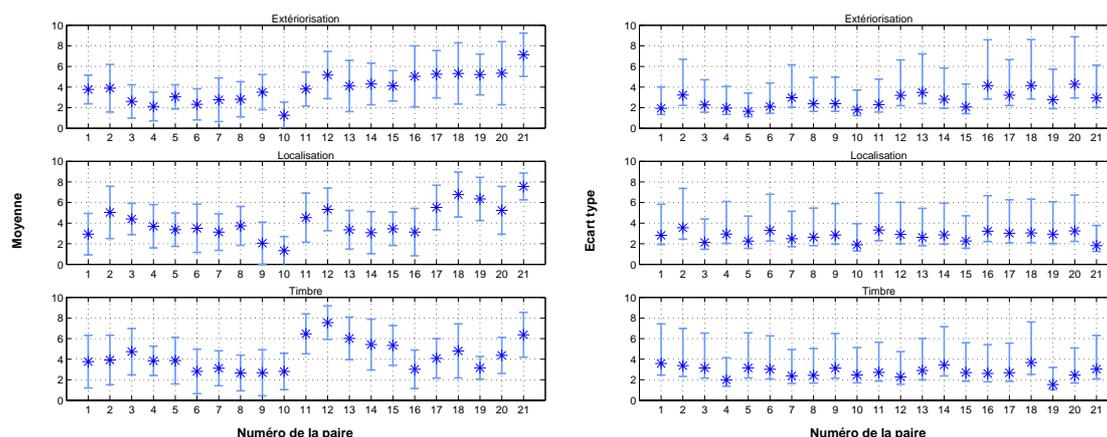


FIGURE II.24 – Encadrement contenant la moyenne théorique (à gauche) et l'écart type théorique (à droite) des distributions des dissemblances entre les stimuli, pour l'ensemble des individus. En abscisse : le numéro de la paire, en ordonnée : la valeur de la dissemblance

II.4.7.3 Analyse de variance

À ce stade, nous devons nous demander dans quelle mesure les résultats de notre test d'écoute sont significatifs, c.-à-d. si les valeurs de dissemblance ne reflètent pas plus une variance entre les individus qu'une réelle variance entre les stimuli. En d'autres termes, nous devons regarder l'ensemble des valeurs de dissemblance, pour chaque axe perceptif, et nous demander si la variance globale est pilotée par la variance entre individus ou entre stimuli. Globalement :

- si tous les individus ont répondu de la même manière (et ont réagi identiquement aux mêmes caractéristiques), la variance globale s'explique par la variance inter stimuli : le test est hautement significatif ;
- si tous les individus ont répondu de manière dispersée, la variance globale s'explique par la variance inter sujets : le test est très peu significatif .

La méthode est décrite dans l'annexe B.1.2. Nous effectuons le test pour l'ensemble des stimuli (soit 21 paires), mais aussi en enlevant les deux stimuli pour lesquels les dissemblances avec tous les autres apparaissent déjà comme importantes, c.-à-d. le stimulus 'ID_tête' et le stimulus 'Voiture'. Comme nous l'avons préconisé, les dissemblances sont normalisées par la variance des réponses de chaque individu. Le tableau II.6 montre le pourcentage de chance pour que la variance globale soit égale à la variance entre les stimuli. Pour l'extériorisation et la localisation, les valeurs de dissemblances restent hautement significatives, même si on enlève les deux stimuli présentant les plus fortes dissemblances avec tous les autres. Il n'en est pas de même en ce qui concerne le jugement du timbre, puisque si on enlève le stimulus 'ID_tête' (et *a fortiori* si on enlève aussi le stimulus 'Voiture'), la variance entre les stimuli restant n'explique que très faiblement la variance globale. Cela veut dire qu'il est très difficile de mettre en évidence des différences de timbre significatives entre les stimuli suivant : 'Ref_ind', 'Ref_autre', 'Diff_ind', 'Diff_autre', et 'Ref_tête'.

Notons que si on enlève le stimulus 'Voiture', le stimulus le plus proche 'Ref_ind' et le stimulus le plus éloigné 'ID_tête', la significativité du test tombe en dessous de la barre des 95 %.

II.4.7.4 Ordonnement des stimuli

Un premier classement des différents stimuli peut être effectué à l'aide de la procédure de Bradley-Terry (voir annexe B.1.1). Cette procédure est applicable uniquement si l'on dispose d'un attribut séman-

TABLEAU II.6 – Analyse de variance : test de comparaison entre la variance globale et la variance entre les stimuli

Stimuli	Tous	Sans Voiture	Sans ID_tête	Sans Voiture et sans ID_tête
Extériorisation	100,0 %	99,9 %	99,9 %	98,2 %
Localisation	100,0 %	98,0 %	100,0 %	96,0 %
Timbre	100,0 %	100,0 %	60,3 %	39,0 %

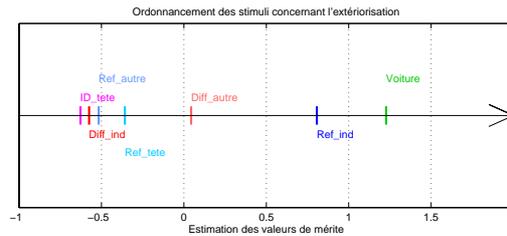


FIGURE II.25 – Ordonnement des stimuli concernant l'extériorisation sur une échelle continue et orientée à l'aide de l'algorithme de Bradley-Terry

tique non symétrique, c.-à-d. mesurable sur une échelle continue et orientée, ce qui est le cas avec le test de l'extériorisation. En effet, les sujets doivent juger si le stimulus A est plus ou moins extériorisé que le stimulus B, alors que pour la localisation et le timbre, on demande de quantifier la différence.

Nous transformons chaque matrice de dissemblance concernant l'extériorisation en une matrice de préférence de la façon suivante : une note comprise entre 0 et 5 signifie que le sujet a jugé le stimulus A plus extériorisé que le stimulus B, et nous attribuons la note 1 à la paire (A/B) ; dans le cas contraire, on attribue la note 0. Les matrices de préférences de chaque sujet sont alors sommées, pour obtenir une matrice globale, prenant en compte l'information fournie par chaque sujet. Nous classons les stimuli sur une échelle d'extériorisation.

Le résultat est fourni figure II.25. Le test de performance se révèle être positif à plus de 95 %. Le stimulus le plus extériorisé est sans surprise 'Voiture', suivi par 'Ref_ind'. Vient ensuite 'Diff_autre', puis un groupe assez compact emmené par 'Ref_tête'. Par contre, il est étonnant que 'Diff_autre' arrive en troisième position.

Cependant, il faut remarquer que le nombre de sujets ayant passé le test est faible. Bien que le test de performance se soit révélé positif, les résultats fournis par cette procédure ne sont pas stables, et l'ordre de certains stimuli peut changer selon le poids accordé aux sujets ('Voiture' et 'Ref_ind' restent néanmoins toujours en tête). Il se peut que l'on ait du mal à établir un ordre valide. Pour confirmer cette hypothèse, nous appliquons une procédure de *bootstrap* [68] qui consiste à tirer n fois au sort un individu, parmi l'ensemble des n sujets (voir annexe B.1.3) ; en effectuant cette opération un très grand nombre de fois (quelques centaines), on construit autant d'échantillons sur lesquels on applique l'algorithme de Bradley-Terry. Cela *simule* un nombre de sujets plus élevé, à partir desquels on obtient une solution bien plus stable qu'avec une population restreinte. C'est donc un moyen de tester l'identité des p_{ij} .

La figure II.26 montre les résultats obtenus en construisant 250 échantillons de bootstrap. Pour chaque échantillon, nous avons appliqué l'algorithme en question, et mémorisé les valeurs de mérite qui permettent de classer les stimuli ; nous représentons ici les histogrammes de ces valeurs de mérite. Comme on le présuait, deux groupes apparaissent :

- 'Voiture' et 'Ref_ind' ;
- 'Diff_autre', 'Ref_tête', 'Ref_autre', 'Diff_ind' et 'ID_tête'.

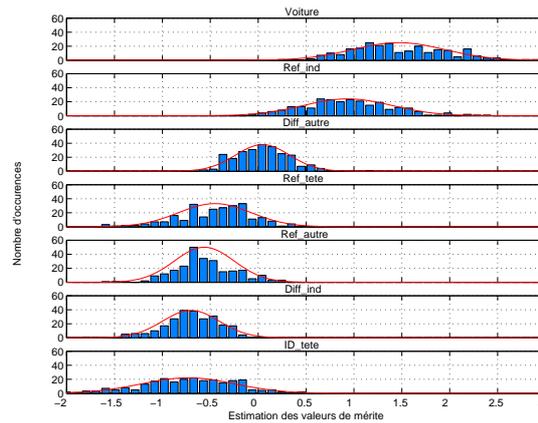


FIGURE II.26 – Ordonnement des stimuli concernant l'extériorisation à l'aide de l'algorithme de Bradley-Terry, effectué sur 250 échantillons de bootstrap

TABLEAU II.7 – Pouvoir de discrimination entre deux stimuli contigus

Paire de stimuli	Voiture Ref_ind	Ref_ind Diff_autre	Diff_autre Ref_tête	Ref_tête Ref_autre	Ref_autre Diff_ind	Diff_ind ID_tête
Probabilité de distinction	27.1 %	72.6 %	17.7 %	8.4 %	7.5 %	0.1 %

La distinction au sein de ces deux groupes peut être impossible puisqu'il y a recouvrement entre les distributions. En assimilant les distributions à des gaussiennes, on peut effectuer un test d'hypothèse, et calculer une erreur de type II [18] : il suffit de calculer la probabilité de détecter la moyenne du stimulus i lorsque la moyenne prise en compte est celle du stimulus $i+1$. Cette probabilité quantifie la *puissance* du test (capacité à discriminer deux stimuli dont les distributions se recouvrent). Nous obtenons le tableau II.7.

La puissance du test est relativement élevée entre 'Ref_ind' et 'Diff_autre', ce qui confirme la scission en deux groupes. Cependant, la discrimination au sein de ces deux groupes de stimuli semble très difficile, comme le laissait présager d'ailleurs l'analyse de variance.

En conclusion, l'algorithme d'ordonnement de Bradley-Terry nous montre que le stimulus restituant le mieux le sentiment d'extériorisation est sans surprise la référence individuelle. Par contre, il ne nous permet pas de choisir entre la référence brute du mannequin et une quelconque correction diffuse.

II.4.7.5 Conclusion

L'analyse statistique classique nous permet de dresser des constatations locales. Nous constatons que l'échelle de notation est utilisée de manière non homogène par les différents sujets, et nous décidons de normaliser chaque matrice de dissemblance par sa variance. Nous constatons également que les notes extraites de la matrice ne sont pas toujours en accord avec le classement demandé *ex nihilo*, et que les jugements ne sont pas toujours concordants entre les sujets.

L'examen des distributions des notes (§ II.4.7.2) permet de dégager des dissemblances importantes entre certains stimuli, et faibles entre d'autres. En particulier, les paires contenant le stimulus 'Voiture' et celles contenant le stimulus 'ID_tête' présentent de grandes différences, alors que les paires contenant le stimulus 'Ref_tête' présentent des différences plus faibles, notamment la paire ('Ref_tête' / 'Diff_autre').

L'analyse de variance (§ II.4.7.3) met en évidence la difficulté du test, puisque les différences de timbre constatées entre tous les stimuli sans 'Voiture' sont plus à mettre sur le compte d'une variance

entre les individus qu'à une variance entre stimuli. La significativité du test chute encore si on enlève le stimulus 'ID_tête'.

Enfin, l'algorithme de Bradley-Terry (§ II.4.7.4) appliqué sur les jugements concernant l'extériorisation, a établi un ordre de proximité au stimulus 'Voiture'. Le stimulus le plus proche est sans conteste notre propre référence individuelle, suivie d'un groupe au sein duquel la distinction est délicate.

II.4.8 Analyse multidimensionnelle

Cette partie traite de l'emploi de méthodes statistiques plus élaborées, c.-à-d. l'analyse multidimensionnelle, ou MDS (MultiDimensional Scaling), déclinée sous trois formes différentes :

- MDSCAL, prenant en compte les résultats d'un seul sujet à la fois, ou une moyenne des sujets ;
- INDSCAL et EXSCAL, prenant en compte les résultats de chacun des sujets ayant passé le test, avec différents degrés de raffinement.

II.4.8.1 Introduction

Nous avons souhaité effectuer un test subjectif conduit sous la forme de comparaisons par paires, ayant pour but de déterminer un ordre sur une série de stimuli. Par souci de simplicité quant à l'extraction des résultats, nous avons déjà séparé les axes perceptifs (extériorisation, localisation et timbre) ; ceux-ci ont été testés séparément. Nous nous situons de fait dans un contexte bien précis, réclamant une analyse adaptée.

Les méthodes exposées précédemment constituent une phase indispensable dans le dépouillement des données, mais cependant insuffisante : en effet, l'étude séparée de chaque individu, ou de chaque dissemblance, laisse de côté les liaisons pouvant exister entre les individus, ou entre les dissemblances, et qui sont souvent l'aspect le plus important. Il faut donc analyser les données en tenant compte de leur caractère multidimensionnel. Il existe pour cela deux familles de méthodes : l'analyse factorielle sur des tableaux de variables ou AF, et l'analyse à échelle multidimensionnelle ou MDS (MultiDimensional Scaling). Ces méthodes ont le même objectif : trouver une configuration de n individus dans un espace de faible dimension. Cependant, les données de départ sont différentes :

- l'analyse factorielle s'exerce sur un ensemble de données multivariées : chaque sujet note une série d'attributs sémantiques ;
- alors que l'analyse MDS suppose que l'on connaît, pour chaque individu, les $\frac{t(t-1)}{2}$ dissemblances entre les t variables.

Le modèle qui sous-tend l'analyse factorielle est basé sur des angles entre vecteurs, alors que celui concernant l'analyse MDS est basé sur des distances entre points, ce qui est plus simple à interpréter. D'autre part, l'analyse factorielle résulte en un plus grand nombre de dimensions, car elle suppose l'existence de relations linéaires entre les variables ; cette hypothèse est assez sévère en regard de données perceptives.

Un autre point important doit être souligné : la récolte de dissemblances entre objets est toujours peu influencée par l'examineur, alors que l'attribution de notes sur une liste d'attributs peut ne pas l'être ; de plus, celle-ci n'est pas toujours significative, à cause du choix des attributs.

Pour notre test, nous choisissons l'analyse MDS. Les dissemblances δ entre les t configurations écoutées permettent à l'analyse de les placer dans un espace à r dimensions ($r \leq t - 1$) pour que les distances entre ces configurations respectent les dissemblances données. De la même manière, elle placerait les villes les unes par rapport aux autres, à partir d'un tableau de distances entre villes. Les configurations qui ont été estimées peu différentes (δ_{jk} petit) seront proches dans cet espace, par contre celles estimées très différentes (δ_{jk} grand) seront éloignées dans ce même espace. Le nombre de dimensions est une inconnue levée à l'aide d'une procédure itérative comportant un critère de décision *a posteriori*.

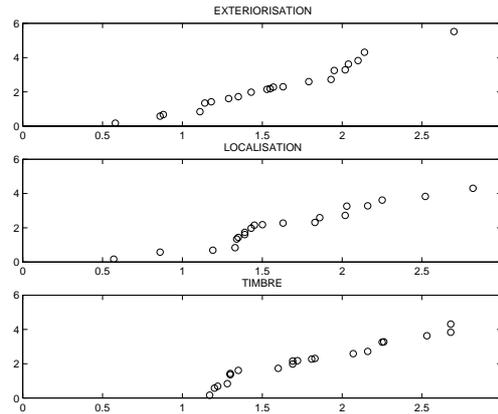


FIGURE II.27 – Diagrammes de Shepard de la solution MDSCAL à une dimension. En abscisse : les valeurs de dissimilarités initiales. En ordonnée : les valeurs de distances entre les points de la configuration finale

Si l'on effectue l'analyse sur une seule matrice de dissimilarité à la fois, donc sur un seul individu ou sur la moyenne des individus, on utilise l'analyse MDSCAL. Si l'on prend en compte séparément les matrices de tous les individus, on utilise l'analyse INDSCAL (INDividual Differences SCALing), ou mieux l'analyse EXSCAL (EXTended INDSCAL).

La première étape indispensable, commune à toutes ces méthodes, consiste à transformer en distances euclidiennes (d) les données collectées sous forme de dissimilarités perceptives (δ). Il s'agit donc de leur faire respecter les trois propriétés suivantes :

1. $d(A_i, A_i) = 0$;
2. $d(A_i, A_j) = -d(A_j, A_i)$;
3. $d(A_i, A_j) + d(A_j, A_k) \geq d(A_i, A_k)$.

Les deux premières sont supposées vraies par construction du test, puisqu'avec t objets on ne soumet au test que $\frac{t(t-1)}{2}$ paires. La troisième réclame une attention particulière, car elle n'est, en général, pas vérifiée ; dans ce cas, la méthode de la constante additive permet d'en déduire une distance euclidienne. Il existe en effet une constante c , telle que $d_{ij} = \delta_{ij} + c$ avec $d_{ii} = 0$, soit euclidienne. On a coutume de choisir $c = \max_{jkl} (\delta_{jl} - \delta_{jk} - \delta_{kl})$.

II.4.8.2 Analyse sur la moyenne des sujets

Cette analyse consiste à utiliser la méthode MDSCAL avec la moyenne des matrices de dissimilarités ; la méthode est décrite en annexe B.2.2. Concernant notre expérience, nous avons limité l'analyse à deux dimensions à cause du petit nombre de stimuli (il est en effet illusoire de chercher à décrire précisément un espace à trois dimensions ou plus avec seulement sept stimuli) ; de plus, nous n'avons pas cherché à nommer les dimensions perceptives, car cela dépasse le cadre de notre étude.

Cette méthode a été utilisée avec le programme SYSTAT^{®3} fonctionnant sous WindowsTM.

Les jugements des dix sujets, normalisés par leur écart type, ont été moyennés sur chacune des composantes perceptives testées, avant d'être injectés dans le programme. L'analyse en une dimension révèle un stress de 0,24 pour l'extériorisation, de 0,19 pour la localisation et de 0,27 pour le timbre. Les diagrammes de Shepard sont présentés figure II.27. Ils sont relativement réguliers, et attestent de la bonne adéquation du modèle aux données initiales.

Le résultat de l'analyse est présenté à la figure II.28. Concernant l'extériorisation, le stimulus le plus proche de l'écoute naturelle semble être 'Ref_autre', puis vient un groupe de stimuli non discernable,

3. <http://www.spss.com>

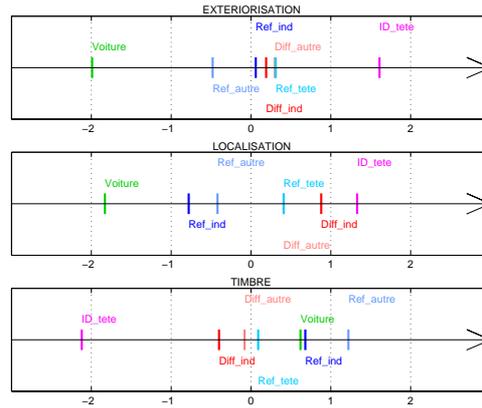
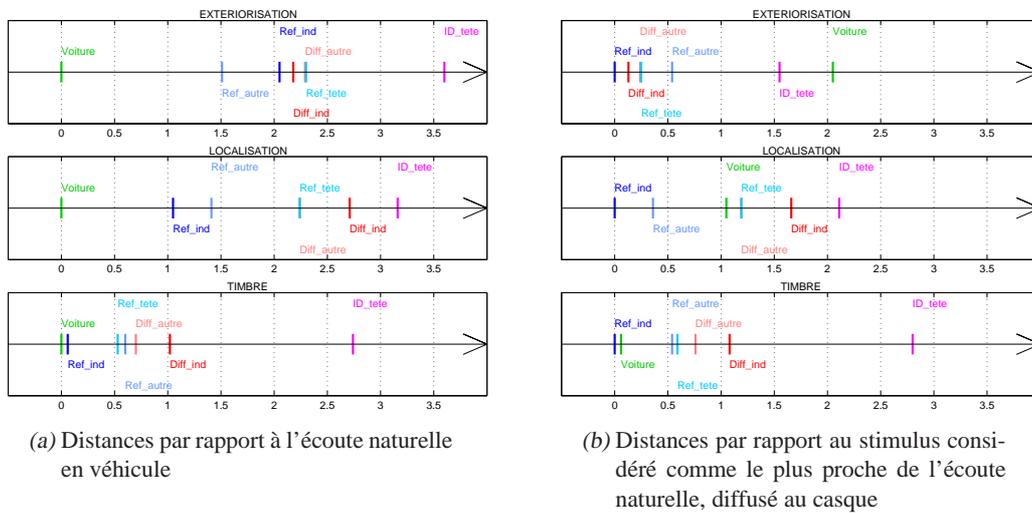


FIGURE II.28 – Résultat de l'analyse MDSCAL à une dimension



(a) Distances par rapport à l'écoute naturelle en véhicule

(b) Distances par rapport au stimulus considéré comme le plus proche de l'écoute naturelle, diffusé au casque

FIGURE II.29 – Distances calculées à partir des points de la configuration finale fournie par MDSCAL à une dimension

puis à l'autre extrémité la correction fournie par HEAD ACOUSTICS. Il est surprenant que le stimulus de référence d'un autre individu arrive en première place. Concernant la localisation, les positions des stimuli reflètent un ordre intuitif, le stimulus de référence du mannequin étant au moins aussi bien placé que les corrections diffuses. Concernant le timbre, le stimulus le plus proche de l'écoute naturelle est sans surprise la référence individuelle, puis la référence du mannequin, avant la référence non individuelle et les corrections diffuses ; la correction ID est très éloigné de l'ensemble des autres stimuli. La représentation à une dimension de la figure II.28 ne faisant pas toujours apparaître le stimulus 'Voiture' à une extrémité de l'axe (il s'agit seulement de coordonnées), nous lui préférons celle de la figure II.29, où les distances sont recalculées à partir des coordonnées de la configuration finale. Le stimulus de référence du mannequin est bien placé, faisant jeu égal avec les corrections diffuses.

Afin d'améliorer l'estimation de l'espace des stimuli, nous effectuons maintenant l'analyse en deux dimensions. Le stress est de 0,09 pour l'extériorisation, 0,08 pour la localisation et 0,12 pour le timbre ; il a diminué avec l'augmentation du nombre de dimensions, et indique une précision plus importante du modèle. Les diagrammes de Shepard sont présentés figure II.30. Comme pour la solution à une dimension, ils ne font pas apparaître de saut, et caractérisent un modèle assez précis.

La figure II.31 montre les résultats de l'analyse à deux dimensions, présentés sous forme de carte ;

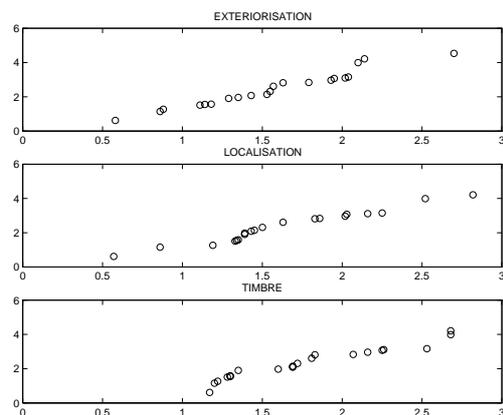


FIGURE II.30 – Diagrammes de Shepard de la solution MDSCAL à deux dimensions

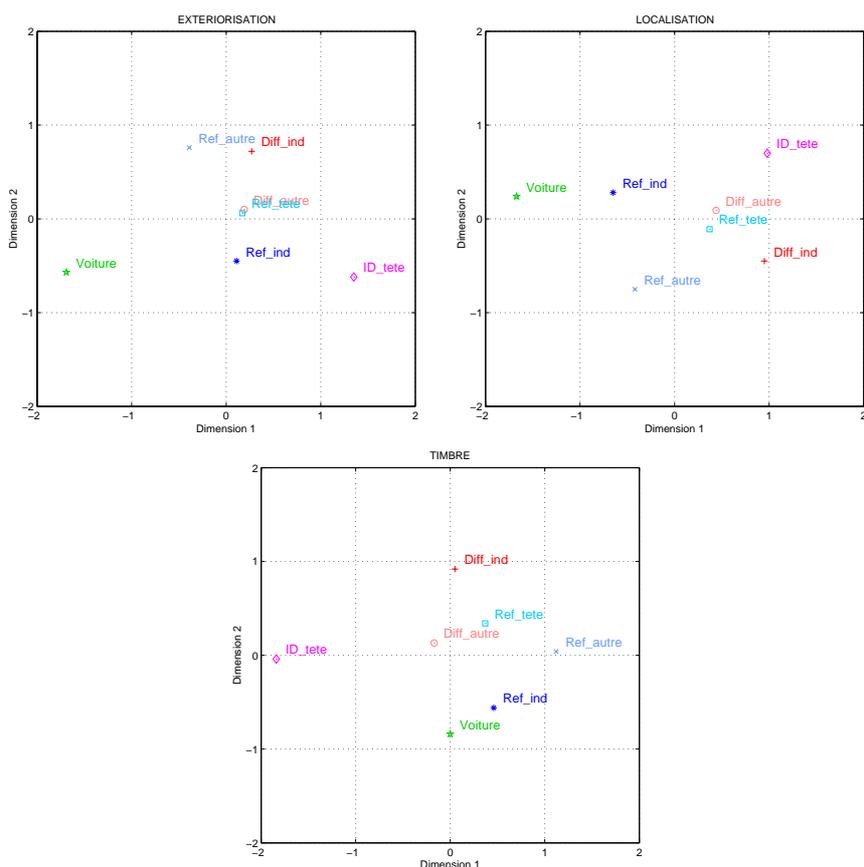


FIGURE II.31 – Résultat de l'analyse MDSCAL à deux dimensions

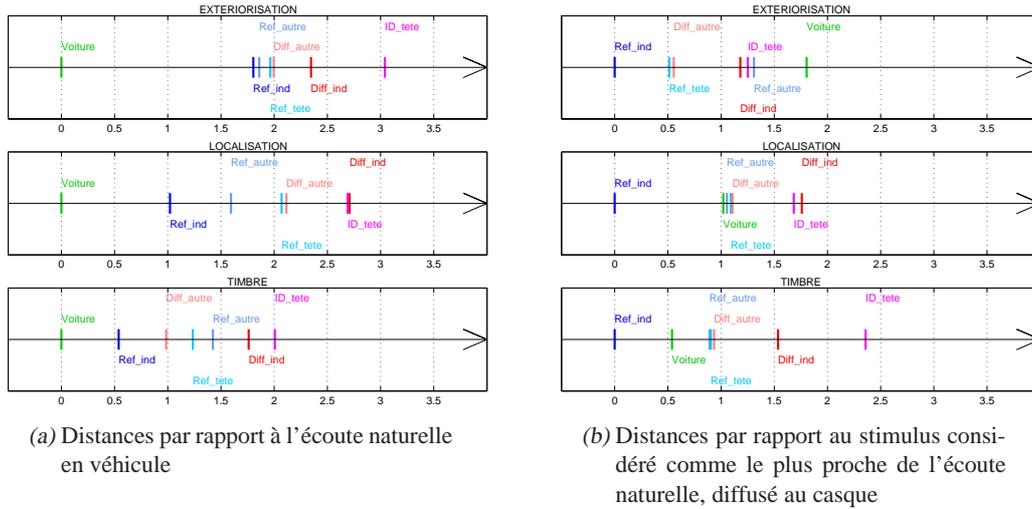


FIGURE II.32 – Distances calculées à partir des points de la configuration finale fournie par MDSCAL à deux dimensions

les distances entre points sont affichées sur la figure II.32. Concernant l’extériorisation, l’analyse en deux dimensions révèle que l’émergence du stimulus le plus proche de l’écoute naturelle est difficile à mettre en évidence (mis à part la référence individuelle) ; tous les stimuli diffusés au casque, à l’exception de la correction ID, sont à égale distance du stimulus ‘Voiture’. Ce résultat est à rapprocher de celui obtenu avec l’algorithme de Bradley-Terry au § II.4.7.4. Remarquons que le passage à deux dimensions a provoqué un déplacement du stimulus ‘Ref_autre’ à l’intérieur du groupe ; comme le stress a diminué, c’est cette situation que nous devons retenir.

Concernant la localisation, les deux stimuli les mieux placés sont la référence individuelle et la référence non individuelle. Puis vient la référence du mannequin et les corrections diffuses ; il semble, comme l’a montré l’analyse à une dimension, que la référence du mannequin soit meilleure que la correction diffuse individuelle.

Concernant le timbre, la correction ID est assez éloignée de l’ensemble des autres stimuli, ce qui corrobore l’impression générale évoquée par les participants à l’issue du test (voir § II.4.7.1). La référence du mannequin est un peu moins bien placée que la correction diffuse non individuelle, mais un peu mieux que la référence d’un autre individu.

Afin d’établir les différentes stratégies de jugement des individus, nous allons effectuer une analyse en composantes principales ou ACP (cf. § B.2.1) sur les résultats fournis par l’analyse MDSCAL à deux dimensions sur chaque individu [262]. Nous cherchons un espace commun à tous les individus, dans lequel nous pourrions placer les différents stimuli [26, 277].

À l’issue de l’analyse MDSCAL, chaque stimulus se trouve décrit par un ensemble de vingt coordonnées (deux par individu). Grâce à l’ACP, nous pouvons réduire ce nombre ; on choisit de le limiter à six. Nous cherchons à connaître le poids accordé par les individus à chacun de ces axes. Il est représentatif d’une manière de juger, puisque les axes fournis par l’ACP sont communs à tous les sujets. Pour cela, nous évaluons la corrélation entre chacun de ces six axes, et les deux axes servant à placer le stimulus pour chaque individu. Cette manière de procéder tient compte du fait que les résultats fournis par MDSCAL sont invariants par rotation.

D’après Saporta [277], la corrélation multiple R_m entre une variable y et p autres variables x^1, x^2, \dots, x^p est la valeur maximale prise par le coefficient de corrélation linéaire entre y et une combinaison linéaire des x^j . Si X désigne la matrice dont les colonnes sont les p variables x^1, x^2, \dots, x^p centrées, et si y est

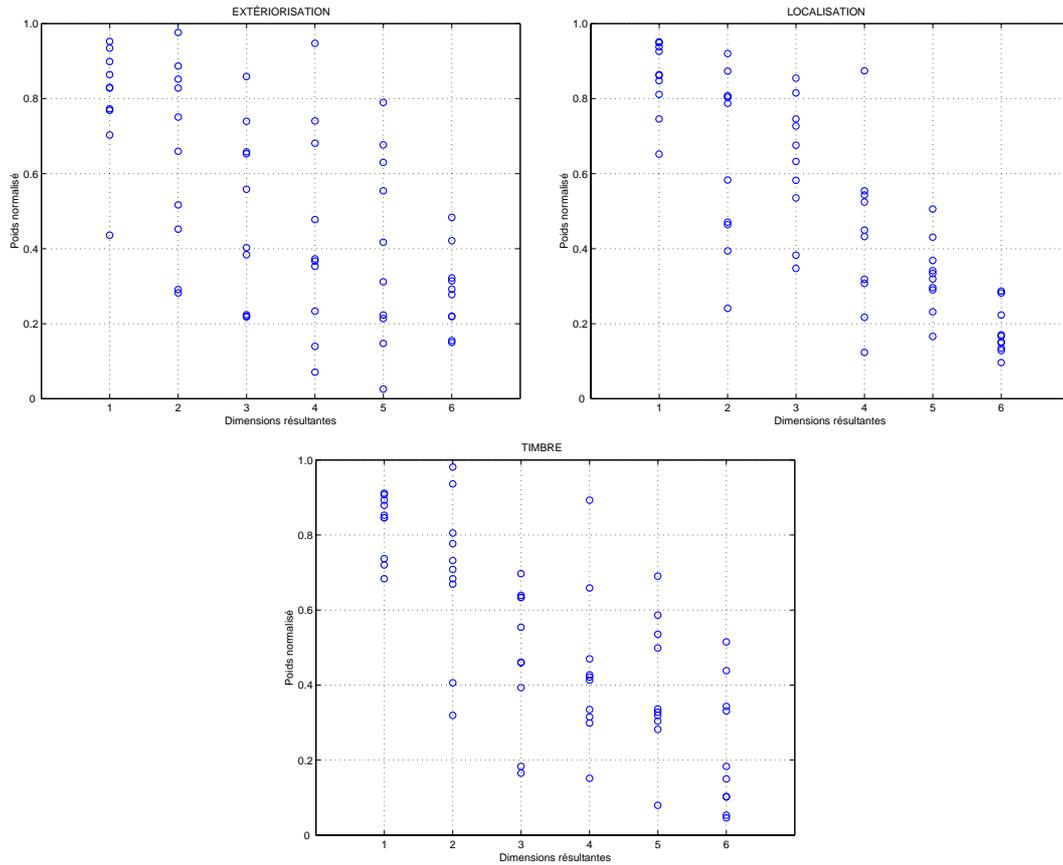


FIGURE II.33 – Poids accordé par les individus à six dimensions communes

centrée, alors

$$R_m^2 = \frac{\mathbf{y}^t [\mathbf{X}(\mathbf{X}^t \mathbf{X})^{-1} \mathbf{X}^t] \mathbf{y}}{\mathbf{y}^t \mathbf{y}}, \quad (\text{II.18})$$

avec $(\mathbf{X}^t \mathbf{X})^{-1}$ une inverse généralisée quelconque de $(\mathbf{X}^t \mathbf{X})$. Le lecteur reconnaîtra $\mathbf{X}(\mathbf{X}^t \mathbf{X})^{-1} \mathbf{X}^t$ comme étant la matrice de l'endomorphisme associé à la projection sur l'espace défini par les p vecteurs \mathbf{x}^p .

La figure II.33 illustre ces corrélations. Les poids accordés par les différents sujets aux axes principaux varient fortement, traduisant l'existence de divergences dans les stratégies de jugement de la dissemblance auditives. Cette analyse est beaucoup plus pertinente qu'un simple examen des corrélations entre les réponses des individus, où la réduction d'information est importante. Ces différences de stratégies nous suggèrent qu'effectuer la moyenne des réponses des individus avant de faire l'analyse MDSCAL revient à éliminer une grande partie de l'information. En effet, si l'on effectue la même démarche pour « l'individu moyen », on remarque aisément sur la figure II.34 qu'il a effectué un jugement simplifié puisqu'il a utilisé un nombre moins important de dimensions. Cet individu moyen représente donc l'information minimale.

En conclusion de cette analyse, nous pouvons dire que chaque individu apporte sa contribution à l'établissement d'un ordre de proximité des différents stimuli par rapport à l'écoute naturelle. Comme nous l'avons pressenti lors de l'analyse statistique classique, chaque individu a sa propre sensibilité, et possède donc des facultés d'appréciation différentes. Effectuer la moyenne des réponses de chaque individu avant de procéder à une analyse MDSCAL réduit l'information, puisque nous ne prenons pas en compte les particularités de chaque individu. Il convient de raffiner notre recherche, et d'utiliser des techniques plus complexes.

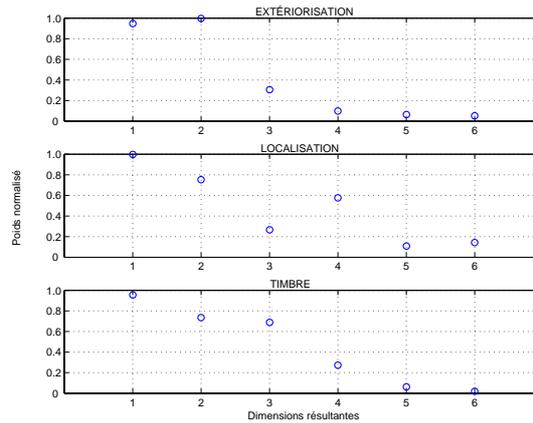


FIGURE II.34 – Poids accordés par « l'individu moyen » aux six dimensions communes à tous les individus

II.4.8.3 Analyse prenant en compte chaque sujet

Afin d'améliorer notre analyse, nous cherchons à introduire dans le modèle l'influence de chaque sujet. Nous allons donc utiliser l'analyse EXSCAL, avec un programme développé à l'IRCAM par le chercheur Suzanne Winsberg ; le principe de cette méthode est décrit en annexe B.2.4.

Comme nous effectuons l'analyse avec sept stimuli (définis dans le tableau II.2), nous limitons le nombre de dimensions recherchées à deux au plus. Les paramètres de modèles avec et sans pondération, avec et sans spécificité, ont été estimés pour chaque attribut perceptif. Les valeurs de la variance expliquée par le modèle, du logarithme de la vraisemblance, de AIC et de BIC sont données dans les tableaux II.8, II.9 et II.10, en fonction du nombre de dimensions de l'analyse. Les poids attribués par les différents sujets aux axes issus de l'analyse ne sont pas tous identiques, et confirme de manière définitive que tous les sujets n'ont pas jugé de la même manière. Dans le choix du meilleur modèle, c.-à-d. celui représentant au mieux les jugements des individus, nous devons donc privilégier les modèles prenant en compte les pondérations individuelles. De plus, nous privilégions le modèle avec la plus petite valeur de BIC, voire en plus la plus petite valeur de AIC en cas de doute. Nous choisissons alors pour :

- l'extériorisation, un modèle pondéré à deux dimensions, sans spécificité ;
- la localisation, un modèle pondéré à deux dimensions, sans spécificité ;
- le timbre, un modèle pondéré à deux dimensions, avec spécificités, car ce modèle attribue une forte spécificité au stimulus 'ID_tête', ce que tous les sujets se sont accordés à reconnaître.

Pour chaque attribut perceptif, nous avons représenté l'espace stimulus de groupe (cf. fig. II.35), et l'espace individuel (cf. fig. II.36). Les distances entre le stimulus 'Voiture' et les autres stimuli sont présentées figure II.37 ; elles ont été calculées en tenant compte des pondérations et spécificités éventuelles (cf. éq. (B.53)).

EXTÉRIORISATION

Nous constatons que le stimulus 'Voiture' se détache nettement des autres stimuli, tout comme le stimulus 'ID_tête'. Remarquons que les stimuli de référence et les corrections individuelles sont globalement situés sur un axe parallèle à l'axe 2. L'examen des poids démontre l'existence de deux groupes de sujets : un groupe est plus sensible à l'axe 1, tandis que l'autre est plus sensible à l'axe 2. Le premier groupe a donc perçu des différences non négligeables entre les stimuli sus-cités, tandis que le deuxième s'est concentré sur la scission avec le stimulus 'Voiture', les différences entre les autres stimuli étant assimilables à du « bruit ». L'examen des distances au stimulus 'Voiture' montre que le stimulus de référence de la tête se situe en troisième place, devant la correction individuelle par rapport au champ diffus.

TABLEAU II.8 – Résultats de l'analyse EXSCAL sur le test concernant l'extériorisation

ndim	log L	AIC	BIC
Sans pondération, sans spécificité			
1	-530	1074	1097
2	-510	1043	1083
3	-507	1046	1099
Avec pondération, sans spécificité			
1	-523	1096	1180
2	-483	1047	1181
3	-468	1046	1230
Sans pondération, avec spécificité			
1	-507	1042	1089
2	-506	1050	1114
3	-506	1057	1134
Avec pondération, avec spécificité			
1	-479	1039	1176
2	-471	1054	1241
3	-457	1056	1294

ndim	N° du stimulus						
	1	2	3	4	5	6	7
Spécificités sans pondération							
1	5,13	6,29	0,00	0,00	0,00	15,89	26,10
2	0,00	5,50	1,58	0,00	2,59	12,87	0,00
3	0,00	5,23	0,00	0,00	0,00	12,95	0,00
Spécificités avec pondération							
1	5,42	6,24	2,54	0,00	2,93	14,43	0,00
2	4,17	0,00	2,27	1,00	4,49	0,00	0,00
3	0,00	3,28	1,71	0,00	0,00	9,95	0,00

TABLEAU II.9 – Résultats de l'analyse EXSCAL sur le test concernant la localisation

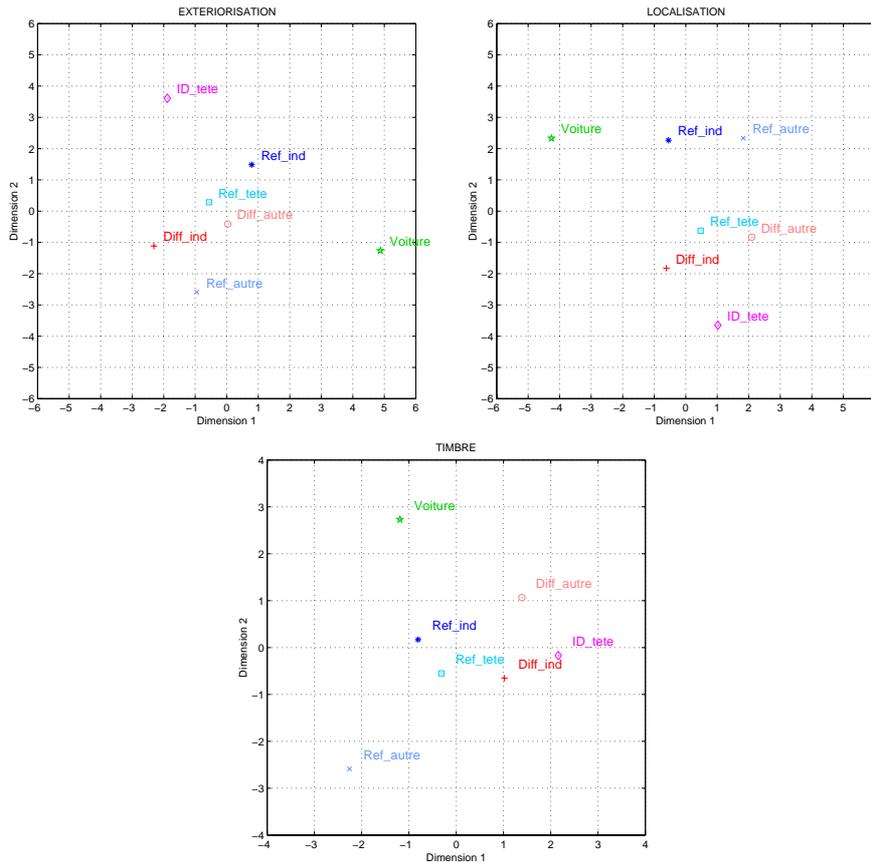
ndim	log L	AIC	BIC
Sans pondération et sans spécificité			
1	-521	1057	1080
2	-508	1040	1080
3	-506	1044	1097
Avec pondération et sans spécificité			
1	-505	1061	1144
2	-479	1038	1172
3	-462	1035	1219
Sans pondération et avec spécificité			
1	-507	1041	1088
2	-506	1050	1113
3	-506	1057	1134
Avec pondération et avec spécificité			
1	-481	1044	1181
2	-462	1036	1223
3	-443	1028	1265

ndim	N° du stimulus						
	1	2	3	4	5	6	7
Spécificités sans pondération							
1	0,00	8,82	3,47	1,80	0,00	7,83	6,07
2	0,00	7,82	0,00	0,00	0,47	8,14	0,00
3	0,00	0,00	0,00	0,00	0,00	7,75	0,00
Spécificités avec pondération							
1	0,00	7,89	6,81	1,85	0,00	9,54	0,00
2	6,72	0,00	6,85	1,13	1,40	4,98	0,00
3	5,68	0,00	0,00	0,00	0,00	8,43	0,00

TABLEAU II.10 – Résultats de l'analyse EXSCAL sur le test concernant le timbre

ndim	log L	AIC	BIC
Sans pondération et sans spécificité			
1	-532	1078	1101
2	-510	1045	1085
3	-506	1043	1097
Avec pondération et sans spécificité			
1	-522	1094	1177
2	-471	1023	1157
3	-444	998	1182
Sans pondération et avec spécificité			
1	-506	1039	1086
2	-504	1046	1110
3	-504	1053	1130
Avec pondération et avec spécificité			
1	-458	999	1136
2	-448	1009	1196
3	-421	985	1222

ndim	N° du stimulus						
	1	2	3	4	5	6	7
Spécificités sans pondération							
1	6,67	0,00	4,58	2,20	2,89	24,35	9,05
2	6,15	0,00	4,62	2,70	0,00	22,79	0,00
3	6,24	0,00	0,00	3,50	0,00	0,00	0,00
Spécificités avec pondération							
1	7,50	3,13	3,03	3,17	4,23	28,75	4,95
2	6,97	0,00	3,78	1,85	4,76	29,26	0,00
3	7,54	0,00	5,78	1,33	4,32	10,42	0,00

FIGURE II.35 – *Espaces stimuli fournis par l'analyse EXSCAL à deux dimensions*

LOCALISATION

On peut considérer, tout comme pour l'extériorisation, que le stimulus 'Voiture' se détache des autres, et que ceux-ci sont globalement situés sur un axe parallèle à l'axe 2. Par contre, les poids attribués par les sujets sont, pour la plupart, groupés autour de la première diagonale : les individus n'ont donc pas favorisé un axe par rapport à l'autre, mais ils ont attribué un facteur de zoom différent. Sans nous laisser influencer par AIC, nous aurions pu choisir un modèle à une seule dimension (les valeurs de BIC sont proches). Rapporté au stimulus 'Voiture', le stimulus de référence de la tête arrive également en troisième place, en faisant jeu égal avec la correction individuelle par rapport au champ diffus.

TIMBRE

En ce qui concerne le timbre, nous avons privilégié un modèle avec spécificités⁴ ; en effet, le stimulus n° 6 (c.-à-d. correspondant aux BIR mesurées sur le mannequin corrigées par le filtre ID proposé par HEAD ACOUSTICS) révèle une forte spécificité, ce qui corrobore l'impression auditive. La représentation de l'espace de groupe le fait apparaître comme étant assez proche des autres stimuli, bien qu'il ne le soit pas. Nous devons interpréter cette représentation comme une projection dans un plan commun à tous les stimuli ; les stimuli ayant une forte spécificité s'en écartent fortement. Le stimulus de référence de la tête arrive ici en deuxième place, largement devant la correction individuelle par rapport au champ diffus.

4. on rappelle qu'une spécificité peut être vue comme un axe expliqué par un stimulus particulier

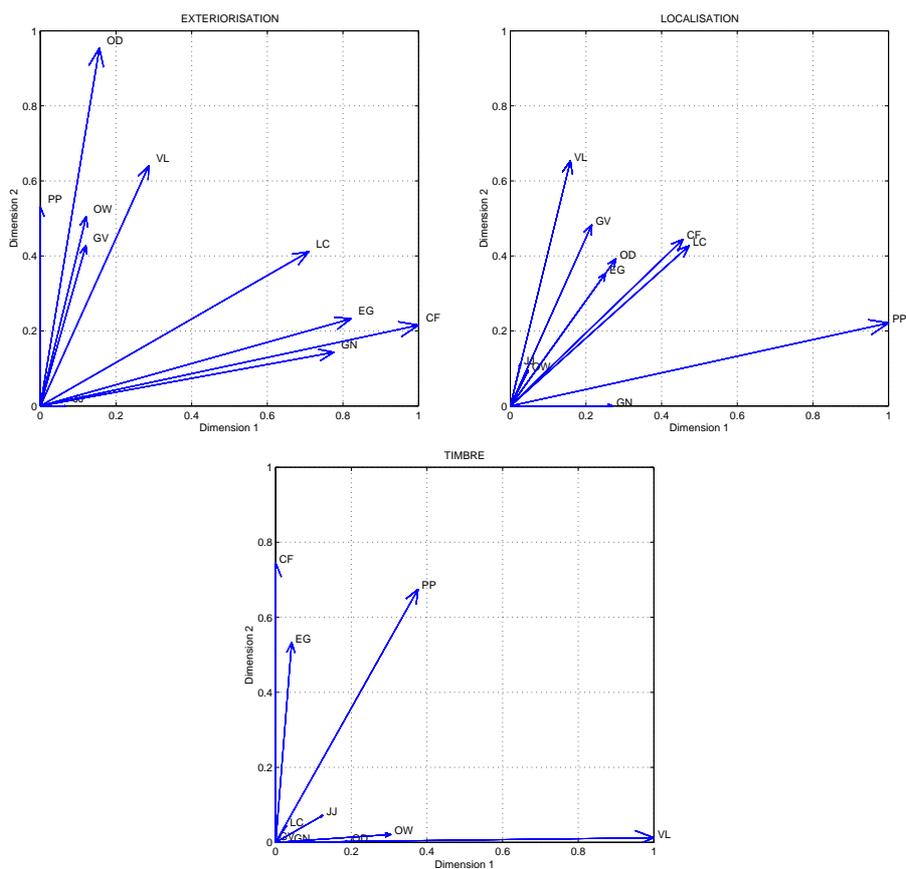


FIGURE II.36 – *Espaces sujets fournis par l'analyse EXSCAL à deux dimensions*

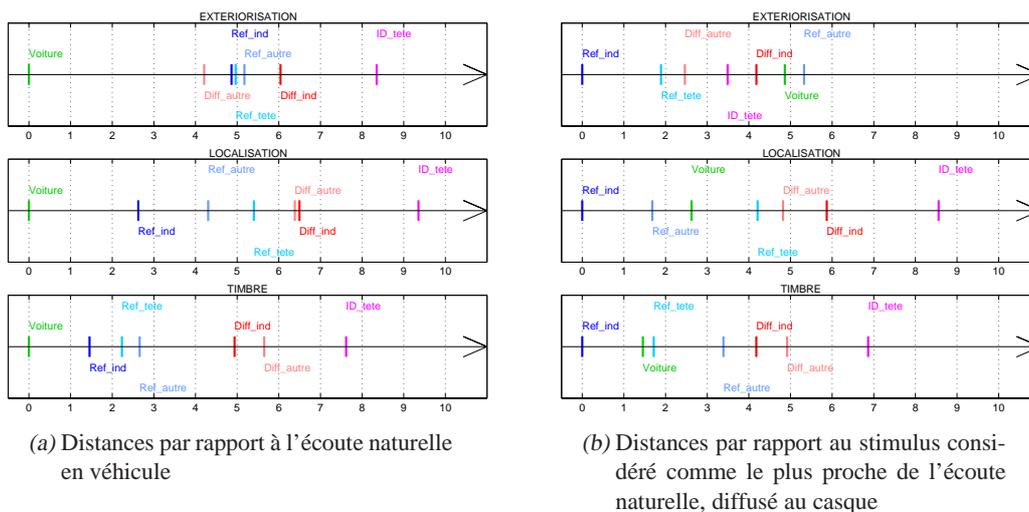


FIGURE II.37 – *Distances calculées à partir des points de la configuration finale fournie par EXSCAL à deux dimensions, en tenant compte des pondérations (et des spécificités pour le timbre)*

II.4.8.4 Conclusion

L'analyse multidimensionnelle des résultats de notre test d'écoute, conduit sous forme de comparaison par paires, nous permet de conclure que le calcul d'une correction individuelle par rapport au champ diffus ne s'impose pas. Toutes les analyses effectuées confirment que le stimulus de référence du mannequin est au moins aussi proche de l'écoute naturelle des haut-parleurs de la voiture. De plus, le calcul de cette correction comporte plusieurs étapes délicates et coûteuses. Enfin, à la différence d'une application de transmission de BIR susceptibles d'être écoutées avec n'importe quel casque, notre étude nous permet de conserver toujours le même casque, et celui-ci possède de bonnes propriétés de couplage avec les oreilles. Nous décidons de rejeter, pour notre application, le calcul de la correction individuelle par rapport au champ diffus.

Une simple mesure des BIR du mannequin, suivie d'une inversion numérique du casque d'écoute (modèle K₁₀₀₀ de marque AKG) toujours mesuré sur le mannequin, permet de rendre compte de manière optimale de la qualité acoustique régnant dans l'habitacle du véhicule.

II.5 Conclusion

Ce second chapitre avait pour but de rechercher une méthode universelle et optimale de restitution sur écouteurs du champ sonore régnant à l'intérieur d'un habitacle de véhicule automobile.

Nous avons d'abord rappelé les principes de la simulation binaurale, et nous avons remarqué que les conditions nécessaires à l'exactitude de sa restitution sont difficilement compatibles avec les exigences de la conception audio des véhicules. Il est en effet impératif de faire comparer différentes installations audio à un très grand nombre de sujets, sans que ceux-ci soient obligés de se porter physiquement dans les différents habitacles ; les tests d'écoute ont lieu en laboratoire, l'écoute se faisant au casque. Nous avons donc décidé de procéder à des enregistrements binauraux à l'aide d'un mannequin acoustique.

Malheureusement, les caractéristiques acoustiques d'un mannequin sont assez éloignées de celles des humains. C'est pourquoi, nous avons élaboré une méthode de correction individuelle de ces enregistrements, en essayant de nous rapprocher d'enregistrements effectués sur nos propres oreilles, à l'aide des HRTF en champ diffus du mannequin et des individus.

Cependant, il apparaît que l'inversion du casque d'écoute mesuré sur le mannequin, préalable à la restitution d'enregistrements effectués sur le mannequin, permet d'enlever certaines caractéristiques propres à celui-ci, surtout lorsque le casque choisi est le K₁₀₀₀ utilisé oreillettes ouvertes. En effet, cette fonction de transfert est très proche de celles calculées à l'aide d'une correction individuelle des enregistrements du mannequin par rapport au champ diffus (suivie d'une inversion du même casque mesuré sur les individus).

Pour lever toute ambiguïté, nous avons réalisé un test d'écoute subjectif. Il avait pour but de déterminer le meilleur traitement à apporter aux mesures de réponses impulsionnelles binaurales effectuées avec notre mannequin acoustique. Il permettait de comparer différents stimuli diffusés au casque avec l'écoute naturelle, c.-à-d. par les haut-parleurs du véhicule.

Confirmant les impressions des sujets, notre analyse statistique a démontré que le stimulus le plus proche de la situation naturelle est sans conteste l'écoute de nos propres BIR, corrigées par le casque d'écoute. Par contre, il est difficile de mettre en évidence une réelle différence entre les BIR du mannequin corrigées par le casque d'écoute mesuré sur le mannequin, et les mêmes BIR du mannequin corrigées individuellement par rapport au champ diffus, et suivies d'une inversion du même casque mesuré sur les individus. Nous confirmons les résultats objectifs.

Comme le principe de la correction individuelle par rapport au champ diffus nécessite, pour chaque personne, le calcul du rapport entre la HRTF en champ diffus de l'individu réalisant l'écoute et la HRTF en champ diffus du mannequin, que cette quantité représente un surcoût important de calcul et nécessite des ajustements délicats, nous décidons de l'abandonner.

La restitution binaurale de la qualité acoustique régnant à l'intérieur d'un véhicule automobile sera donc réalisée à l'aide de mesures de BIR effectuées sur notre mannequin acoustique, suivies d'une inversion numérique spécifique de notre casque d'écoute K_{1000} utilisé oreillettes ouvertes.

CHAPITRE III

Amélioration de l'écoute embarquée

III.1 Introduction

Les travaux de caractérisation objective de la restitution audio en habitacle permettent de dégager des pistes concernant les traitements à mettre en place pour améliorer l'écoute. Le premier traitement envisagé est une correction de la réponse en fréquence de l'habitacle. Cette correction est rendue nécessaire par la présence de résonances assez marquées, causant de profondes altérations de la réponse en fréquence, et perçues subjectivement comme un effet « tonneau ». Ces résonances sont dues en particulier au petit volume de l'habitacle ; elles sont très importantes en basses fréquences, c.-à-d. en dessous de la fréquence de Schroeder qui est de l'ordre de plusieurs centaines de Hertz, zone dans laquelle le recouvrement modal est faible.

Mais une correction fréquentielle n'est pas suffisante ; la présence de surfaces réfléchissantes telles que les vitres fait émerger de fortes réflexions, conduisant à détériorer encore un peu plus l'image sonore. Outre une correction fréquentielle, un second traitement envisagé est une correction temporelle qui permet, d'une part, de modifier la direction *apparente* de provenance des sources sonores, et d'autre part, de substituer à l'effet d'habitacle une réverbération plus conséquente.

Dans un mode très simplifié, il est possible d'aligner les temps d'arrivée et les niveaux énergétique des ondes sonores provenant des différents haut-parleurs. Mais par le biais d'une inversion à phase totale des réponses impulsionnelles de l'habitacle, il est possible de les rendre transparentes. Cela permet de recentrer l'image sonore (en créant des haut-parleurs virtuels) et d'annuler l'effet des premières réflexions.

D'autre part, si on désire substituer à l'effet d'habitacle une autre réverbération, il faut d'abord mettre en œuvre une procédure de compensation ; celle-ci *peut* faire appel à la déconvolution, mais ce n'est pas la seule méthode (cf. la procédure de compensation de contexte du Spatialisateur® [140, 238]). Puis il faut faire appel à une méthode de synthèse en temps réel du nouvel effet de salle. Remarquons que si la réverbération désirée est conséquente, il est plus facile de travailler directement par masquage.

Enfin, bien que les grandes salles soient capables de générer une sensation d'espace sonore, il n'en est pas de même pour les petites, spécialement les habitacles de voiture. Nous avons donc élaboré un troisième traitement électronique des signaux pour augmenter la sensation d'espace perçue par les occupants d'un véhicule. Il convient cependant de distinguer plusieurs composantes de la sensation d'espace, chacune nécessitant un traitement particulier pour l'augmenter.

Nous n'avons pas pour ambition de corriger les éventuelles non-linéarités du système audio existant [74] ; ces distorsions peuvent apparaître au niveau de l'amplificateur ou des haut-parleurs eux-mêmes,

principalement si on tente de dépasser la limite de puissance imposée. Cependant, on peut rendre au système un comportement grossièrement linéaire, en filtrant de manière particulière le signal avant sa diffusion par les haut-parleurs [83, 84].

III.2 Correction du timbre

III.2.1 Cadre de l'étude

Nous désirons réaliser une correction de la réponse en fréquence d'un habitacle de véhicule automobile. Compte tenu des observations réalisées pendant l'étape de caractérisation, la question se pose de la stratégie à adopter pour mener l'égalisation. En effet, l'ordre de grandeur du support temporel d'une réponse impulsionnelle d'habitacle de véhicule automobile est de 50 ms environ. Et l'analyse temps-fréquence de ces réponses impulsionnelles montre que le temps de réverbération décroît en basses fréquences pour atteindre une valeur constante de 50 ms à partir de 800 Hz (cf. § I.5.2). Par abus de langage, on parlera de réverbération, malgré la durée très courte du phénomène, bien que l'on doive plutôt parler d'« effet d'habitacle ».

Pour les salles de taille classique, on a l'habitude de considérer que les contributions du son direct et de la réverbération jouent des rôles indépendants [181]. Cette perception peut être liée à la distribution temporelle des réflexions (précoces/tardives), et à leur distribution spatiale (localisées/diffuses). Cependant, dans le cas d'un habitacle de véhicule automobile, les phénomènes de masquage et d'intégration temporels justifient que l'on s'interroge sur la subsistance de cette distinction dans un cas où le temps est très ramassé.

Il peut donc paraître *a priori* inutile de distinguer la réponse spectrale liée au son direct de celle liée à l'habitacle. Nous devons chercher à répondre à cette question, afin de savoir si il est indispensable de dissocier les corrections fréquentielles, c.-à-d. effectuer une égalisation différente pour chacune des deux contributions spectrales. Pour s'en assurer, un test psychoacoustique a été réalisé à partir de réponses impulsionnelles synthétiques permettant de contrôler de manière fine les contenus spectraux du son direct et de l'effet d'habitacle. Bien que non totalement réalistes, des conditions de laboratoire permettent néanmoins de calibrer beaucoup plus finement une expérience. Nous avons donc utilisé des réponses impulsionnelles de synthèse issues du Spatialisateur[®], configuré pour traduire les principales caractéristiques de la restitution du son dans un habitacle de véhicule automobile (cf. § III.2.3); on se place alors dans le cadre d'une qualité acoustique de type voiture. Et nous avons fait varier la distribution temps-fréquence-énergie, c.-à-d. les contenus fréquentiels relatifs du champ direct et du champ réverbéré.

L'enjeu est de savoir si, dans le contexte de réponses courtes, les sujets distinguent, à spectre total constant, des variations de répartition temps-fréquence s'accompagnant d'une distribution spatiale.

Chaque réponse impulsionnelle ainsi créée représente un stimulus; nous décidons de comparer entre eux six stimuli, soit quinze paires. Ce test est reconduit dans trois conditions, chacune se différenciant par les variations fréquentielles du temps de réverbération, et par la diffusion spatiale de la réverbération.

III.2.2 Configurations envisagées

Les différentes configurations de notre test d'écoute sont rassemblées dans le tableau III.1; il s'agit des niveaux énergétiques relatifs du contenu fréquentiel du son direct et de la réverbération, dans trois bandes de fréquences: basse, moyenne et haute. On remarque que le rapport énergétique entre l'onde directe et le champ réverbéré est constant pour chaque catégorie A ou B:

- $\{-6; 0; +6\}$ dB pour la configuration A, dans les trois bandes de fréquences $\{BF; MF; HF\}$;
- $\{+6; 0; -6\}$ dB pour la configuration B.

TABLEAU III.1 – Réglages énergétiques pour le test d'écoute sur l'influence perceptive du contenu fréquentiel relatif du son direct et de la réverbération.

	A		B		
	Son direct	Réverbération	Son direct	Réverbération	
1			1A	1B	$\begin{pmatrix} 1A & 1B & 2A & 2B & 3A & 3B \\ 1A & & & & & \\ 1B & 1 & & & & \\ 2A & 2 & 3 & & & \\ 2B & 4 & 5 & 6 & & \\ 3A & 7 & 8 & 9 & 10 & \\ 3B & 11 & 12 & 13 & 14 & 15 \end{pmatrix}$
2			2A	2B	
3			3A	3B	

(a) Valeurs énergétiques fixées dans trois bandes de fréquences : basses, moyennes et hautes

(b) Numéro des paires de stimuli

D'autre part, l'énergie totale, somme de l'énergie du son direct et de la réverbération, est constante pour chaque configuration 1, 2 ou 3 :

- $\{+1 ; +3 ; +7\}$ dB environ pour la configuration 1, ce qui correspond à un réglage nettement passe-haut ;
- $\{+4 ; +3 ; +4\}$ dB environ pour la configuration 2 ;
- $\{+7 ; +3 ; +1\}$ dB environ pour la configuration 3, ce qui correspond à un réglage nettement passe-bas.

Les réglages énergétiques des stimuli oscillent donc autour d'un rapport énergétique entre l'onde directe et le champ réverbéré qui vaut toujours 0 dB dans la bande MF.

En d'autres termes, on peut considérer trois familles de modifications :

1. Modification globale du spectre de l'énergie : c'est le cas des paires 2 $\{1A;2A\}$, 5 $\{1B;2B\}$, 9 $\{2A;3A\}$, 14 $\{2B;3B\}$, 7 $\{1A;3A\}$ et 12 $\{1B;3B\}$.
2. Energie totale identique, avec échange entre le son direct et la réverbération : c'est le cas des paires 1 $\{1A;1B\}$, 6 $\{2A;2B\}$ et 15 $\{3A;3B\}$.
3. Affectation particulière de l'énergie sur le son direct ou la réverbération :
 - modifications de la densité d'énergie du son direct, avec compensation par une moindre modification inverse sur la réverbération : c'est le cas des paires 4 $\{1A;2B\}$ et 13 $\{2A;3B\}$; ou inversement : c'est le cas des paires 3 $\{2A;1B\}$ et 10 $\{3A;2B\}$;
 - modifications uniquement de la densité d'énergie du son direct ou de la réverbération : c'est le cas des paires 8 $\{1B;3A\}$ et 11 $\{1A;3B\}$.

Trois ensembles de configurations, correspondant à trois conditions différentes, seront construits. Dans chaque ensemble, le réglage énergétique sera identique à ce qui est décrit précédemment, la variabilité étant introduite par les variations fréquentielles du temps de réverbération, et par la diffusion spatiale de la réverbération (cf. tab. III.2).

Dans les deux premières conditions, nous prenons des valeurs du temps de réverbération proches de celles rencontrées lors de l'analyse temps-fréquence de véritables réponses impulsionnelles d'habitacles, à savoir $\{0,3 ; 0,05 ; 0,05\}$ sec. dans les trois bandes de fréquences ; dans la condition *Référence*, nous prenons des valeurs identiques, à savoir $\{0,1 ; 0,1 ; 0,1\}$ sec. Ces deux ensembles de valeurs permettront de juger de l'influence d'une répartition non homogène du temps de réverbération dans les trois bandes de fréquences.

TABLEAU III.2 – Variations du temps de réverbération (en sec.) et de la diffusion spatiale de la réverbération au cours des trois conditions d'écoute

	Tr BF	Tr MF	Tr HF	Réverbération
Condition Habitable	0,30	0,05	0,05	diffuse
Condition Frontale	0,30	0,05	0,05	localisée frontale
Condition Référence	0,10	0,10	0,10	diffuse

En ce qui concerne la diffusion spatiale de la réverbération, nous choisissons, d'une part, une réverbération diffuse pour les conditions *Référence* et *Habitable*, et d'autre part, une réverbération localisée frontalement pour le second. Ces deux réglages permettront de juger de l'influence de l'aspect diffus de l'effet d'habitable.

Le but du test est d'étudier l'influence de ces différents réglages sur notre perception, dans le cas d'une écoute en habitacle automobile. Afin de s'approcher de cette situation, nous devons maintenant procéder au réglage du Spatialisateur®.

III.2.3 Réglage du Spatialisateur®

Une description détaillée du processeur de réverbération est donnée en annexe E.1. On rappelle que l'effet de salle généré repose sur une modélisation temps-fréquence de la réponse impulsionnelle. Celle-ci est composée de quatre zones temporelles OD, R1, R2 et R3, correspondent à l'onde directe, les premières réflexions, le début de la réverbération tardive, et la réverbération tardive par elle-même. L'énergie de chaque section temporelle est réglable en trois bandes de fréquences, grâce à des filtres « double shelving » implantés sous forme d'égaliseurs paramétriques.

Dans notre cas, les limites des zones temporelles sont modifiées pour approcher au mieux la description temporelle d'une réponse impulsionnelle d'habitable de véhicule, pour laquelle on considère que la partie diffuse débute aux alentours de 5 ms après le son direct. Les limites sont donc fixées comme suit :

$$\begin{aligned} R1 &= [1 - 6] \text{ ms} \\ R2 &= [4 - 10] \text{ ms} \\ R3 &= [5 - \infty] \text{ ms.} \end{aligned}$$

Les valeurs des fréquences de transition des filtres « double shelving » du Spatialisateur® (cf. § E.1.3) sont laissées aux valeurs par défaut, c.-à-d. $f_1=250$ Hz et $f_2=8000$ Hz. Les gains imposés aux filtres sont doublés en basses et hautes fréquences, puisque la valeur du gain à la fréquence de transition est égale à la moitié de l'excursion totale imposée. Par exemple, pour la configuration 1A, on désire une excursion de ± 6 dB entre les bandes BF et HF, et pour cela les gains imposés aux filtres doivent être de ± 12 dB.

D'autre part, la densité modale, c.-à-d. le nombre de modes par Herz, doit avoir une valeur minimale égale au quart du temps de réverbération ; nous la fixons donc à 0,05.

Afin de commuter rapidement entre les différentes configurations, nous utilisons le système de messages de Max®, envoyés à quatre filtres du Spatialisateur® :

- les filtres *face* et *omni* appartenant au module *Source* traitant des effets que l'on veut donner à la source sonore ;
- les filtres *direct* et *room* appartenant au module *Room* traitant des effets que l'on veut donner à l'effet de salle.

Les filtres *face* et *direct* ont une influence sur le son direct, tandis que les filtres *omni* et *room* alimentent la réverbération. La mise en série de ces filtres permet d'atteindre les configurations énergétiques visées pour notre expérience, comme expliqué figure III.1. Par exemple, en imposant des gains de ± 12 dB,

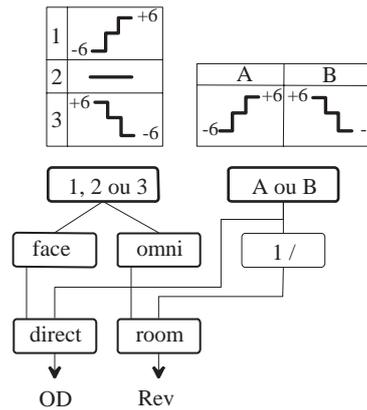


FIGURE III.1 – Mise en cascade des filtres du Spatialisateur®

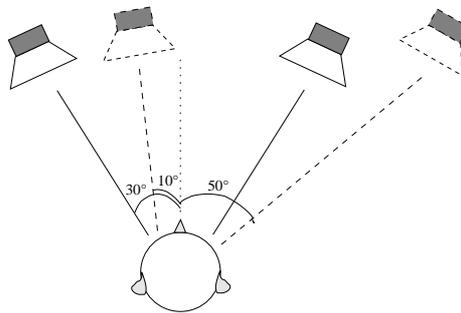


FIGURE III.2 – Dispositif de restitution utilisé pour le test d'écoute en laboratoire

on obtient une excursion de ± 6 dB entre les deux fréquences de transition sur lesquelles sont centrées les bandes BF et HF. Le niveau de la bande MF reste toujours égal à 0 dB.

III.2.4 Protocole expérimental

Le système de diffusion choisi est transaural, c.-à-d. sur deux haut-parleurs placés à $\pm 30^\circ$ par rapport au plan médian de la tête de l'auditeur, avec un dispositif d'annulation de la diaphonie gauche-droite (plus connu sous le nom de *crosstalk cancellation*). Afin de se rapprocher des conditions d'écoute des haut-parleurs avant dans un véhicule, nous avons simulé deux haut-parleurs virtuels avec des directions de provenance réglées à 10° à gauche et 50° à droite (cf. fig. III.2). Pour les conditions *Habitacle* et *Référence*, les canaux de sortie du filtre réverbérant sont conservés tels quels, c.-à-d. non corrélés ; ils génèrent une réverbération diffuse. Pour la condition *Frontale*, les canaux de sortie du module Room sont sommés avant de rentrer dans le module Pan~, de façon à ce que la réverbération soit localisée dans la direction de provenance des sources virtuelles. Le dispositif est situé dans un studio d'écoute isolé des sources sonores environnantes.

Dix-sept sujets ont écouté les quarante-cinq paires (quinze paires par condition), sur un extrait d'une symphonie de Haydn dont le contenu fréquentiel et dynamique a été jugé suffisamment large. Il était demandé aux sujets de juger de la différence globale entre les deux stimuli d'une même paire, et de rapporter cette différence sur une échelle continue, mais non quantifiée visuellement, allant de *identique* à *différent* (cf. fig. III.3) ; intrinsèquement, l'échelle va de 0 à 100. Le test s'est déroulé en double aveugle. Les sujets pouvaient commuter à tout instant entre les deux configurations à tester. La durée du test a été très variable entre les différents individus : d'une demi-heure à presque deux heures ! Remarquons que le jugement des sujets fournit une information de *dissemblance* et non de *distance*, puisque l'in-

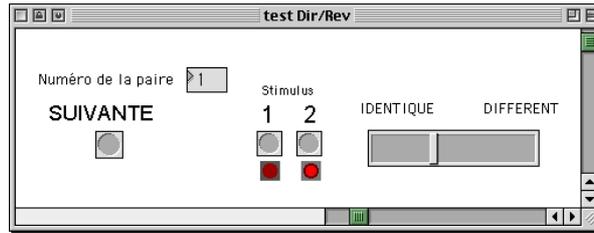


FIGURE III.3 – Interface du test d'écoute réalisé pour étudier l'influence perceptive du contenu fréquentiel relatif du son direct et du champ réverbéré

égalité triangulaire ne sera pas forcément respectée ; nous sommes, pour l'instant, dans un espace non euclidien. Pour chaque sujet, le programme informatique de test fournit trois matrices de dissemblance, une par condition. On suppose qu'il n'y a pas de différence entre le stimulus A et lui-même, et que la dissemblance entre les stimuli A et B est identique à celle entre les stimuli B et A. Chaque matrice de dissemblance possède donc des 0 sur la diagonale, et sa forme est triangulaire (l'autre moitié étant reconstruite si besoin).

III.2.5 Analyse objective

L'analyse objective des différents stimuli de notre test d'écoute permet de vérifier que ceux-ci sont conformes aux spécifications que nous nous sommes fixées.

Les réponses impulsionnelles générées par le Spatialisateur® ont donc été examinées avec la procédure d'analyse temps-fréquence développée au chapitre précédent. Les paramètres d'analyse sont les suivants :

- fenêtre de Blackmann-Harris d'ordre quatre de 512 points, soit 11,6 ms à 44,1 kHz, avec un pas d'avancement de 128 points, soit 2,9 ms ;
- TFD sur 512 points, soit une résolution fréquentielle de 86 Hz ;
- début de la réverbération choisi à 5 ms du son direct.

À partir du relief de décroissance, nous calculons, outre la densité fréquentielle du temps de réverbération, la densité fréquentielle de l'énergie totale, ainsi que celle de l'énergie réverbérée (notée Rev), c.-à-d. 5 ms après le son direct ; nous en déduisons, par soustraction, la densité fréquentielle de l'énergie du son direct équivalent (notée OD).

Conformément à la procédure décrite au chapitre précédent, la résolution fréquentielle est d'abord réduite à sept bandes d'octaves de 177 à 11314 Hz, avant d'être ramenée à seulement trois valeurs correspondant aux moyennes des niveaux dans les bandes d'octaves {1 ; 2}, {3 ; 4 ; 5} et {6 ; 7}. Les bandes de fréquences sur lesquelles se repose notre test d'écoute se composent comme donc suit :

- BF regroupe les octaves centrées sur 177 Hz et 353 Hz ;
- MF regroupe les octaves centrées sur 707 Hz, 1414 Hz et 2828 Hz ;
- HF regroupe les octaves centrées sur 5657 Hz et 11314 Hz.

III.2.5.1 Configurations idéales

Les valeurs idéales que nous voulions nous imposer, à savoir ± 3 dB ou ± 6 dB dans deux bandes de fréquences basses et hautes, ne seront pas exactement atteintes. En effet, le procédé de réglage des filtres du Spatialisateur® consiste à spécifier les cinq paramètres de réglage d'un filtre de type « double shelving » (cf. § E.1.3) ; notamment, les valeurs des gains en basses et hautes fréquences correspondent aux valeurs à $f = 0$ et $f = F_{ech}/2$. Et lorsque nous effectuons la réduction de la résolution fréquentielle exposée précédemment (cf. fig. III.4), nous obtenons trois valeurs légèrement différentes, regroupées dans le tableau III.3. De ce fait, nous ne pouvons pas atteindre les configurations idéales exposées dans

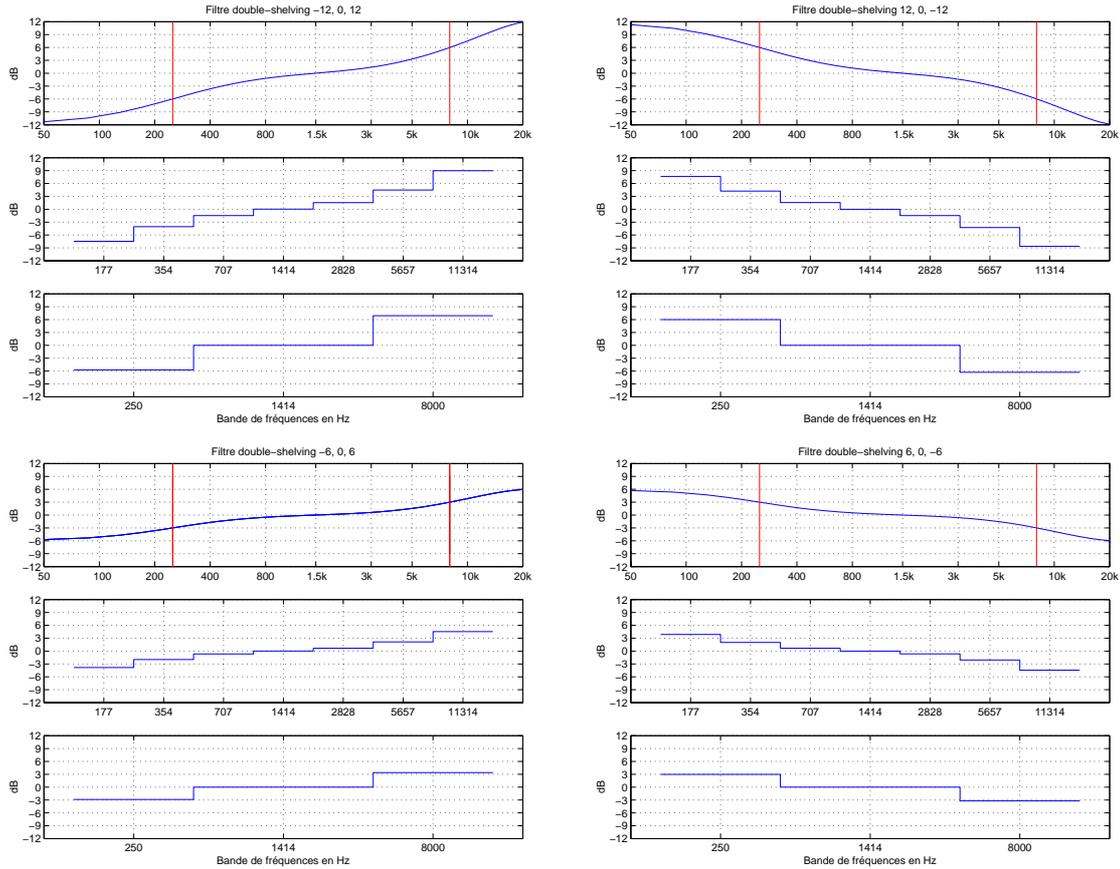


FIGURE III.4 – Fonctions de transfert pleine échelle des filtres « double shelving », puis réduction en sept bandes d'octaves, et moyenne en trois valeurs finales

le tableau III.1, mais des valeurs approchées à moins de 1 dB, ce qui ne nuira en rien à la pertinence de notre test ; ces valeurs sont rassemblées dans le tableau III.4.

III.2.5.2 Configurations réelles

Nous calculons d'abord le temps de réverbération pour chacun des six stimuli de la condition *Habitacle* (cf. fig. III.5(a)), et nous constatons qu'il est en accord avec les spécifications que nous nous étions fixées au § III.2.2.

En ce qui concerne les niveaux énergétiques du son direct équivalent et de la réverbération (cf. fig. III.5(c)), il faut tenir compte du réglage imposé au temps réverbération ; en effet, la valeur en basses fréquences est beaucoup plus élevée que celles en moyennes et hautes fréquences (d'un facteur 6, soit 7,8 dB), et ces niveaux influent directement sur celui de la section diffuse du Spatialisateur[®], à savoir la section R3. En particulier, en l'absence d'influence de tout autre filtre, les écarts de gain entre les trois sections fréquentielles du temps de réverbération doivent se retrouver sur la section R3. Cependant, dans notre étude, comme les sections R1, R2 et R3 du Spatialisateur[®] se recouvrent partiellement, ce n'est pas exactement le cas ; la différence de gain est plus faible (cf. fig. III.5(c)). De plus, l'énergie totale n'est pas totalement identique entre les configurations A et B (cf. fig. III.5(e)). Dans la suite de l'analyse nous vérifierons que ces différences n'entraînent pas de modifications perceptibles, notamment par comparaison avec la condition *Référence*.

Puis nous effectuons le même genre d'analyse pour les six stimuli de la condition *Référence*. Les résultats sont rassemblés sur les figures III.5(b), III.5(d), et dans le tableau III.6. Les résultats sont par-

TABLEAU III.3 – Paramètres de réglage des filtres « double shelving », et valeurs finales issues de la réduction de la résolution fréquentielle en trois bandes de fréquences BF, MF et HF ; les valeurs sont exprimées en dB

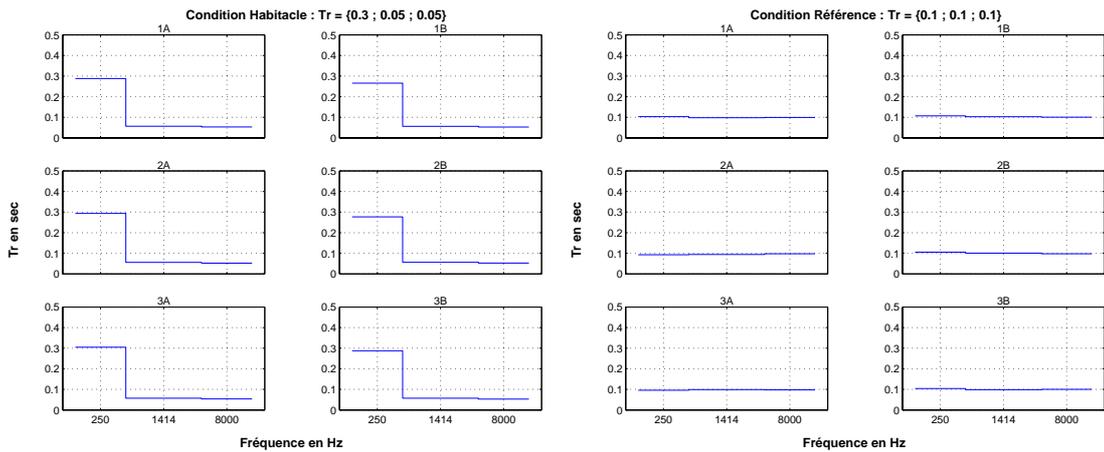
Paramètres de réglage des filtres « double shelving »	-12,0 0,0 +12,0	+12,0 0,0 -12,0	-6,0 0,0 +6,0	+6,0 0,0 -6,0
Réduction du gain en trois valeurs	-5,7 0,0 +6,9	+6,0 0,0 -6,3	-2,8 0,0 +3,4	+3,0 0,0 -3,2
Configurations idéales	-6,0 0,0 +6,0	+6,0 0,0 -6,0	-3,0 0,0 +3,0	+3,0 0,0 -3,0

TABLEAU III.4 – Niveaux énergétiques BF, MF et HF envisageables à partir d'une spécification énergétique sous forme de filtres « double shelving », exprimés en dB. L'énergie totale est la somme de l'énergie du son direct et de celle de la réverbération tardive

Stimulus	1A			1B		
Energie totale	+1,0	+3,0	+7,7	+1,0	+3,0	+7,7
Energie directe	-5,7	0,0	+6,9	0,0	0,0	0,0
Energie réverbérée	0,0	0,0	0,0	-5,7	0,0	+6,9
Stimulus	2A			2B		
Energie totale	+4,0	+3,0	+4,3	+4,0	+3,0	+4,3
Energie directe	-2,8	0,0	+3,4	+3,0	0,0	-3,2
Energie réverbérée	+3,0	0,0	-3,2	-2,8	0,0	+3,4
Stimulus	3A			3B		
Energie totale	+7,0	+3,0	+0,9	+7,0	+3,0	+0,9
Energie directe	0,0	0,0	0,0	+6,0	0,0	-6,3
Energie réverbérée	+6,0	0,0	-6,3	0,0	0,0	0,0

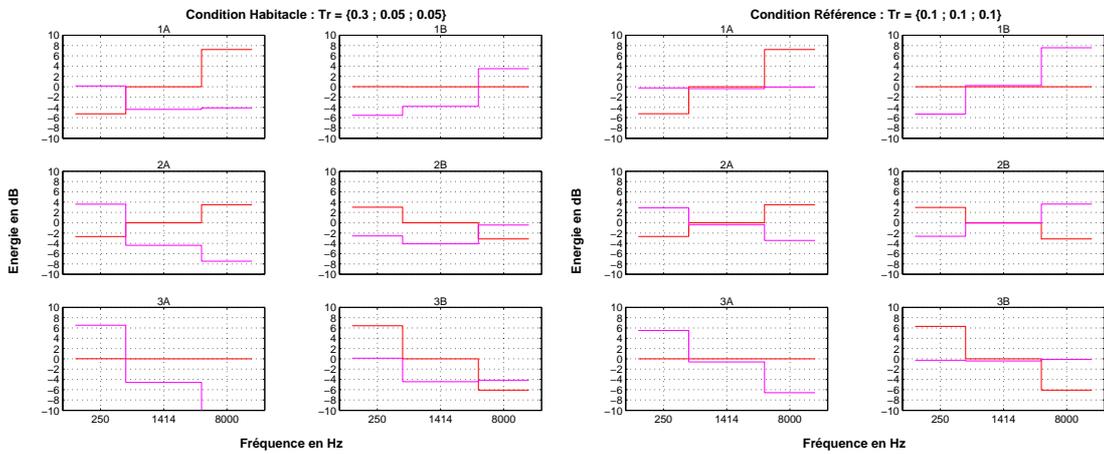
TABLEAU III.5 – Niveaux énergétiques de la condition Habitable, issus d'une analyse temps-fréquence et réduits à trois valeurs, exprimées en dB

Stimulus	1A			1B		
Energie totale	+1,2	+1,4	+7,5	+1,1	+1,5	+5,1
Energie directe	-5,3	0,0	+7,2	0,0	0,0	0,0
Energie réverbérée	+0,1	-4,4	-4,1	-5,6	-3,80	3,5
Stimulus	2A			2B		
Energie totale	+4,5	+1,4	+3,8	+4,1	+1,4	+1,4
Energie directe	-2,7	0,0	+3,5	+3,0	0,0	-3,1
Energie réverbérée	+3,6	-4,4	-7,5	-2,6	-4,1	-0,4
Stimulus	3A			3B		
Energie totale	+7,4	+1,3	+0,4	+7,3	+1,3	-2,0
Energie directe	0,0	0,0	0,0	+6,4	0,0	-6,1
Energie réverbérée	+6,5	-4,6	-10,6	+0,1	-4,4	-4,1



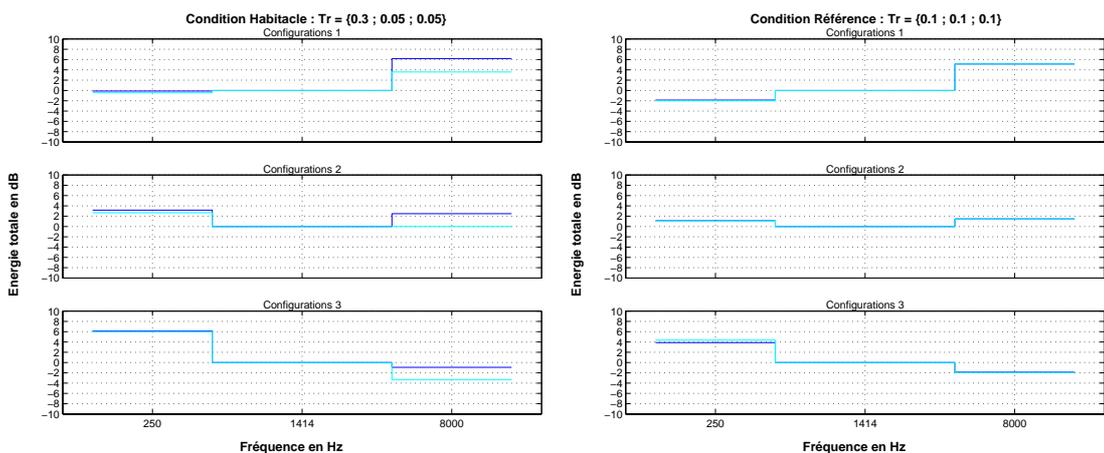
(a) Temps de réverbération pour la condition *Habitacle*

(b) Temps de réverbération pour la condition *Référé*



(c) Énergie du son direct équivalent (rouge) et de la réverbération (magenta), la condition *Habitacle*

(d) Énergie du son direct équivalent (rouge) et de la réverbération (magenta), pour la condition *Référé*



(e) Énergie totale des configurations A (bleu) et B (cyan), pour la condition *Habitacle*

(f) Énergie totale des configurations A (bleu) et B (cyan), pour la condition *Référé*

FIGURE III.5 – Temps de réverbération et niveaux énergétiques des stimuli des conditions *Habitacle* et *Référé*, réduits à trois valeurs après avoir été spécifiés sous forme de filtre « double shelving »

TABLEAU III.6 – Niveaux énergétiques de la condition Référence, issus d'une analyse temps-fréquence et réduits à trois valeurs, exprimées en dB

Stimulus	1A			1B		
Energie totale	+1,0	+2,8	+8,0	+1,1	+3,1	+8,3
Energie directe	-5,2	0,0	+7,2	0,0	0,0	0,0
Energie réverbérée	-0,2	-0,4	0,0	-5,3	+0,2	+7,6
Stimulus	2A			2B		
Energie totale	+4,0	+2,8	+4,3	+4,0	+3,0	+4,5
Energie directe	-2,7	0,0	+3,5	+3,0	0,0	-3,1
Energie réverbérée	+2,9	-0,4	-3,4	-2,6	0,0	+3,6
Stimulus	3A			3B		
Energie totale	+6,6	+2,7	+0,9	+7,2	+2,8	+0,8
Energie directe	0,0	0,0	0,0	+6,3	0,0	-6,1
Energie réverbérée	+5,5	-0,6	-6,6	-0,3	-0,4	-0,1

TABLEAU III.7 – Analyse de variance pour les trois conditions d'écoute

Condition Habitacle					
	Variance	ddl	Variance / ddl	f	Pr
Stimuli	1,308×10 ⁵	14	9343,0	20,76	100,0
Erreur	1,080×10 ⁵	240	450,1		
Total	2,388×10 ⁵	254			
Condition Frontale					
	Variance	ddl	Variance / ddl	f	Pr
Stimuli	1,158×10 ⁵	14	8268,0	19,18	100,0
Erreur	1,034×10 ⁵	240	431,1		
Total	2,192×10 ⁵	254			
Condition Référence					
	Variance	ddl	Variance / ddl	f	Pr
Stimuli	0,993×10 ⁵	14	7095,0	14,37	100,0
Erreur	1,185×10 ⁵	240	493,6		
Total	2,178×10 ⁵	254			

faitement en accord avec les valeurs du tableau III.4, et l'énergie totale est parfaitement identique entre les configurations A et B (cf. fig. III.5(f)).

Nous ne procédons pas à l'analyse objective des stimuli de la condition *Frontale*, puisque les configurations énergétiques et le temps de réverbération sont en tout point identiques à ceux de la condition *Habitacle* ; seule change la diffusion spatiale de la réverbération.

III.2.6 Analyse statistique classique

III.2.6.1 Analyse de variance

Nous devons d'abord nous demander si les résultats des tests sont significatifs, c.-à-d. si les sujets ont réellement entendu des différences, et n'ont pas répondu au hasard. Nous effectuons pour cela une analyse de la variance des résultats. Cette méthode est décrite dans l'annexe B.1.2.

L'idée est de tester l'hypothèse selon laquelle les moyennes sur les individus sont identiques entre les paires. Pour cela, on forme le rapport entre la variance due aux stimuli et la variance résiduelle (différence entre la variance totale et la variance due aux stimuli) ; comme ces deux variances sont des variables aléatoires indépendantes, leur rapport suit une loi F de Fisher. Il ne reste plus qu'à calculer la probabilité de rejet de l'hypothèse nulle. Les résultats sont rassemblés dans le tableau III.7. Nous en concluons que les trois tests d'écoute sont globalement significatifs.

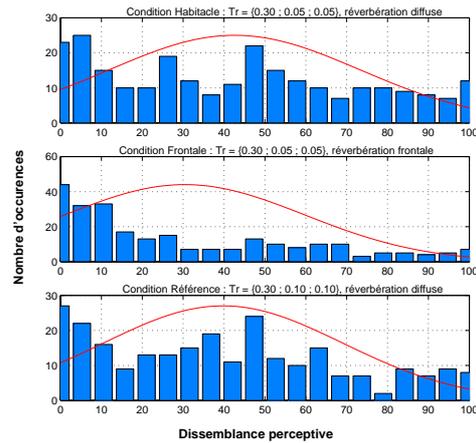


FIGURE III.6 – Distributions de l'ensemble complet des jugements de dissemblance, sur toutes les paires et tous les sujets ; superposition d'une gaussienne de même moyenne et de même écart type.

TABLEAU III.8 – Moyennes et écarts types de l'ensemble des individus sur toutes les paires

Condition Habitacle	23 < moyennes < 60 19 < écarts types < 42
Condition Frontale	14 < moyennes < 52 12 < écarts types < 39
Condition Référence	23 < moyennes < 66 19 < écarts types < 37

III.2.6.2 Analyse des distributions des jugements de dissemblance

Les trois distributions de l'ensemble des dix-sept sujets sur l'ensemble des quinze paires sont présentées figure III.6. Aucune des distributions n'est gaussienne, ce qui montre que les sujets n'ont pas – *a priori* – répondu au hasard ; ce fait important sera confirmé dans la suite de notre étude. Pour chaque condition, les notes comprises entre 0 et 10, reflétant une dissemblance quasi nulle, forment un grand ensemble, spécialement pour la condition *Frontale* ; ceci peut provenir de sujets n'étant pas en mesure de focaliser leur attention sur ce qu'on a voulu leur faire écouter.

Pour la condition *Habitacle*, mis à part la forte proportion des notes faibles, la distribution est à peu près régulière, avec une forte proportion de notes aux alentours de 25 et 48. Il en est de même pour la condition *Référence*, avec une forte proportion de notes aux alentours de 38 et 48. Par contre, il n'en est pas de même pour la condition *Frontale*, pour laquelle l'ensemble des sujets a détecté des différences plutôt faibles.

Intéressons-nous ensuite aux distributions de chaque individu. Elles ne sont pas uniformes, traduisant le fait que les sujets ont réellement remarqué des différences entre les paires, et n'ont pas répondu au hasard. Les moyennes entre les paires varient assez fortement d'un individu à l'autre, surtout pour la localisation et le timbre ; il en est de même pour les écarts types. Le tableau III.8 rassemble les résultats.

Nous mettons en évidence des différences importantes dans les comportements des différents sujets : certains sont plus sensibles que d'autres à tel ou tel aspect, ce qui se traduit par une moyenne des notes plus ou moins importante. D'autre part, nous remarquons que les sujets n'utilisent pas tous l'échelle de notation de manière identique : certains sujets n'en ont utilisé que la moitié, et d'autres ne l'ont pas utilisé de manière homogène. Il est donc nécessaire de *normaliser* les notes de chaque individu par leur

TABLEAU III.9 – Moyennes et écarts types de l'ensemble des paires sur tous les individus

Condition Habitable	8 < moyennes < 83 13 < écarts types < 30
Condition Frontale	9 < moyennes < 75 6 < écarts types < 31
Condition Référence	11 < moyennes < 72 13 < écarts types < 31

variance, ce qui revient à donner à chaque individu la même importance quelle que soit sa dispersion : on rétablit l'équilibre entre les individus en leur donnant à tous une variance unité.

L'examen des distributions de chaque paire sur tous les individus fait apparaître une forme proche d'une gaussienne. Le nombre de sujets commence à être suffisant pour appliquer le théorème de la limite centrale. Nous confirmons nos observations par un test de comparaison avec une distribution gaussienne idéale, de même moyenne et de même écart type ; ce test fait intervenir une loi du χ^2 à seize degrés de libertés. En choisissant comme hypothèse nulle l'identité entre les deux lois, et un facteur de risque relativement important, nous pouvons considérer la plupart des distributions testées comme gaussiennes.

Les moyennes interindividuelles des notes varient fortement d'une paire à une autre ; il en est de même pour les écarts types. Le tableau III.9 rassemble les résultats. Remarquons que les écarts types sont assez importants, en regard de la moyenne ; ceci peut s'expliquer par les différences d'utilisation de l'échelle de notation, et laisse présager une certaine difficulté dans l'extraction de résultats fiables.

Poursuivons l'analyse des distributions par la recherche de l'intervalle contenant les moyennes et écarts types *théoriques* des distributions des notes des différentes paires, sur tous les individus. Les bornes de ces deux intervalles ont déjà été précisées au § II.4.7.2 ; un calcul rigoureux de ces bornes suppose que les distributions sont gaussiennes, mais on peut considérer les résultats fiables lorsque l'effectif de la population est au moins égal à une dizaine d'individus [18].

La figure III.7 présente les résultats concernant les moyennes et les écarts types des distributions des notes. Les quantités estimées sont représentées par des *, tandis que les intervalles contenant les quantités théoriques sont représentés par des barres verticales. Le risque est choisi égal à 5 %.

Les intervalles contenant les moyennes théoriques des dissemblances permettent de mettre en évidence les précautions à prendre lors de conclusions trop hâtives en regardant uniquement les valeurs des moyennes elles-mêmes. Si les intervalles contenant les moyennes théoriques de deux paires se chevauchent, il est possible qu'il n'y ait pas de différences entre ces paires. C'est le cas par exemple des paires 7 et 8 pour la condition *Habitable*, et de plus de la moitié des paires de la condition *Frontale* ! En ce qui concerne les écarts types, bien que ceux-ci soient assez importants, les intervalles contenant les valeurs idéales sont relativement faibles, ce qui indique une certaine stabilité des résultats ; notamment, pour la condition *Habitable*, l'écart type des dissemblances des paires 2, 5, 9 et 14 est le plus faible, et l'intervalle contenant l'écart type théorique est très petit.

Terminons l'analyse des distributions par l'examen des diagrammes en boîtes, encore appelés *box plot*. Il s'agit d'une représentation schématique qui montre les principales caractéristiques d'une variable numérique en utilisant les quartiles ; si l'on suppose connue la distribution théorique d'une variable aléatoire X de densité de probabilité $f(x)$, alors la fonction de répartition vaut par définition $F(x) = \int_{-\infty}^x f(t) dt$, et les quartiles Q_1 , Q_2 et Q_3 sont définis par $F(Q_1) = 0.25$, $F(Q_2) = 0.50$ et $F(Q_3) = 0.75$. La partie centrale de la distribution est représentée par une boîte de largeur arbitraire et dont la longueur correspond à l'intervalle entre le premier et le troisième quartile. On trace alors à

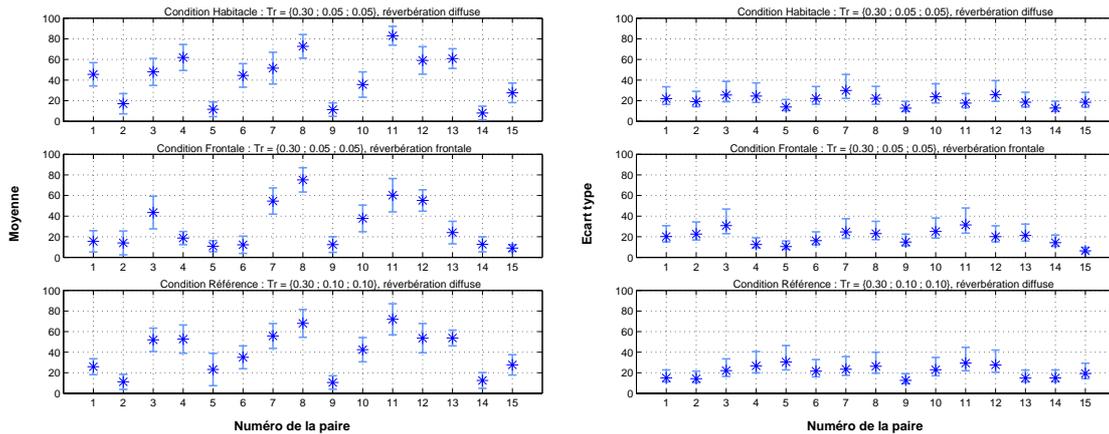


FIGURE III.7 – Encadrement contenant la moyenne théorique (à gauche) et l'écart type théorique (à droite) des distributions des dissemblances entre les stimuli, pour l'ensemble des individus. En abscisse : le numéro de la paire, en ordonnée : la valeur de la dissemblance

l'intérieur la position de la médiane, correspondant au deuxième quartile. La boîte est complétée par des « moustaches » correspondant aux valeurs adjacentes :

- adjacente supérieure, c.-à-d. la plus grande valeur inférieure à $Q_3 + 1.5(Q_3 - Q_1)$;
- adjacente inférieure, c.-à-d. la plus petite valeur supérieure à $Q_1 - 1.5(Q_3 - Q_1)$.

Les valeurs extérieures représentées par des + sont celles qui sortent des « moustaches ». Enfin, la boîte comporte des entailles représentant l'intervalle de confiance à 95% de la médiane. Un tel graphique facilite la comparaison de plusieurs distributions ; en particulier, il est plus aisé de comparer des diagrammes en boîtes que de comparer des histogrammes.

Nous avons utilisé des diagrammes en boîtes pour comparer les distributions des valeurs de dissemblances de chaque paire sur tous les individus.

CONDITION HABITACLE (cf. fig. III.8)

Modification globale de la densité d'énergie

Les paires 2 {1A;2A}, 5 {1B;2B}, 9 {2A;3A} et 14 {2B;3B} correspondent, dans le tableau III.2(a), à un changement de ligne, la colonne restant identique. Il s'agit donc d'un changement d'énergie totale d'une graduation, avec inversion de l'énergie entre la section OD et la section Rev ; la modification énergétique est identique sur chacune des deux sections. Les dissemblances perçues se situent à un niveau très faible, voisin de 5 sur une échelle de 100. On peut expliquer cette valeur par le fait que le changement énergétique sur chaque section est trop faible pour permettre de différencier les stimuli, bien que la modification énergétique totale soit de +3 dB en BF et -3 dB en HF.

Les paires 7 {1A;3A} et 12 {1B;3B} correspondent à un changement de deux lignes, la colonne restant identique. Il y a changement d'énergie totale de deux graduations, avec inversion de l'énergie entre les deux sections. Les dissemblances perçues sont voisines de 55. Le changement de spectre est ici important, et il est réparti identiquement entre OD et Rev. Les différences sont donc nettement plus importantes que pour les paires 2, 5, 9 et 14. La dissemblance au sein de la paire 7 est moins importante qu'au sein de la paire 12 car Rev est plus faible en hautes fréquences et OD plus faible en basses fréquences. La dissemblance au sein de la paire 7 est du même ordre qu'au sein de la paire 1 car le changement sur OD est identique même si l'énergie de Rev a fortement augmenté ou diminué en hautes fréquences (les changements sur Rev sont comme « gommés » par les changements sur OD). Les dissemblances au sein de la paire 12 sont plus fortes qu'au sein de la paire 15 car cette fois les différences sur OD laissent passer les changements sur Rev (de plus il y a modification du spectre de l'énergie totale

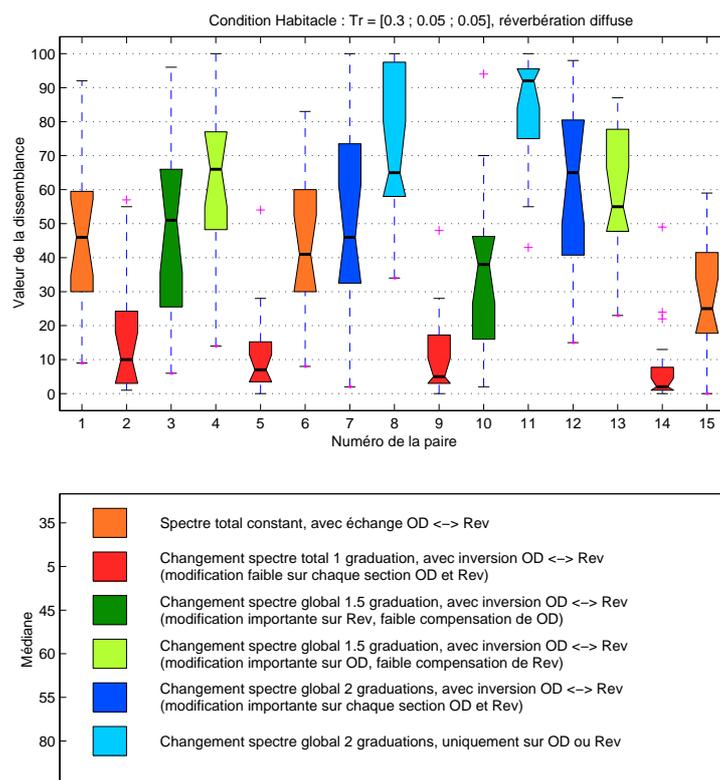


FIGURE III.8 – Diagrammes en boîtes pour la condition d'écoute Habitable

au sein de la paire 12, et non au sein de la paire 15).

Echange d'énergie entre le son direct et la réverbération

Les paires 1 {1A;1B}, 6 {2A;2B} et 15 {3A;3B} correspondent, dans le tableau III.2(a), à un changement de colonne, la ligne restant identique. L'énergie totale est identique, mais il y a échange (ou compensation) énergétique de deux graduations entre les sections OD et Rev (ce qui est perdu par l'une est gagné par l'autre). Les dissemblances perçues se situent à un niveau plus élevé, voisin de 35. Le changement énergétique sur chaque section est suffisamment important pour que les différences soient audibles. Les différences au sein des configurations de type 1 sont plus importantes qu'au sein des configurations de type 3 ; ces différences peuvent s'expliquer par les variations fréquentielles du temps de réverbération : l'échange d'énergie en moyennes et hautes fréquences entre OD et Rev n'est pas exact.

Affectation particulière de l'énergie

Les paires 3 {2A;1B} et 10 {3A;2B} correspondent, dans le tableau III.2(a), à un changement diagonal d'un étage. Il y a changement d'énergie totale d'une graduation (le spectre de Rev est modifié d'une graduation et demie, tandis que le spectre de OD est modifié d'une demi-graduation dans l'autre sens). Les dissemblances perçues sont voisines de 45. La modification énergétique sur la section Rev est plus importante que dans la première série : elle est parfaitement audible, et la légère compensation de OD a tendance à atténuer les différences. Comme dans la deuxième série, une diminution trop importante de la section Rev en HF réduit les différences. Par contre, le changement étant plus important, et la compensation par OD plus faible, les différences sont plus fortes.

Les paires 4 {1A;2B} et 13 {2A;3B} correspondent également à un changement diagonal d'un étage. Il y a également changement d'énergie totale d'une graduation, mais c'est le spectre de OD qui est modifié d'une graduation et demie, tandis que le spectre de Rev est modifié d'une demi-graduation dans

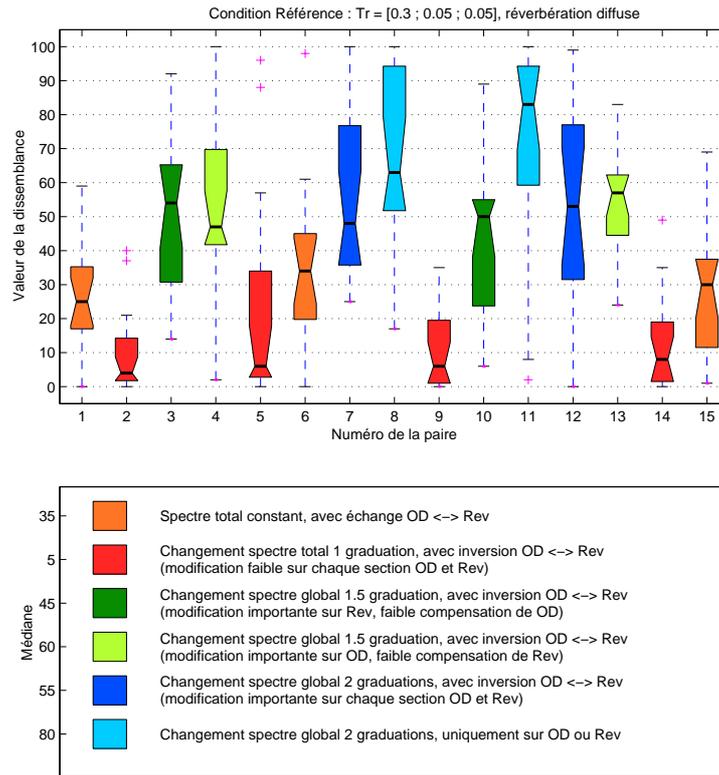


FIGURE III.9 – Diagrammes en boîtes pour la condition d'écoute Référence

l'autre sens. Les dissemblances perçues sont de l'ordre de 60. Pour ces paires, le changement énergétique s'effectue principalement sur OD, et les différences sont plus importantes que lorsqu'il s'effectue sur Rev. Comme dans les séries précédentes, lorsque le niveau de Rev devient trop faible, les dissemblances sont moins audibles. La dissemblance au sein de la paire 4 est plus faible qu'au sein de la paire 11 car les différences sur OD sont plus faibles et en partie compensées par Rev ; on peut tenir le même raisonnement pour les paires 3 et 8.

Les paires 8 {1B;3A} et 11 {1A;3B} correspondent à un changement de deux lignes et deux colonnes en diagonale. Il y a changement d'énergie totale de deux graduations, mais séparément sur OD et Rev. Les distances perçues sont les plus fortes, de l'ordre de 80. Les dissemblances sont plus importantes que pour les paires 7 et 12. La dissemblance au sein de la paire 11 est plus importante qu'au sein de la paire 8 car les changements ont lieu sur OD.

En conclusion, on peut dire qu'il est déjà possible de construire un estimateur de la dissemblance globale, en ajoutant séparément les modifications spectrales apportées au son direct et à la réverbération, plutôt qu'en tenant compte uniquement de la modification spectrale globale. À ce stade de notre analyse, il semble que le contenu spectral du son direct et celui de la réverbération doivent être considérés comme deux variables indépendantes ; toutefois, il semble nécessaire d'introduire une pondération privilégiant le contenu spectral du son direct. Cette hypothèse sera confirmée par l'analyse multidimensionnelle.

CONDITION RÉFÉRENCE (cf. fig. III.9)

Les dissemblances entre les paires 1, 6 et 15 sont légèrement réduite par rapport à la condition précédente ; ceci peut s'expliquer par la constance du temps de réverbération. La dissemblance de la paire 12 est plus faible que dans la condition *Habitacle* car le niveau de Rev est plus élevé en HF.

Le changement de spectre total influe directement notre perception, mais pour un même changement

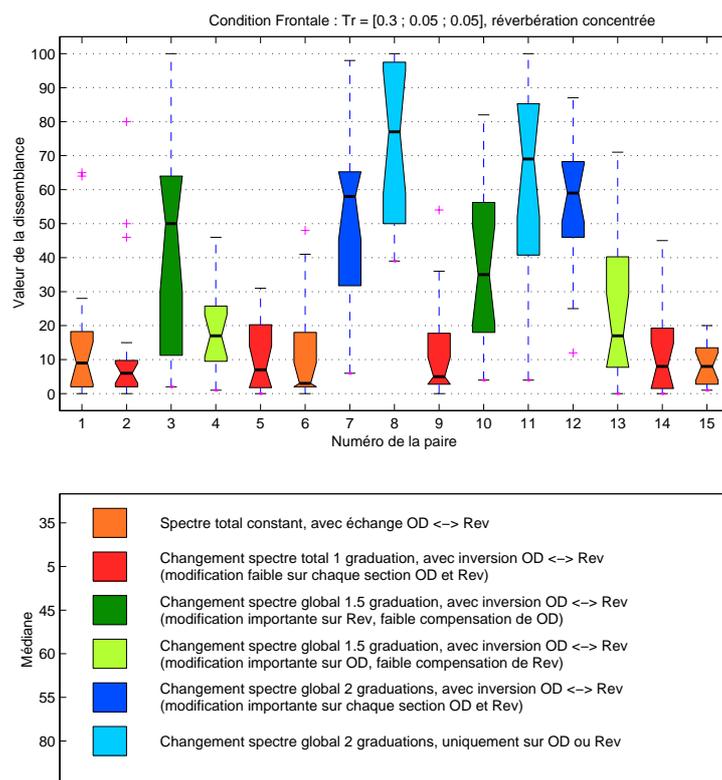


FIGURE III.10 – Diagrammes en boîtes pour la condition d'écoute Frontale

de spectre total, notre perception est gouvernée par l'énergie du son direct. Entre les paires 7, 12, 8 et 11, nous effectuons un changement de spectre total de deux graduations : les dissemblances sont identiques entre les paires 7 et 12 (la modification énergétique est la même), mais elles sont plus importantes pour la paire 8, et surtout pour la paire 11. Entre les paires 8 et 12, les dissemblances sont globalement identiques : que l'on effectue un changement de spectre total réparti sur OD et Rev, ou uniquement sur Rev, les influences sont faibles. Par contre, entre les paires 7 et 11, les différences sont importantes : on perçoit plus facilement les changements sur OD. Ajoutons que ce phénomène s'amplifie lorsque le temps de réverbération n'est plus constant.

Les observations sont donc globalement identiques à la condition *Habitacle*, à quelques nuances près. Aussi peut-on conclure que les variations fréquentielles du temps de réverbération introduites pour la condition *Habitacle* n'ont pas modifié sensiblement la structure perceptive du test, en dépit de valeurs objectives différentes de la condition *Référence*.

CONDITION FRONTALE (cf. fig. III.10)

Les performances de discrimination sont nettement réduites par rapport aux conditions précédentes et la structure perceptive est modifiée. Les changements de spectre sur OD et Rev, sous forme d'échange (c.-à-d. les paires 1, 6 et 15), ne donnent plus lieu à une perception différente ; ceci démontre l'importance de l'aspect spatial de la réverbération. On constate aussi une inversion de l'effet préférentiel de OD par rapport à Rev : les dissemblances au sein des paires 3 et 10 sont supérieures à celles au sein des paires 4 et 13 ; il en est de même pour la paire 8 par rapport à la paire 11. Ceci peut s'expliquer par le fait que si la réverbération est diffusée par des canaux de sortie non corrélés, ce qui est le cas dans les conditions *Référence* et *Habitacle*, les colorations éventuelles sont atténuées. Inversement, ces colorations sont amplifiées si la réverbération est diffusée par un seul canal.

TABLEAU III.10 – Paramètres de décision concernant le choix du nombre de dimensions du modèle d'analyse INDSCAL : stress, et variance expliquée

ndim	Condition Habitable		Condition Frontale		Condition Référence	
	stress	VE	stress	VE	stress	VE
1	0,40	0,26	0,32	0,52	0,40	0,29
2	0,17	0,65	0,18	0,65	0,17	0,62
3	0,13	0,66	0,14	0,65	0,12	0,63

En conclusion, c'est donc grâce à la distribution spatiale que les sujets ont pu discriminer les différences spectrales des contributions directe et réverbérées dans les conditions *Habitable* et *Référence*, et ce en dépit d'un support temporel très court. Lorsque les contributions directe et réverbérée sont localisées de manière identique la perception du timbre devient globale.

III.2.7 Analyse multidimensionnelle

L'analyse statistique classique nous a permis de dresser des constatations locales quant aux résultats de nos trois tests d'écoute. Comme souvent dans de pareils tests, l'échelle de notation n'a pas été utilisée de manière uniforme par tous les individus. L'examen des distributions des jugements de dissemblance, notamment par le biais des diagrammes en boîtes, permet de repérer globalement les paires jugées plus différentes que d'autres. Les paires de stimuli présentant une modification importante de la densité énergétique du son direct ont été jugées très différentes, et le phénomène s'amplifie lorsque le temps de réverbération n'est pas constant ; par contre, ces différences diminuent si la réverbération n'est plus diffuse. En outre, l'analyse de variance, étape indispensable à toute analyse statistique d'un test de comparaison par paires, se révèle largement positive.

Cependant, les méthodes d'analyse exposées ci-dessus se montrent insuffisantes pour affiner les conclusions, car elles ne prennent pas en compte les liaisons existant entre les individus, ou entre les dissemblances, et qui peuvent revêtir un aspect important. Il faut donc analyser les données en tenant compte de leur caractère multidimensionnel, et surtout de la sensibilité propre de chaque sujet.

Nous n'avons pas pu avoir recours à la méthode d'analyse EXSCAL, décrite au § B.2.4, à cause du faible nombre de stimuli (six stimuli, soit quinze paires), car les résultats n'auraient pas été fiables. Nous avons donc utilisé la méthode INDSCAL (cf. § B.2.3), moins riche, car bien qu'elle prenne en compte les particularités de jugement de chaque sujet, elle ne permet pas d'introduire des spécificités propres à chaque stimulus. Cette méthode a été utilisée avec le programme SYSTAT[®] 1 [322] fonctionnant sous Windows[™].

L'objectif de cette méthode est de reconstruire un espace à plusieurs dimensions dans lequel on va placer les stimuli. Afin de déterminer le nombre de dimensions optimal, il faut chercher à rendre le stress minimal, et la variance expliquée maximale. D'après les résultats rassemblés dans le tableau III.10, nous constatons que la troisième dimension n'apporte pas d'amélioration pour l'adéquation du modèle aux données ; de plus, il est parfaitement illusoire de décrire un espace à trois dimensions avec aussi peu de stimuli. Nous choisissons donc une représentation en deux dimensions.

CONDITION HABITACLE

Sur le plan sujets (cf. fig. III.11(a)), on remarque que les sujets sont assez regroupés, ce qui indique une concordance certaine dans les jugements. La corrélation entre les poids des différents sujets sur les deux axes vaut 0,98, et permet donc d'envisager une rotation de la configuration du plan stimuli.

1. <http://www.spssscience.com/SYSTAT>

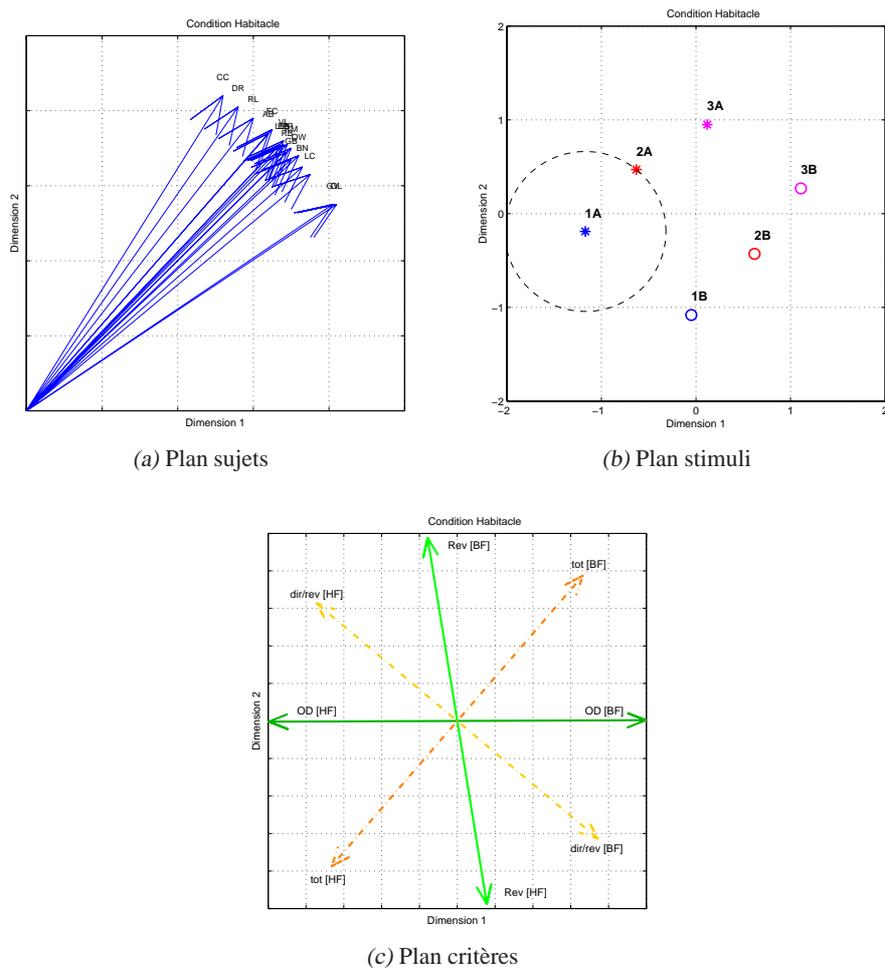


FIGURE III.11 – Résultats de l'analyse INDSCAL en deux dimensions pour la condition d'écoute Habitacle

Sur le plan stimuli (cf. fig. III.11(b)), la variance expliquée totale est de 0,65 ; la variance expliquée par chacun des axes représente respectivement 49% et 51% de la variance expliquée totale. Celle-ci, notée R^2 , est calculée comme étant le carré de la moyenne des corrélations sur l'ensemble des sujets entre les produits scalaires initiaux et ceux estimés par le modèle (cf. § B.2.3) ; le pourcentage de variance pour chaque axe est estimé en pondérant la variance expliquée totale R^2 par le poids normalisé moyen que les sujets ont accordé à cet axe.

Les segments de droite représentant les modifications du rapport entre la densité spectrale d'énergie du son direct et celle de la réverbération, à énergie totale constante (c.-à-d. les segments [1A-1B], [2A-2B] et [3A-3B]) sont parallèles entre eux, et orientés approximativement le long de la deuxième diagonale. Les segments de droite représentant les modifications de l'énergie totale, alors que le rapport entre la densité spectrale d'énergie du son direct et celle de la réverbération reste identique (c.-à-d. les segments [1A-2A-3A] et [1B-2B-3B]) sont également parallèles entre eux, et orientés approximativement le long de la première diagonale. Ces deux ensembles de segments sont orthogonaux entre eux ; les deux critères objectifs sont donc indépendants dans le plan. Ces deux critères étant objectivement non corrélés dans la construction des configurations, la valeur de S_{12}^* égale à 0,04 permet de conclure à l'indépendance des facteurs perceptifs correspondants (S_{12}^* mesure la quantité d'information conservée par les axes 1 et 2 par rapport à la quantité d'information injectée *a priori* par les valeurs des critères des configurations).

La même conclusion peut être tirée pour les segments [1A-3B] et [1B-3A] qui représentent respectivement les variations de la densité spectrale d'énergie du son direct à énergie réverbérée constante, et les variations de la densité spectrale d'énergie de la réverbération à énergie directe constante. On peut ajouter que ces segments de droite sont approximativement parallèles aux axes fournis par l'analyse.

En accord avec les valeurs brutes de dissemblances et un paramètre interne de l'analyse INDSCAL (en l'occurrence la constante ajoutée aux dissemblances pour leur faire respecter l'inégalité triangulaire), il est possible d'estimer la limite inférieure de perception d'une différence ; celle-ci est représenté par un cercle centré sur le stimulus 1A. Grâce à ce seuil, il est aisé de constater que tous les stimuli peuvent être discriminés à l'exception de ceux faisant partie des paires {1A;2A}, {2A;3A}, {1B;2B} et {2B;3B}.

Le plan critères (cf. fig. III.11(c)) permet de visualiser la pertinence des critères objectifs dans le domaine perceptif. Plus la norme d'un vecteur est grande, plus il est globalement corrélé avec le plan stimuli, et plus sa direction se rapproche de celle d'un axe, plus il traduit objectivement l'effet perceptif caractérisé par cet axe. Les vecteurs dessinés dans ce plan ont comme composantes les corrélations entre les valeurs des critères objectifs et les coordonnées des configurations sur les différents axes.

Les axes fournis par l'analyse sont donc très fortement corrélés à la densité spectrale d'énergie du son direct, et à celle de la réverbération. Les premières diagonales sont fortement corrélés à la densité spectrale de l'énergie totale et à celle du rapport énergétique entre le son direct et la réverbération. Les positions respectives des critères confirment que les variations fréquentielles de l'énergie du son direct sont non corrélées avec les variations fréquentielles de l'énergie de la réverbération, et que les variations fréquentielles de l'énergie totale sont non corrélées avec les variations fréquentielles du rapport énergétique entre le son direct et la réverbération.

Une analyse en composantes principales (cf. § B.2.1) a été effectuée sur les coordonnées de chacun des stimuli dans le plan stimuli issu de l'analyse. Elle permet de détecter l'axe supportant le maximum de variance : il est à $+9^\circ$ par rapport à l'axe 1, et rassemble 57% de la variance des coordonnées. On peut donc effectuer une rotation de -9° de la configuration des stimuli dans le plan pour obtenir une variance maximale le long du premier axe.

À ce stade, aucune conclusion ne peut être donnée quant au lien entre les critères objectifs et les facteurs perceptifs. Comme on peut effectuer une rotation des stimuli dans le plan, on peut choisir de les expliquer aussi bien par le couple {énergie du son direct ; énergie de la réverbération} que par le couple

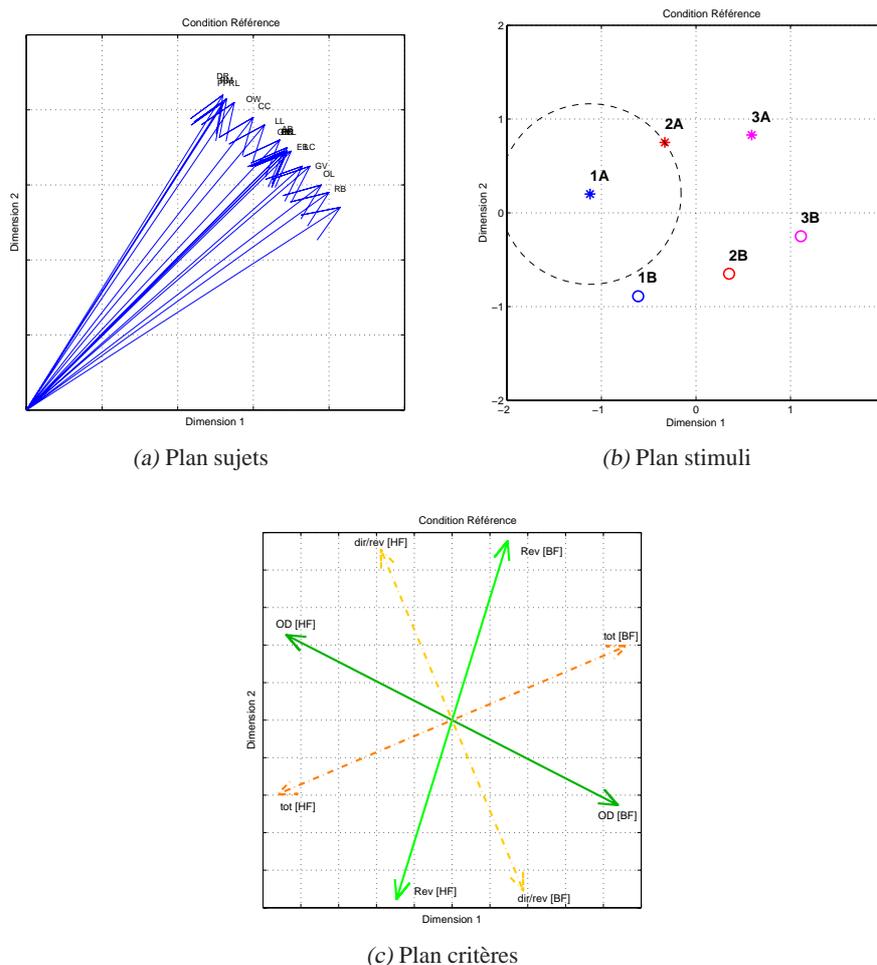


FIGURE III.12 – Résultats de l'analyse INDSICAL en deux dimensions pour la condition d'écoute Référence

{énergie totale ; rapport énergétique entre le son direct et la réverbération}. On ne peut pas déterminer les axes explicatifs.

Comme la distance entre les stimuli 1A et 3B est supérieure à la distance entre les stimuli 1B et 3A, cette condition *Habitacle* montre que, pour les stimuli considérés, nous sommes plus sensibles aux changements du contenu spectral du son direct qu'à ceux de la réverbération. Cependant l'existence des deux dimensions et la comparaison des distances avec l'estimation du seuil de discrimination montre que les sujets ont opéré, de manière sous-jacente, une distinction entre les modifications spectrales de la contribution directe et de la contribution réverbérée.

CONDITION RÉFÉRENCE

Les conclusions sont en tout point identiques à celles de la condition *Habitacle*. Seul le plan critères se distingue (cf. fig. III.12(c)) : bien que les critères objectifs soient toujours fortement corrélés avec le plan stimuli, les axes issus de l'analyse ne sont plus corrélés avec l'un ou l'autre des critères objectifs. Par ailleurs on observe que les discriminations des paires jouant sur l'échange des timbres entre les contributions directes et réverbérées sont un peu moins aisées que dans la condition *Habitacle*. On peut conclure que les variations fréquentielles du temps de réverbération, telles que rencontrées dans un habitacle de véhicule, vont donc faciliter la discrimination entre les contenus fréquentiels direct et réverbéré.

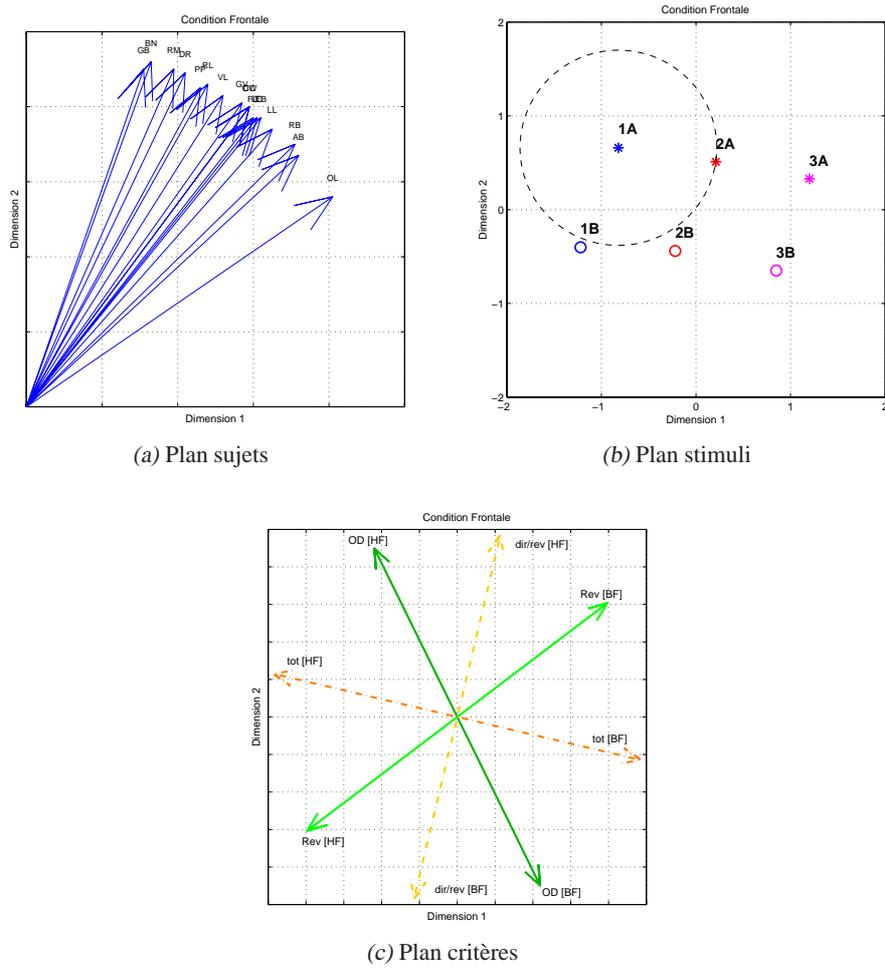


FIGURE III.13 – Résultats de l'analyse INDSCAL en deux dimensions pour la condition d'écoute Frontale

CONDITION FRONTALE

Sur le plan sujets (cf. fig. III.13(a)), on remarque que les sujets sont un peu moins regroupés que dans la condition *Habitacle*, mais la corrélation entre les poids des différents sujets sur les axes est égale à 0,94, valeur qui permet toujours d'envisager une rotation de la configuration des stimuli dans le plan correspondant.

Sur le plan stimuli (cf. fig. III.13(b)), la variance expliquée totale est de 0,65 ; la variance expliquée par chacun des axes représente respectivement 67% et 33% de la variance expliquée totale.

La configuration des stimuli dans le plan est globalement identique à la condition *Habitacle*, avec une prépondérance plus forte des changements importants de l'énergie totale. La valeur de S_{12}^* voisine de 0 confirme, comme dans la condition *Habitacle*, l'indépendance des facteurs perceptifs décrits par les axes du plan. Mais nous devons remarquer dans cette condition une influence très forte de la constante additive (cf. l'algorithme de INDSCAL) : alors que sur le diagramme en boîtes, seules les dissemblances entre les paires {1A;3A}, {1B;3B}, {1B;3A} et {1A;3B} ont réellement une importance, la perception des changements dans le rapport entre l'énergie du son direct et l'énergie de la réverbération étant la plus importante, elle conduit à construire obligatoirement une configuration à deux dimensions.

Le respect de l'inégalité triangulaire impose d'augmenter « artificiellement » les distances entre les paires {1A;1B}, {2A;2B}, {3A;3B} et {1A;2A}, {2A;3A}, {1B;2B}, {2B;3B}. La valeur du seuil de perception de la différence nous montre qu'il n'est plus possible d'envisager une discrimination au sein de ces paires, correspondant aux modifications de timbre opérées soit sur le son direct soit sur la réverbération.

Sur le plan critères (cf. fig. III.13(c)), on remarque, comme dans la condition *Habitacle*, que tous les critères sont fortement corrélés avec le plan stimuli. Le premier axe de l'analyse est corrélé fortement avec l'énergie totale, tandis que le second est corrélé au rapport direct sur réverbéré. Cependant, compte tenu de la remarque précédente, cette deuxième dimension a peu de signification puisqu'elle ne s'explique que par l'ajout artificiel de la constante additive.

Une analyse en composantes principales a également été effectuée sur les coordonnées de chacun des stimuli dans le plan issu de l'analyse. L'axe supportant le maximum de variance est le premier axe issu de l'analyse ; il rassemble cette fois 74% de la variance totale des stimuli.

Si on suppose que, comme dans la condition *Habitacle*, tous les sujets sont regroupés dans le plan sujets, les configurations du plan stimuli peuvent tourner ; on ne peut donc pas déterminer les axes préférés, qui seraient là où les sujets entendent des différences. Pour pouvoir le faire, il faudrait trouver deux directions dans lesquelles certains sujets vont dilater le plan stimuli, et d'autres non ; ces deux axes seraient donc significatifs.

Mais comme les stimuli des deux conditions sont construits de manière identique, on peut assimiler cette nouvelle condition à un « deuxième sujet » sur lequel on aurait effectué une analyse MDSCAL. En regardant les variances des configurations, on s'aperçoit que, dans cette condition, il y a forte dilatation du premier axe issu de l'analyse ; ce premier axe est fortement prépondérant, alors que le deuxième est beaucoup moins important : la condition *Frontale* est fortement influencée par l'énergie totale, alors que l'importance du rapport énergétique entre le son direct et la réverbération a chuté. Cela permet de *choisir* un couple de critères plutôt qu'un autre : on choisit d'expliquer le plan stimuli par le couple de critères indépendants {énergie totale ; rapport énergétique entre le son direct et la réverbération}.

Pour la condition *Frontale*, on conclut à une indissolubilité entre la densité spectrale de l'énergie du son direct et celle de la réverbération (et non masquage car dans ce cas on aurait repéré une forte influence du son direct) : on n'entend plus les variations de timbre du son direct, ni celles de la réverbé-

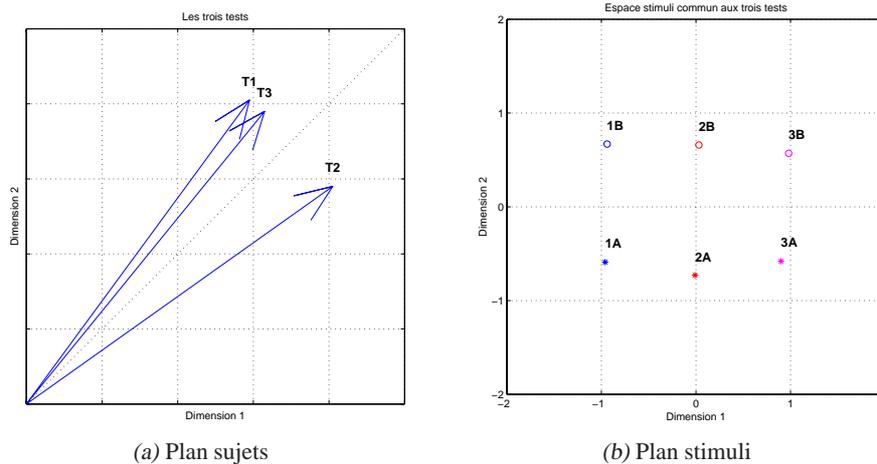


FIGURE III.14 – Résultats de l'analyse INDSCAL effectuée à partir des résultats de l'analyse de chacun des trois conditions

ration, mais uniquement celles de l'énergie totale. En outre, ce test confirme que l'aspect spatial de la réverbération fait partie intégrante de celle-ci.

Remarque générale concernant les trois tests : les sujets étant fortement regroupés dans le plan sujets, avec des valeurs de corrélation supérieures à 0,90, on aurait pu se contenter d'une analyse MDSCAL effectuée à partir d'une moyenne des matrices de dissemblances de tous les sujets. Il s'avère que les résultats sont en tout point identiques, ce qui confirme le fait que les sujets ont tous utilisé avec la même force les dimensions perceptives sous-jacentes pour effectuer leur jugement.

ANALYSE COMPLÉMENTAIRE

Pour chaque condition, les distances entre stimuli ont été calculées à partir des coordonnées des stimuli fournies par l'analyse INDSCAL, et ont été assimilées à des dissemblances récoltées par un test d'écoute, avant d'être injectées dans une analyse INDSCAL, comme si l'on disposait de trois sujets. Le plan stimuli calculé (cf. fig. III.14(a)) est semblable à celui de chacune des trois conditions individuelles ; il représente l'espace commun aux trois conditions. La variance expliquée totale est de 0,95 ; la variance expliquée par chacun des axes représente respectivement 47% et 53% de la variance expliquée totale. Cet espace commun est orienté selon deux axes représentant la densité spectrale de l'énergie totale et celle du rapport de la densité spectrale de l'énergie du son direct et de celle de la réverbération. Il nous semble donc plus opportun d'expliquer les facteurs perceptifs par le couple de critères objectifs {énergie totale ; rapport énergétique entre le son direct et la réverbération}.

Le plan sujets (cf. fig. III.14(b)) montre que l'espace de la condition *Frontale* est, par rapport aux deux autres conditions, plus dilaté vers la densité spectrale de l'énergie totale, ce qui confirme la prépondérance de ce critère dans la condition *Frontale*.

III.2.8 Conclusions du test d'écoute

Trois tests d'écoute ont été réalisés, dans des conditions proches de celles rencontrées dans un habitacle automobile : temps de réverbération très faible (mais six fois plus élevé dans les basses fréquences), rapport énergétique important entre le son direct et la réverbération, position d'écoute non symétrique par rapport à celle du dispositif électroacoustique de restitution. Plusieurs configurations ont été testées, grâce auxquelles on a pu s'interroger sur l'influence du temps de réverbération, et sur l'influence de l'aspect

spatial de la réverbération. Pour chacune de ces configurations, six stimuli ont été comparés par paire ; les stimuli ont été construits de manière identique pour chaque condition : variations fréquentielles de l'énergie totale, du son direct et/ou de la réverbération.

La condition *Habitacle* (Tr plus élevé dans les basses fréquences, réverbération diffuse) montre que, dans les conditions d'écoute représentées par les six stimuli testés, notre perception peut être expliquée à l'aide de deux facteurs perceptifs représentant deux critères objectifs indépendants, deux entités sur lesquelles on est parfaitement capable de discerner les différences de spectre : l'énergie totale, et le rapport énergétique entre le son direct et la réverbération. En outre, notre perception est plutôt gouvernée par le son direct ; nous sommes plus sensible à un changement fréquentiel sur le son direct, qu'au même changement sur la réverbération.

La condition *Référence* (Tr identique dans les trois bandes, réverbération diffuse) confirme les impressions de la condition *Habitacle*, mais les différences entre les stimuli sont plus subtiles, car le contenu fréquentiel de la réverbération n'est pas «soutenu» par les variations du Tr.

La condition *Frontale* (Tr différent dans les trois bandes, identique à la condition *Habitacle*, réverbération frontale) montre que nous ne sommes plus en mesure de distinguer les modifications fréquentielles effectuées soit sur le son direct soit sur la réverbération : il y a fusion (mais pas masquage) entre les deux. Nous ne percevons plus que les changements de l'énergie totale. La distinction entre l'énergie du son direct et celle de la réverbération, mise en évidence lors de la condition *Habitacle*, s'appuie donc sur la distribution spatiale.

Il ressort de cette étude que, malgré la durée très courte de l'effet d'habitacle, notre perception distingue la répartition temps-fréquence de l'énergie, et n'est pas uniquement sensible au contenu fréquentiel global ; cette distinction repose sur l'aspect spatial de la réverbération. Dans l'optique de traitements de correction fréquentielle, il est donc souhaitable d'égaliser séparément le son direct et l'effet de l'habitacle. Si une seule égalisation est possible, pour des raisons de coût de calcul par exemple, il est plus pertinent de la réaliser sur l'entité adéquate (son direct ou effet d'habitacle) pour corriger un défaut, que de la réaliser globalement sur l'énergie totale.

III.2.9 Implantation en véhicule

Nous envisageons d'effectuer une correction différente sur le son direct et le champ réverbéré. Dans ce paragraphe, nous cherchons à mettre à plat le spectre résultant, bien que ce ne soit pas perceptivement le meilleur but à atteindre dans un habitacle automobile. Il est en effet largement admis que les occupants d'un véhicule ont tendance à apprécier une certaine augmentation de niveau dans le bas medium du spectre [38] ; ceci peut s'expliquer soit par le masquage de ces basses fréquences par le bruit du moteur et le bruit de roulement, soit par le manque d'enveloppement sonore dû au faible volume de l'habitacle (cf. § III.4.1). Soit cette forme particulière, non uniforme, du spectre est considérée comme cible à atteindre lors du processus de calcul du filtre inverse (auquel cas elle est intégrée dans le filtre inverse), soit elle est considérée comme une correction supplémentaire. En effet, une fois les différents spectres mis à plat, rien n'empêche le concepteur du système audio d'apporter une correction supplémentaire, rendue nécessaire par la définition d'un cahier des charges issu d'un test auprès de la clientèle. Cette correction est le plus souvent implantée sous forme de filtres paramétriques [57, 245] ; on dispose alors d'une grande liberté pour sculpter la forme du spectre.

Toutes les corrections envisagées dans ce paragraphe sont des égalisations à phase minimale, c.-à-d. qu'elles ne concernent que le module des fonctions de transfert correspondantes (cf. § II.3.2.1) [229]. En particulier, l'excès de phase est laissé intact ; la réverbération n'est donc pas corrigée (puisqu'elle est incluse dans l'excès de phase [43]), de même que la localisation des sources sonores. Il s'agit donc uniquement d'une correction de *timbre*.

Étant donnée la configuration du dispositif électroacoustique de restitution sonore par rapport au conducteur ou au passager avant, nous considérons que l'auditeur est plus sensible au son direct provenant des haut-parleurs avant qu'à celui provenant des haut-parleurs arrière (en fait, le son direct est censé venir de l'avant). Nous cherchons donc à inverser le spectre du son direct provenant des haut-parleurs avant, et celui de la réverbération provenant des haut-parleurs arrière.

Pour l'inversion du spectre du son direct, la correction à appliquer sur les haut-parleurs avant, notée d , s'écrit

$$d \times OD = \text{constante}, \quad (\text{III.1})$$

avec OD la densité d'énergie du son direct mesuré sur les haut-parleurs avant ; la correction est réalisée dans le domaine fréquentiel, après avoir pris les précautions qui s'imposent (voir ci-dessous).

Nous voulons aussi inverser le spectre de la réverbération (ou plutôt de l'effet d'habitacle). Heureusement, les haut-parleurs arrière nous offrent un degré de liberté supplémentaire qui va nous permettre de mener l'égalisation du champ diffus. Il faut cependant tenir compte de la contribution de l'ensemble des haut-parleurs, en particulier ceux de l'avant auxquels on vient d'apporter une correction. En d'autres termes, la difficulté provient du fait que le son direct émis par les haut-parleurs avant « alimente » également la réverbération. En appliquant le principe selon lequel on peut ajouter de l'énergie, mais pas enlever, la correction à appliquer sur les haut-parleurs arrière, notée r , s'écrit

$$d \times \text{Rev}_{\text{avant}} + r \times \text{Rev}_{\text{arrière}} = \text{constante}, \quad (\text{III.2})$$

avec Rev la densité d'énergie de la réverbération ; elle est également réalisée dans le domaine fréquentiel.

La densité spectrale de l'énergie du son direct équivalent et celle de la réverbération sont déduites de l'analyse temps-fréquence des réponses impulsionnelles acoustiques de l'habitacle, mesurées entre chaque haut-parleur et un microphone omnidirectionnel placé approximativement au centre de la tête de l'auditeur (cf. § I.5.4). Plus précisément, le calcul est effectué grâce au relief de décroissance intégré :

- la densité spectrale de l'énergie de la réverbération est « lue » sur le relief de décroissance intégré, 5 ms après l'arrivée du son direct (c'est donc l'énergie restant dans la réponse impulsionnelle, au temps considéré) ;
- la densité spectrale du son direct équivalent est alors calculée comme étant la différence entre la densité spectrale de l'énergie totale et la densité spectrale de l'énergie de la réverbération précédemment calculée.

III.2.9.1 Correction optimisée pour le conducteur

Choisissons dans un premier temps une correction calculée pour le conducteur, selon le schéma de la figure III.15. L'inversion des spectres est effectuée dans le domaine spectral, après avoir pris les précautions élémentaires suivantes (cf. § II.3.2.2) :

1. Effectuer un lissage sur une échelle fréquentielle logarithmique, c.-à-d. à Q constant, en accord avec les propriétés de l'audition.
2. Limiter la dynamique, afin d'éviter les résonances ou anti-résonances trop marquées.
3. Limiter l'étendue fréquentielle, afin d'éviter l'amplification erronée des très basses ou très hautes fréquences.

Une fois l'inversion effectuée dans le domaine fréquentiel, se pose le problème de la modélisation soit par un filtre non récursif à réponse impulsionnelle finie (RIF), soit par un filtre récursif à réponse impulsionnelle infinie (RII).

Dans le cas de la synthèse d'un filtre RIF, la méthode la plus employée est la méthode des fenêtres (cf. § C.1.6). Elle fournit un filtre stable (puisque sans pôle), et assez « fouillé » pour peu que le nombre de coefficients soit assez important.

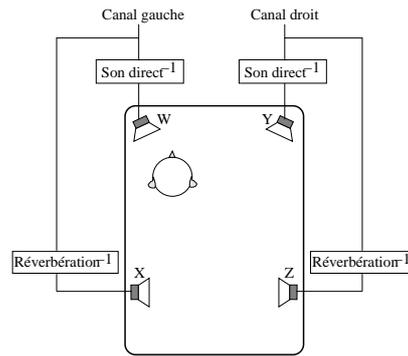


FIGURE III.15 – Schéma de l'implantation des filtres correctifs du timbre

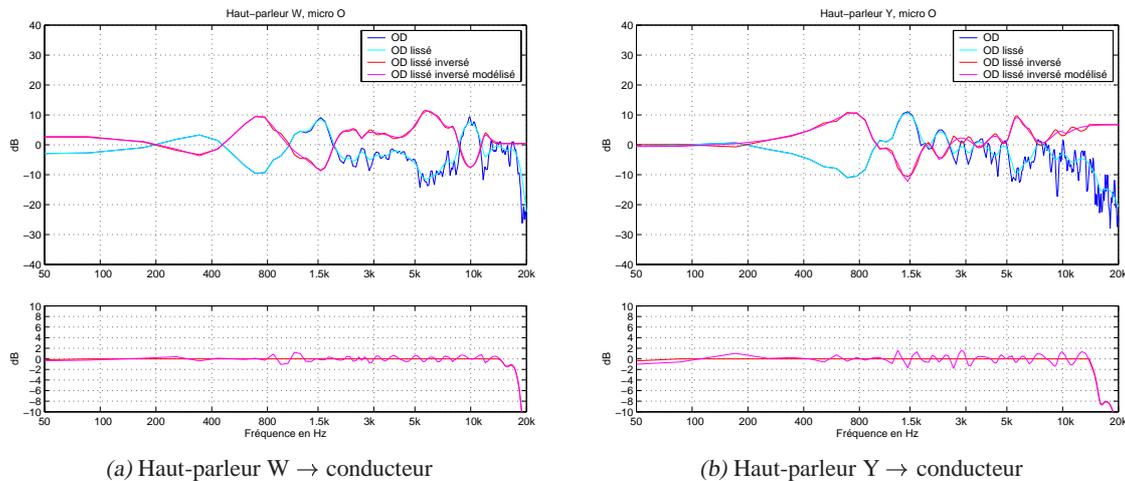


FIGURE III.16 – Inversion du son direct équivalent provenant des deux haut-parleurs avant ; comparaison entre une modélisation du filtre inverse sous forme RIF et sous forme RII d'ordre 16

En revanche, elle nécessite l'utilisation d'une technique de filtrage efficace (cf. § C.1.5), car le coût de calcul est très élevé.

Dans le cas de la synthèse d'un filtre RII, la méthode la plus efficace que nous avons testée est celle de Steiglitz-MacBride (cf. § C.1.7). L'algorithme ne converge pas toujours, et le filtre généré n'est pas nécessairement stable, mais l'erreur de modélisation reste généralement très faible. L'ordre du modèle n'excède jamais 20, et il est le plus souvent de 12 ou 16.

Le coût de calcul de l'opération de filtrage est de $2N + 1$ opérations par échantillons, où N est l'ordre du filtre.

En tout état de cause, l'étape d'inversion et de modélisation RIF ou RII requiert un certain nombre d'ajustements qui ne peuvent être automatisés.

La figure III.16 présente les résultats de l'inversion de la densité spectrale de puissance du son direct équivalent, calculé à partir d'une analyse temps-fréquence des réponses impulsionnelles des haut-parleurs avant-gauche et avant-droit dans un véhicule de type Safrane. Les filtres inverses sont appliqués uniquement aux haut-parleurs avant.

La figure III.17 présente les résultats de l'inversion de la densité spectrale de puissance de la réverbération, calculé à partir d'une analyse temps-fréquence des réponses impulsionnelles des haut-parleurs

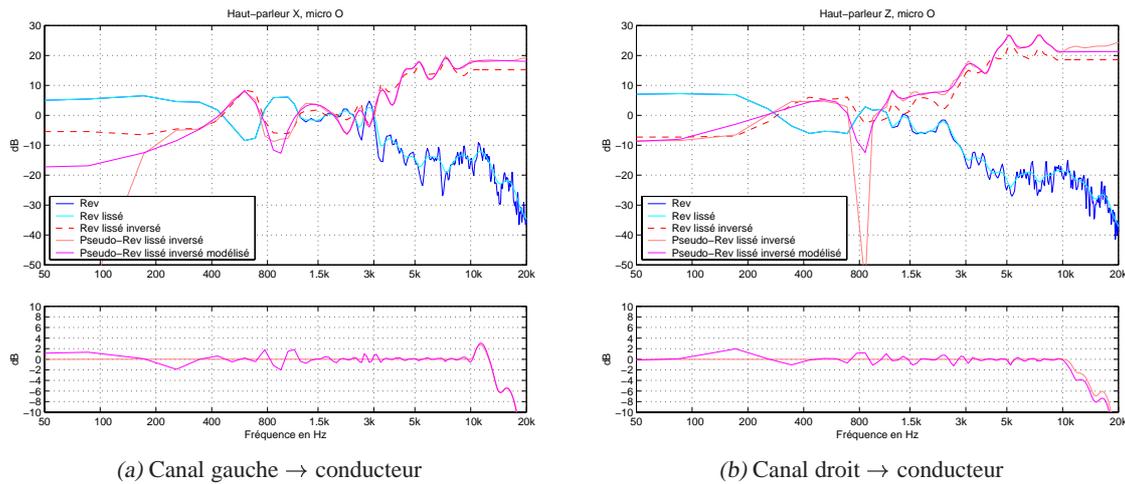


FIGURE III.17 – Inversion globale de l'effet d'habitacle, réalisé sur les haut-parleurs arrière ; comparaison entre une modélisation du filtre inverse sous forme RIF et sous forme RII d'ordre 16

arrière-gauche et arrière-droit dans le même véhicule. Les filtres inverses sont calculés de la manière indiquée à l'équation (III.2) ; ils sont appliqués uniquement aux haut-parleurs arrière.

On peut remarquer sur ces deux figures que l'étape de modélisation par un filtre récursif est hautement souhaitable, dans la mesure où le coût de calcul est très faible comparé à celui d'un filtre transverse, et l'erreur de modélisation est ici inférieure à 2 dB. Le nombre d'opérations par échantillons est de 33 multiplications-accumulations par filtre inverse d'ordre 16, soit 132 pour les quatre filtres.

Ces filtres ont été implantés dans le logiciel Max/MSP[®], et un test informel a été réalisé en aveugle dans le véhicule en question : de nombreux auditeurs, assis en position conducteur, ont écouté le système audio de base et le système audio égalisé par les filtres précédemment calculés ; il était possible de commuter instantanément entre le système audio brut et le système audio égalisé. Différents signaux sonores ont été utilisés : musique pop, musique classique, et dialogue entre un homme et une femme. Tous les sujets ayant écouté les deux configurations se sont prononcés en faveur du système corrigé.

Néanmoins, nous devons formuler plusieurs remarques. D'abord, les paramètres de réglage de l'inversion des filtres sont délicats à régler, en particulier il faut veiller à ne pas pousser le processus trop loin en fréquence (14 kHz sur l'avant et 10 kHz sur l'arrière), afin de ne pas trop remonter le niveau des aigus qui deviennent alors très agressifs ; de plus, la bande passante des haut-parleurs arrière est, par construction, bien plus faible que celle des haut-parleurs avant, et augmenter de façon excessive le niveau des fréquences élevées conduit inévitablement à un effet de distorsion.

Ensuite, nous avons constaté qu'il était souhaitable de diminuer le niveau global des haut-parleurs arrière, sous peine de détruire l'image sonore générale ; en effet, la présence de ces haut-parleurs, une fois la correction implantée, devient très importante. Mais dans ce cas, la correction du champ diffus n'est plus exacte.

Ce qui nous amène à penser que, comme nous avons affecté un rôle particulier aux haut-parleurs avant et arrière, il faut maintenir une cohérence dans cette démarche. En particulier, le rôle des haut-parleurs arrière comme haut-parleurs d'effet « surround » se doit d'être confirmé en leur ajoutant un retard et une certaine décorrélation.

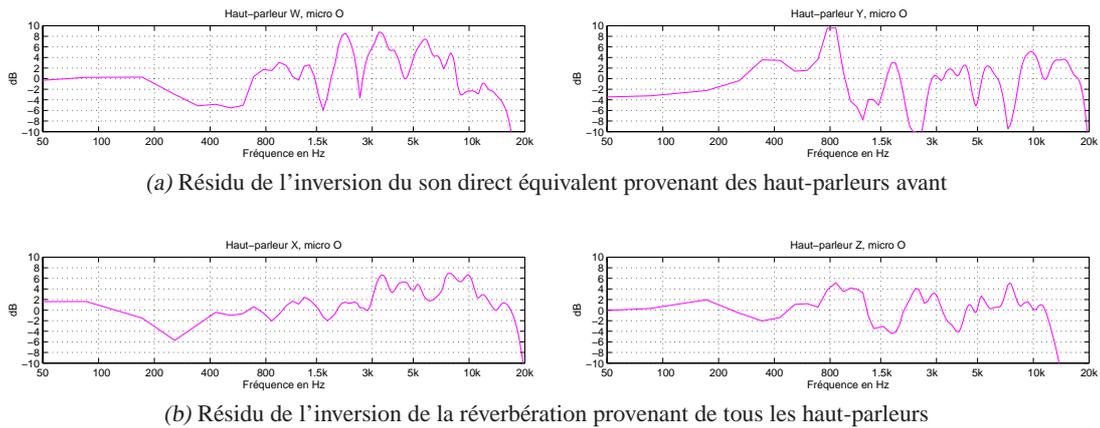


FIGURE III.18 – Fonction de transfert calculée lorsque les filtres de corrections optimisés pour le conducteur sont appliqués aux fonctions de transfert mesurées pour le passager avant

III.2.9.2 Influence d'une correction optimisée pour un autre passager

Il convient enfin de se demander quelle est l'influence sur les autres passagers de la correction calculée pour le conducteur. On s'aperçoit que la position du microphone de mesure est très sensible, c.-à-d. qu'un déplacement d'une dizaine de centimètres est largement suffisant pour modifier radicalement le spectre énergétique total. Cette modification vient principalement de l'influence sur les premières millisecondes de la propagation (que nous appelons son direct équivalent), la réverbération étant moins touchée ; de plus, les basses fréquences sont plus affectées que les hautes fréquences, car le taux de recouvrement modal est assez faible en basses fréquences (cf. § I.4.3). En ce qui concerne la réverbération, c'est plutôt le contraire : les basses fréquences ont globalement la même allure, tandis que les hautes fréquences (supérieures à la fréquence de Schroeder) conservent leur caractère aléatoire ; ceci est conforme au modèle statistique développé au § I.4.2.2, et a été vérifié par Baskind [13].

Nous avons donc utilisé la correction précédente, optimisée pour le conducteur, et nous avons calculé son impact pour un auditeur à la position du passager avant (cf. fig. III.18). Il apparaît que le contenu spectral est loin d'être celui recherché ; à l'audition, les effets peuvent être assez néfastes. Il ne nous semble pas opportun d'utiliser des filtres de corrections en fréquence autres que ceux optimisés pour une place déterminée. On peut envisager le calcul d'une « fonction de transfert moyenne » à partir de celles calculées pour un ensemble de places, et tenter d'optimiser les filtres inverses pour qu'ils compensent les défauts les plus importants mesurés pour cet ensemble de places. Mais il semble plus efficace de mettre en œuvre des techniques d'optimisation telles que celles décrites au § III.3.5.3. L'idée consiste à utiliser un algorithme adaptatif d'égalisation multipoints avec pondérations, dans lequel l'importance de l'erreur à la position d'écoute primordiale est accentuée ; la dégradation de l'égalisation aux autres points de mesure peut être maintenue dans des limites acceptables. Une application de ce principe à l'égalisation en basses fréquences est donnée dans [69, 246].

Cependant, d'après les informations en notre possession, la présence d'une seule personne à bord d'un véhicule représente près de 90 % de son utilisation. Dans ces conditions, la mise en œuvre de tels algorithmes à la fois complexes et coûteux en calcul ne semble pas une nécessité absolue.

III.3 Correction de l'image sonore

III.3.1 Cadre de l'étude

L'amélioration apportée à l'écoute en véhicule que nous avons développée dans la section précédente était une correction fréquentielle : elle ne concernait que la modification du timbre. Elle a été implantée

sous forme de filtres à phase minimale ; de plus, pour avoir un coût de calcul le plus faible possible, nous avons choisi une structure réursive.

Il convient également de chercher à corriger l'image sonore. Nous distinguons dans cette partie la spatialisation délivrée par le signal radio lui-même de celle liée aux conditions de reproduction.

La première correspond aux effets stéréophoniques de temps et d'intensité qui permettent de localiser latéralement et en profondeur une source sonore entre deux haut-parleurs placés à $\pm 30^\circ$ par rapport à l'auditeur. Ces effets sont soit inhérents au système de prise de son, soit ajoutés lors de l'étape de post-production [198, 274] ; ils peuvent être variés suivant le style de musique, mais, dans le cas d'un signal de parole, il est toujours fait en sorte que la source soit perçue de manière frontale.

La seconde correspond aux effets induits par la position des haut-parleurs par rapport à l'auditeur ; or justement, dans un habitacle automobile, le champ sonore est gravement affecté par la position très particulière du système électroacoustique par rapport à l'auditeur, géométrie pour le moins éloignée du triangle équilatéral de la stéréophonie classique. Les informations stéréophoniques contenues dans le signal radio sont donc dégradées, et il convient de chercher à les restaurer.

Dans un habitacle de véhicule, la présence de surfaces réfléchissantes telles que les vitres fait émerger de fortes réflexions, conduisant à détériorer encore un peu plus l'image sonore. Pour ces raisons, une égalisation complète du système audio d'un véhicule ne comprend pas seulement la mise à plat des réponses en fréquence, mais aussi des modifications temporelles. Celles-ci sont nécessaires pour modifier la direction *apparente* de provenance des sources sonores.

Pour localiser une source sonore, le système auditif humain est sensible à la fois aux différences de temps d'arrivée des ondes sonores (en basses fréquences), aux différences de niveaux (en moyennes fréquences), et aux différences spectrales (en hautes fréquences) [54, chap. 1] [20].

Une correction temporelle, associée à une correction de niveau, permet bien entendu d'aligner les temps d'arrivée et les intensités des ondes sonores provenant des sources localisées à des distances différentes (même si cela n'est théoriquement valable qu'en champ libre). Mais des filtres numériques peuvent également être employés pour réaliser un *déplacement virtuel* des sources sonores physiques (c.-à-d. les haut-parleurs du véhicule), et donner l'impression à un auditeur d'écouter une paire de haut-parleurs placés dans une position symétrique par rapport au plan médian de sa tête, restaurant ainsi le triangle équilatéral parfait de l'écoute stéréophonique (cf. Nelson [233, 234, 236]).

L'objectif de cette étude est cependant plus large, puisque nous avons pour ambition de substituer à l'effet d'habitacle un effet de salle nettement plus approprié à l'écoute de la musique, par exemple un bon studio d'écoute voire une salle de concert ; ce qui veut dire que des réflexions discrètes et une réverbération doivent être recréées, pour être en accord avec l'impression spatiale recherchée. Si la réverbération désirée est plus importante que l'effet d'habitacle, il n'est théoriquement pas nécessaire d'annuler l'effet d'habitacle, qui sera masqué par la réverbération créée ; mais si les conditions d'écoute désirées sont plus proches de celle d'un studio d'écoute quasiment anéchoïque, ne pas annuler l'effet d'habitacle peut être à l'origine d'une perception subjective très différente de l'effet recherché, principalement en régime transitoire.

À la différence de Farina [78, 79, 82], nous avons choisi de séparer d'une part l'étape de modification virtuelle de la position des sources sonores, et d'autre part l'étape d'ajout de réverbération ; de cette manière, nous pouvons garder un contrôle sur chacun de ces processus, c.-à-d. l'inversion de la matrice de transfert électroacoustique de l'habitacle qui comprend l'annulation de la réverbération, et le réglage perceptif de la réverbération qui sera effectué grâce au Spatialisateur® (© IRCAM) (cf. § E.1.1). Mais comme Farina, et à la différence de Nelson, nous gardons bien séparées les modifications temporelles et fréquentielles, ce qui permet, entre autres, de minimiser la longueur des filtres de correction.

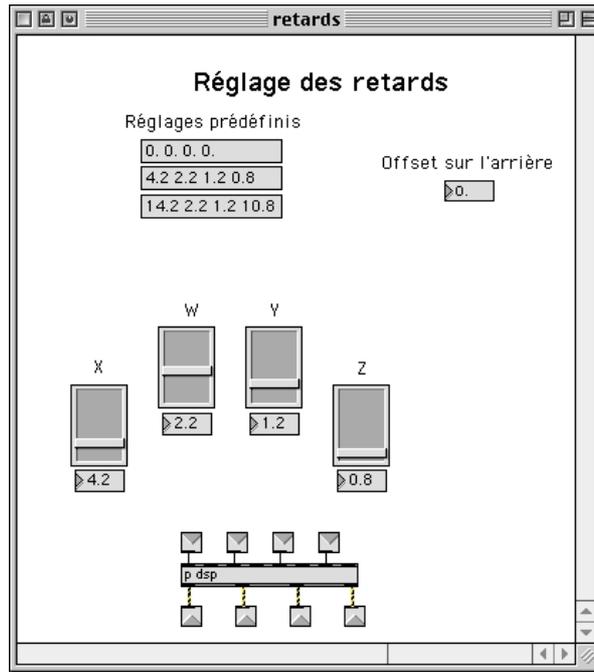


FIGURE III.19 – Interface de traitements en temps-réel réalisée pour optimiser le réglage des valeurs des retards à appliquer aux signaux alimentant les différents haut-parleurs

III.3.2 Alignement des sources sonores existantes

Sans utiliser le moindre filtre, il est possible d'améliorer l'image sonore perçue. L'ajout de retards *et* de gains sur chacun des signaux provenant aux haut-parleurs permet de créer une image parfaitement frontale pour le conducteur (face à l'auditeur et à hauteur de la tête). Le principe consiste à retarder les signaux sonores provenant des haut-parleurs les plus proches, et à leur appliquer une atténuation. Les valeurs des retards et des gains peuvent être calculées à l'aide du temps d'arrivée mesuré sur les réponses impulsionnelles acoustiques entre chacun des haut-parleurs et un microphone omnidirectionnel placé approximativement au centre de la tête d'un auditeur. Mais en tout état de cause, ces valeurs requièrent un ajustement délicat qui ne peut être que manuel. D'où l'utilité de disposer d'une plateforme de traitements fonctionnant en temps réel, et pouvant être facilement connectée au système audio d'un véhicule. La modification de l'image engendrée par une variation de ces retards de l'ordre de quelques dixièmes de millisecondes est très importante, principalement en ce qui concerne le retard appliqué au haut-parleur le plus proche du conducteur.

Nous avons créé une interface de réglage dans le logiciel Max/MSP^{®2} fonctionnant sous MacOS. La figure III.19 présente l'interface pour le réglage des valeurs de retards ; une interface similaire existe pour le réglage des valeurs de gains. Il est possible de choisir une configuration, et d'optimiser le réglage de manière manuelle pour chaque signal sonore ; de plus, il est possible de rajouter un offset global pour les haut-parleurs arrière. Le traitement des échantillons sonores est bien entendu effectué avant la diffusion par les haut-parleurs de la voiture. L'offset général sur les haut-parleurs arrière est important dans la mesure où il permet de retarder l'émission des haut-parleurs arrière par rapport aux haut-parleurs avant, ce qui tend à minimiser les phénomènes de localisations parasites vers l'arrière ; de plus, cela contribue déjà à créer une sensation d'enveloppement sonore (cf. les pondérations spatiales des réflexions au § III.4).

Un test informel a été effectué avec plusieurs dizaines d'auditeurs, avec un signal de parole provenant d'une diffusion radiophonique. Comme dans la section précédente, il était possible de commuter instantanément entre le système audio d'origine et celui modifié ; les sujets n'ont montré aucune hésita-

2. <http://www.cycling74.com/products/maxmsp.html>

tion à accorder une préférence supérieure au système modifié, même avec un signal de musique. Mais si cette technique permet bien de simuler des haut-parleurs situés à égale distance de l'auditeur, elle ne permet pas de placer avec précision une source sonore virtuelle dans l'habitacle.

Nous avons donc poussé plus loin le concept de recentrage de l'image sonore, et nous avons utilisé la technique du PanPot binaural-transaural dans le but de synthétiser deux sources virtuelles à $\pm 30^\circ$, ce qui permet de recréer les conditions d'écoute idéales de la stéréophonie classique.

Nous avons d'abord appliqué cette technique dans le cas simple d'une reproduction avec un système électroacoustique non centré situé en chambre anéchoïque, c.-à-d. sans effet de salle. Nous verrons que l'absence de symétrie du système de reproduction nécessite de mener un calcul d'inversion complète (module et phase) de réponses impulsionnelles binaurales ; or ce problème n'admet pas de solution mathématique exacte, car de tels filtres inverses sont par définition instables (ou stables, mais infinis et non causaux) ! Nous étudierons principalement la technique d'optimisation au sens des moindres carrés, et nous nous intéresserons à différentes techniques de régularisation : régularisation par limitation de la dynamique du module, par filtrage, par troncature ou par amortissement de la décomposition en valeurs singulières des filtres directs.

Puis nous avons appliqué ce principe dans le cas d'une restitution dans un habitacle automobile. Cela pose une difficulté supplémentaire, car la technique du PanPot binaural-transaural est théoriquement étudiée pour être écoutée en chambre anéchoïque. Et même si la réverbération naturelle d'un habitacle automobile peut être considérée comme faible, elle n'est pas complètement absente, et la présence de premières réflexions extrêmement prononcées nuit considérablement à la qualité de la restitution. Il faut donc inverser (annuler) l'acoustique de l'habitacle, ce qui cause des contraintes supplémentaires dans le calcul des filtres transauraux.

III.3.3 Principe de création de sources sonores virtuelles sur haut-parleurs

La synthèse binaurale (cf. chapitre 2), ou simulation de la localisation de sources sonores en trois dimensions en utilisant les propriétés fréquentielles du codage directionnel de l'oreille humaine, est théoriquement conçue pour être écoutée au travers d'un casque d'écoute [206]. Il est théoriquement possible d'obtenir une reproduction sonore tridimensionnelle sur deux haut-parleurs, à condition que le signal binaural soit « decodé » au travers d'une matrice de filtrage inverse (cf. fig. III.20). Cette technique, dite *transaurale*, a pour effet d'annuler l'onde acoustique transmise de chaque haut-parleur vers l'oreille opposée ; on parle d'élimination de la diaphonie transaurale, traduction de *crossstalk cancellation* [15, 46, 53, 106, 204, 232]. En basses fréquences néanmoins, ce processus n'est efficace que si chaque filtre de la matrice inverse a la capacité de fournir une amplification suffisante ; en effet, la différence entre la fonction de transfert d'un haut-parleur vers l'une et l'autre oreille est très faible en basses fréquences, et la matrice de transfert à inverser est par conséquent presque singulière. Ce problème est particulièrement important lorsque les deux haut-parleurs sont proches l'un de l'autre, comme c'est le cas pour le dipôle stéréophonique [160].

Ainsi, si on accepte une contrainte sur la position et l'orientation de la tête de l'auditeur, il est théoriquement possible de reproduire, sur un dispositif stéréophonique conventionnel, des sons latéraux, arrières, en élévation, ou encore diffus (réverbération). En pratique, lorsqu'il est possible de respecter les contraintes de positionnement, la reproduction en mode transaural permet de restituer, au moyen de deux haut-parleurs frontaux, une localisation robuste de la source sonore virtuelle dans un secteur angulaire de $\pm 120^\circ$ environ dans le plan horizontal.

Nous rappellerons dans un premier temps le principe général du décodage transaural pour un signal encodé en binaural, et restitué en chambre anéchoïque ; nous verrons qu'une simplification importante peut être réalisée si le dispositif de restitution est symétrique par rapport au plan médian de la tête de l'auditeur. Puis nous étendrons ce principe au cas d'un dispositif de restitution non symétrique, plus proche de la situation existant dans un habitacle automobile.

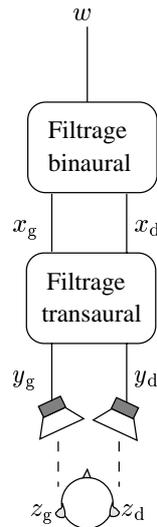


FIGURE III.20 – Encodage binaural d'une source sonore monophonique, et décodage transaural avant sa diffusion sur deux haut-parleurs

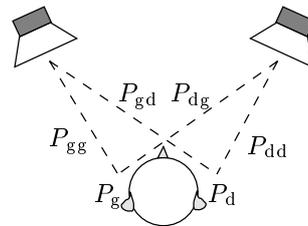


FIGURE III.21 – Chemins croisés pour une écoute sur haut-parleurs

III.3.3.1 Principe des système transauraux

Rappelons d'abord le schéma de perception de deux sources sonores dans l'espace, qui sont constituées dans notre cas par deux haut-parleurs, l'auditeur étant placé sur un des sommets du triangle équilatéral (cf. fig. III.21) Chaque oreille de l'auditeur ne reçoit pas seulement le son du haut-parleur situé dans sa direction, mais aussi celui du haut-parleur situé à l'opposé. La pression reçue par chaque oreille s'écrit

$$\begin{cases} P_g &= P_{gg} + P_{dg} \\ P_d &= P_{gd} + P_{dd} \end{cases} \quad (\text{III.3})$$

Ces termes supplémentaires P_{dg} et P_{gd} ont pour conséquence l'inadaptation du système traditionnel à l'écoute binaurale sur haut-parleur. L'objectif est donc d'éliminer ces termes supplémentaires.

La plupart des équations exprimées dans cette section se rapportent aux systèmes linéaires, et aux fonctions de transfert qui relient les signaux d'entrée et de sortie. On suppose que la structure algébrique dans laquelle nous nous situons est un espace vectoriel réel de dimension finie. Pour faire en sorte que l'opération de convolution puisse être écrite comme une multiplication, ces équations sont exprimées dans le domaine fréquentiel, et, par souci de simplicité, la dépendance fréquentielle des variables a été omise. Les signaux scalaires sont notés en minuscules, et les fonctions de transfert en majuscules ; les vecteurs et les matrices sont notés en gras, les vecteurs en minuscules et les matrices en majuscules, conformément à la tradition en vigueur. Nous adoptons les notations suivantes :

x	:	signal à l'entrée du filtre transaural, après encodage binaural
y	:	signal émis par le haut-parleur
z	:	signal à l'entrée de l'oreille
indice g	:	signal gauche
indice d	:	signal droit
indice gg	:	signal allant du haut-parleur gauche à l'oreille gauche
indice gd	:	signal allant du haut-parleur gauche à l'oreille droite
indice dg	:	signal allant du haut-parleur droit à l'oreille gauche
indice dd	:	signal allant du haut-parleur droit à l'oreille droite

En sommant les contributions de chaque source au niveau de chaque oreille, on obtient

$$\begin{cases} z_g &= G_{gg}y_g + G_{dg}y_d \\ z_d &= G_{gd}y_g + G_{dd}y_d \end{cases}, \quad (\text{III.4})$$

avec G_{xy} les fonctions de transfert entre un haut-parleur et une oreille ; matriciellement, on écrit

$$\mathbf{z} = \mathbf{G}\mathbf{y}, \quad (\text{III.5})$$

avec

$$\mathbf{z} = \begin{pmatrix} z_g \\ z_d \end{pmatrix}, \quad \mathbf{G} = \begin{pmatrix} G_{gg} & G_{dg} \\ G_{gd} & G_{dd} \end{pmatrix}, \quad \mathbf{y} = \begin{pmatrix} y_g \\ y_d \end{pmatrix};$$

pour l'instant, nous n'avons toujours pas introduit le filtrage transaural (on a $y = x$).

Les signaux z_g et z_d sont considérés comme étant mesurés par un microphone omnidirectionnel placé à l'entrée du conduit auditif et en conditions « blocked ear », c.-à-d. avec le conduit auditif bouché. Les fonctions G_{xy} incluent la réponse en fréquence du haut-parleur, l'atténuation de l'air, et la fonction de transfert de l'oreille proprement dite. \mathbf{G} peut donc être factorisé

$$\mathbf{G} = \mathbf{H}\mathbf{P}, \quad (\text{III.6})$$

avec

$$\mathbf{H} = \begin{pmatrix} H_{gg} & H_{dg} \\ H_{gd} & H_{dd} \end{pmatrix}, \quad \text{et} \quad \mathbf{P} = \begin{pmatrix} P_g & 0 \\ 0 & P_d \end{pmatrix};$$

\mathbf{H} est la matrice des HRTF, et \mathbf{P} est la matrice de la fonction de transfert du haut-parleur en champ libre : chaque terme P_x peut être décomposé comme le produit de la fonction de transfert du haut-parleur lui-même et de la fonction de transfert de la propagation dans l'air (c.-à-d. du haut-parleur au centre de la tête). On suppose que le haut-parleur affecte identiquement les deux oreilles.

On souhaite cependant que le son perçu corresponde au signal binaural, ce qui se traduit par les égalités

$$\begin{cases} z_g &= x_g \\ z_d &= x_d \end{cases}, \quad (\text{III.7})$$

et nécessite l'introduction d'un étage supplémentaire de filtrage. Matriciellement, l'opération peut s'écrire

$$\mathbf{y} = \mathbf{F}\mathbf{x}, \quad (\text{III.8})$$

et on désire que

$$\mathbf{z} = \mathbf{G}\mathbf{F}\mathbf{x} = \mathbf{x}, \quad (\text{III.9})$$

d'où

$$\mathbf{F} = \mathbf{G}^{-1} = \mathbf{P}^{-1}\mathbf{H}^{-1} = \begin{pmatrix} F_{gg} & F_{dg} \\ F_{gd} & F_{dd} \end{pmatrix}.$$

\mathbf{P}^{-1} associe un filtre inverse à chaque sortie de haut-parleur et s'écrit

$$\mathbf{P}^{-1} = \begin{pmatrix} \frac{1}{P_g} & 0 \\ 0 & \frac{1}{P_d} \end{pmatrix}. \quad (\text{III.10})$$

En pratique, l'égalisation du haut-parleur peut être omise si ceux-ci sont identiques et de bonne qualité ; en revanche, si l'auditeur n'est pas situé à égale distance des deux haut-parleurs, il est nécessaire de retarder et d'atténuer le signal alimentant le haut-parleur le plus proche.

Dans la suite de l'exposé, nous omettrons la matrice \mathbf{P} pour ne considérer que la matrice de HRTF inverses, c.-à-d.

$$\mathbf{G} = \mathbf{H} \quad \text{et} \quad \mathbf{F} = \mathbf{H}^{-1}. \quad (\text{III.11})$$

Grâce à un calcul très simple d'inversion de matrice, on obtient

$$\mathbf{F} = \frac{1}{\Delta} \begin{pmatrix} H_{dd} & -H_{dg} \\ -H_{gd} & H_{gg} \end{pmatrix}, \quad \text{avec } \Delta = H_{gg}H_{dd} - H_{gd}H_{dg}. \quad (\text{III.12})$$

Δ est le déterminant de la matrice \mathbf{H} ; le déterminant inverse est commun à tous les termes, et il détermine la stabilité du filtre inverse. Cependant, comme c'est un facteur commun, il affecte l'égalisation globale, et non l'effet d'inversion des trajets croisés. Lorsque le déterminant est nul à une fréquence, la matrice \mathbf{H} est singulière et ne peut être inversée.

III.3.3.2 Cas d'un dispositif de restitution symétrique

Nous rappelons ici les équations régissant le principe du système transaural, tel qu'il est couramment décrit dans la littérature.

Le filtre de Schroeder Présentons d'abord le filtrage proposé par Atal et Schroeder [9, 283, 284], dans lequel intervient une simplification très importante : on considère le système auditif humain symétrique par rapport au plan médian. Comme on suppose également le dispositif de restitution symétrique, on écrit alors

$$\begin{cases} H_{gg} = H_{dd} = S \\ H_{gd} = H_{dg} = A \end{cases}, \quad (\text{III.13})$$

avec S la fonction de transfert entre un haut-parleur et l'oreille située du même côté (ou oreille ipsilatérale), et A la fonction de transfert entre un haut-parleur et l'oreille située du côté opposé (ou oreille contralatérale). On obtient donc

$$\begin{cases} y_g = \frac{S}{S^2 - A^2} (x_g - \frac{A}{S} x_d) \\ y_d = \frac{S}{S^2 - A^2} (-\frac{A}{S} x_g + x_d) \end{cases}, \quad (\text{III.14})$$

et finalement en posant $C = -S/A$

$$\begin{cases} y_g = \frac{1}{S(1-C^2)} (x_g + C x_d) \\ y_d = \frac{S}{S(1-C^2)} (C x_g + x_d) \end{cases}. \quad (\text{III.15})$$

Ces relations permettent de déterminer le schéma correspondant au système transaural (cf. fig. III.22). Schroeder a montré que le filtre C est causal (on voit que A possède un retard supérieur à celui de S) ; C est réalisable sous forme de filtre RIF, et donc C^2 aussi ; par conséquent, $\frac{1}{(1-C^2)}$ est réalisable en plaçant C^2 dans une boucle de retour. Le filtre $\frac{1}{S}$ n'est pas *a priori* causal, mais comme on a un nombre fini de points pour la réponse impulsionnelle, on retarde S , ce qui n'a pas d'inconvénients puisqu'il intervient de façon identique sur les deux canaux.

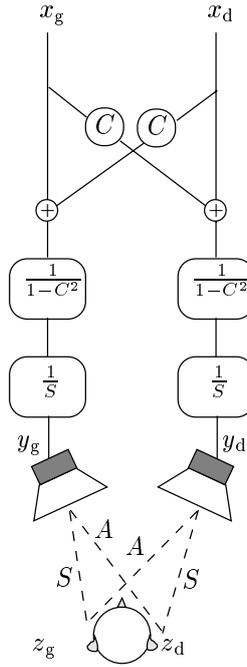


FIGURE III.22 – Schéma du système transaural imaginé par Atal et Schroeder

Le filtre shuffler Présentons ensuite les structures imaginées par Cooper et Bauck [46], et déduites du filtre de Schroeder. Introduisons pour cela

$$\begin{cases} S' = \frac{S}{S^2 - A^2} \\ A' = -\frac{A}{S^2 - A^2} \end{cases}, \quad (\text{III.16})$$

ce qui permet d'écrire

$$\begin{cases} y_g = S'x_g + A'x_d \\ y_d = A'x_g + S'x_d \end{cases}; \quad (\text{III.17})$$

on obtient alors le filtre en treillis équivalent (cf. fig. III.23(a)).

À partir de cette représentation, ils déduisent la structure du filtre *shuffler* en posant

$$\begin{cases} N' = \frac{S' - A'}{2} \\ P' = \frac{S' + A'}{2} \end{cases};$$

en omettant le facteur 2 (qui n'est autre qu'un gain global), ces expressions sont égales à

$$\begin{cases} N' = \frac{1}{N} = \frac{1}{S-A} \\ P' = \frac{1}{P} = \frac{1}{S+A} \end{cases}, \quad (\text{III.18})$$

et l'on peut finalement écrire

$$\begin{cases} y_g = (x_g - x_d)N' + (x_g + x_d)P' \\ y_d = -(x_g - x_d)N' + (x_g + x_d)P' \end{cases}, \quad (\text{III.19})$$

ce qui permet d'en déduire le filtre shuffler équivalent (cf. fig. III.23(b)). On obtient donc une structure

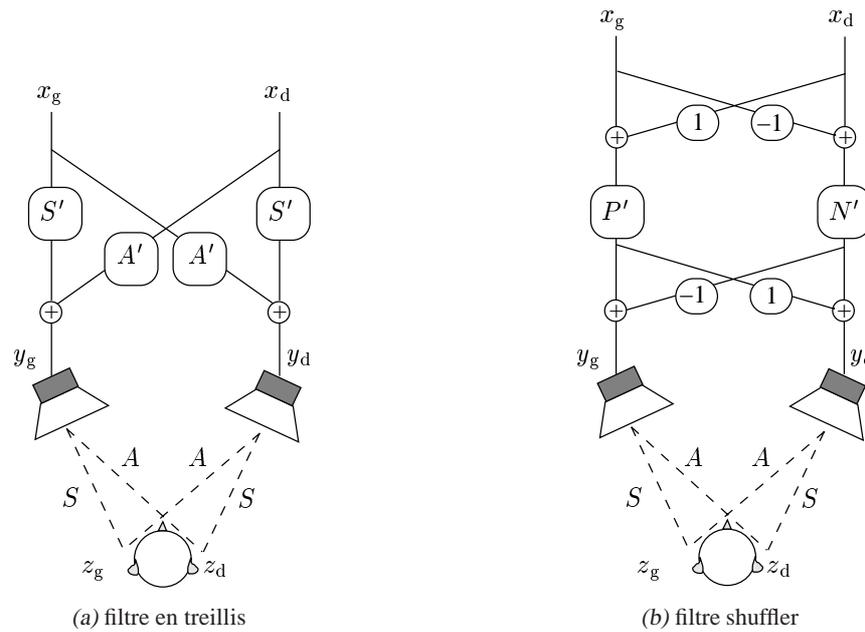


FIGURE III.23 – Schéma du système transaural imaginé par Cooper et Bauck

beaucoup plus simple puisqu'il n'y a plus que deux filtres à synthétiser.

L'appellation de cette structure vient de ce qu'on effectue la somme et la différence des signaux d'entrée binauraux avant d'effectuer le filtrage (on effectue un mélange des signaux) ; la combinaison est effectuée en sens inverse pour obtenir les signaux de sortie alimentant les haut-parleurs. L'opération de mélange des signaux d'entrée est effectuée par la matrice unitaire

$$\mathbf{U} = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix} ; \quad (\text{III.20})$$

on vérifie que $\mathbf{U} = \mathbf{U}^{-1} = \mathbf{U}^t$, que les vecteurs propres de la matrice symétrique \mathbf{H}^{-1} sont les vecteurs $(1; 1)$ et $(1; -1)$, et que les valeurs propres associées sont $\{\frac{1}{S+A}; \frac{1}{S-A}\}$. Donc la matrice de mélange \mathbf{U} diagonalise la matrice symétrique \mathbf{H}^{-1} par la transformation

$$\mathbf{H}^{-1} = \mathbf{U}^{-1} \begin{pmatrix} \frac{1}{S+A} & 0 \\ 0 & \frac{1}{S-A} \end{pmatrix} \mathbf{U}, \quad (\text{III.21})$$

ce qui correspond, toujours au facteur 2 près, à l'équation (III.19).

Cooper et Bauck introduisent également le concept de *phase minimale jointe*. Pour être à phase minimale jointe, tous les filtres d'un même ensemble doivent avoir le même excès de phase (cf. § C.1.3), et cet excès de phase doit pouvoir être assimilé à un retard pur indépendant de la fréquence. Si on retire ce facteur commun à tous les excès de phase, on obtiendra un ensemble de filtres à phase minimale. Donc,

- si l'excès de phase est identique, alors n'inverser que le module ne détériore pas la transauralisation, puisqu'on rajoute un excès de phase commun, mais le système n'est pas transparent ;
- si cet excès de phase est un retard pur, alors le système est neutre à un retard près, identique sur les deux voies.

Ce concept est mis à profit dans la structure shuffler, car on peut vérifier expérimentalement que les filtres N et P ont quasiment le même excès de phase (il s'agit d'une constatation purement heuristique). Une conséquence pratique de ce comportement est que les filtres N et P sont définis uniquement par leur module $|N|$ et $|P|$, la phase minimale étant déduite à l'aide de la transformée de Hilbert (cf. § C.1.3). Le calcul des filtres inverses N' et P' se réduit donc à une inversion de module, opération nettement plus simple. En outre, on peut implanter ces filtres sous forme RII, structure peu coûteuse et efficace (cf. § C.1.7).

III.3.3.3 Cas d'un dispositif de restitution non symétrique

Lorsque la position de l'auditeur n'est pas symétrique par rapport au système électroacoustique à inverser, les principes simplificateurs ne peuvent plus s'appliquer.

Reprenons ici la formulation utilisée par Gardner [91] et introduite par Møller [205] ; en divisant le numérateur et le dénominateur par $H_{gg}H_{dd}$, l'équation (III.12) devient

$$\mathbf{F} = \begin{pmatrix} \frac{1}{H_{gg}} & 0 \\ 0 & \frac{1}{H_{dd}} \end{pmatrix} \begin{pmatrix} 1 & -\text{ITF}_d \\ -\text{ITF}_g & 1 \end{pmatrix} \frac{1}{1 - \text{ITF}_g \text{ITF}_d}, \quad (\text{III.22})$$

avec

$$\text{ITF}_g = \frac{H_{gd}}{H_{gg}} \quad \text{et} \quad \text{ITF}_d = \frac{H_{dg}}{H_{dd}} \quad (\text{III.23})$$

les fonctions de transfert interaurales (ou Interaural Transfer Function). L'examen de l'équation (III.22) permet d'expliquer le processus de l'inversion des trajets croisés. Le travail est en effet réalisé par les termes $-\text{ITF}$ sur la seconde diagonale de la matrice de droite. Ces termes effectuent en quelque sorte une « prédiction » des trajets croisés, et envoient une composante d'annulation hors phase dans le canal opposé. Par exemple, la voie droite du signal d'entrée est convoluée avec ITF_d qui prédit le trajet croisé atteignant l'oreille gauche, et le résultat est soustrait du signal de sortie gauche. Le terme commun $1 - \text{ITF}_g \text{ITF}_d$ effectue quant à lui une compensation des trajets croisés d'ordre supérieurs : le fait que chaque composante d'annulation des trajets croisés transite elle-même par l'oreille opposée doit être annulé ; ce terme est surtout significatif en basses fréquences. La matrice de gauche est diagonale ; il s'agit essentiellement d'une égalisation en hautes fréquences qui facilite la perception de sources latérales ou arrières. L'utilisation des ITF pour prédire les trajets croisés vers l'oreille contralatérale nécessite que chaque signal de sortie soit égalisé par la HRTF ipsilatérale.

La matrice \mathbf{H} est inversible si et seulement si elle est non singulière, c.-à-d. si son déterminant Δ est non nul. Comme \mathbf{H} est fonction de la fréquence, il est possible que la matrice inverse \mathbf{H}^{-1} n'existe seulement que pour certaines bandes de fréquences où la matrice \mathbf{H} est non singulière. En d'autres termes, si la matrice \mathbf{H} est mal conditionnée à certaines fréquences, cela conduira à une faible valeur de Δ , et le module de $1/\Delta$ sera très grand. En pratique, il faut faire appel à des méthodes de régularisation pour limiter le module de $1/\Delta$, et dans ces bandes de fréquences, la matrice inverse alors calculée ne sera qu'une approximation de la véritable matrice inverse.

On peut s'interroger sur la causalité et la stabilité d'un tel système [91]. Le terme à prendre en compte est $1/(1 - \text{ITF}_g \text{ITF}_d)$.

Pour que le système soit causal, il faut que la somme des deux retards interauraux soit supérieure à zéro, c.-à-d.

$$\text{ITD}_g + \text{ITD}_d > 0, \quad (\text{III.24})$$

avec ITD_x le retard interaural correspondant. Cette condition est réalisée (pour un modèle de tête sphérique) lorsque l'angle θ de la tête avec le plan médian est tel que $-90 < \theta < 90$.

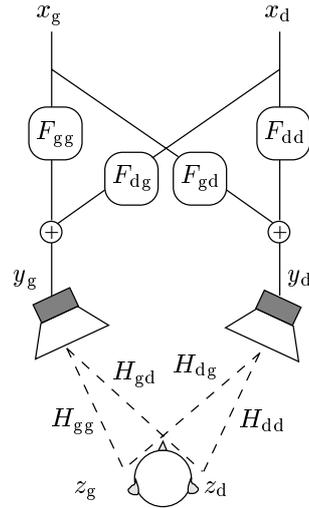


FIGURE III.24 – Schéma général du système transaural pour un dispositif de restitution quelconque

Pour que le système soit stable, il faut que les pôles de la fonction de transfert soient à l'intérieur du cercle unité, ce qui se traduit par l'inéquation

$$|\text{ITF}_g(e^{2j\pi f})| |\text{ITF}_d(e^{2j\pi f})| < 1 \quad \forall f. \quad (\text{III.25})$$

Les ITF rendent compte du masquage de la tête, et peuvent être considérées comme un filtre passe-bas ; la condition énoncée ci-dessus est donc également réalisée lorsque l'angle θ de la tête avec le plan médian est tel que $-90 < \theta < 90$.

Nous allons maintenant présenter trois structures d'implantation des filtres transauraux pour un dispositif de restitution non symétrique.

Structure en treillis Reprenons l'équation (III.9) et l'équation (III.12) ; on peut écrire

$$\begin{cases} y_g = F_{gg}x_g + F_{dg}x_d \\ y_d = F_{gd}x_g + F_{dd}x_d \end{cases}, \quad (\text{III.26})$$

avec

$$\begin{bmatrix} F_{gg} = \frac{H_{dd}}{\Delta} & ; & F_{dg} = -\frac{H_{dg}}{\Delta} \\ F_{gd} = -\frac{H_{gd}}{\Delta} & ; & F_{dd} = \frac{H_{gg}}{\Delta} \end{bmatrix}.$$

La structure la plus générale correspondant à cette équation est celle représentée figure III.24 ; il s'agit d'une structure en treillis.

Structure récursive La structure récursive, introduite par Iwahara et Mori [132], et modifiée par Gardner [91] pour introduire les fonctions de transfert interaurales, est schématisée figure III.25(a). Les équations la régissant sont

$$\begin{cases} z_g = A(y_g + \frac{C}{D}z_d) \\ z_d = D(y_d + \frac{B}{A}z_g) \end{cases}; \quad (\text{III.27})$$

les coefficients de cette équation sont calculés pour satisfaire l'équation (III.12), ce qui donne

$$\begin{bmatrix} A = \frac{1}{H_{gg}} & ; & B = -\frac{H_{gd}}{H_{gg}} \\ D = \frac{1}{H_{dd}} & ; & C = -\frac{H_{dg}}{H_{dd}} \end{bmatrix}.$$

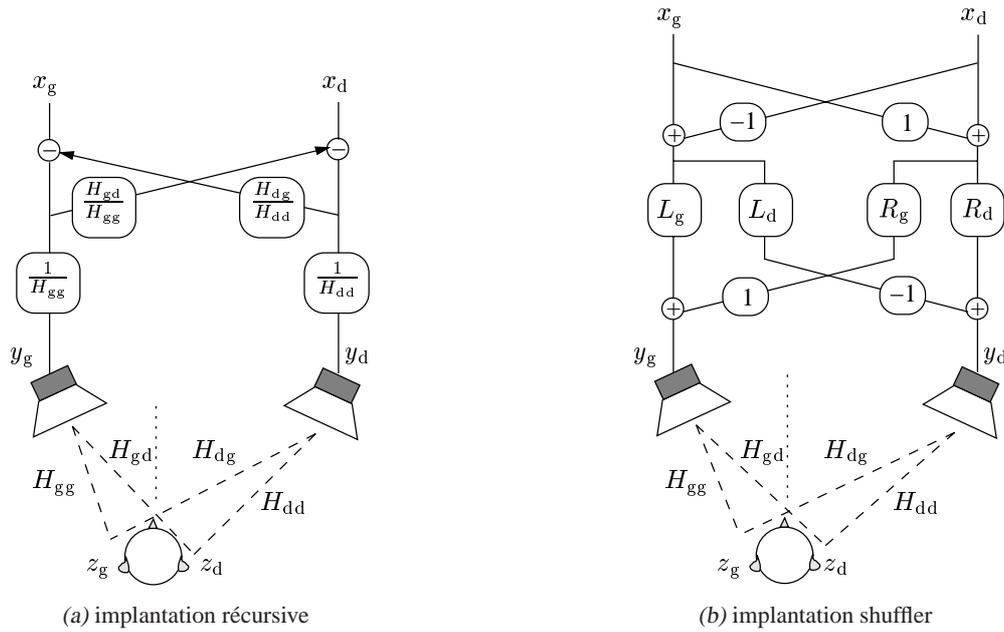


FIGURE III.25 – Schémas particuliers du système transaural pour un dispositif de restitution quelconque

L'interprétation de cette implantation est assez intuitive : chaque canal est convolué par la fonction de transfert interaurale appropriée, effectuant ainsi une prédiction du trajet croisé atteignant l'oreille contralatérale ; le trajet croisé est annulé en injectant l'opposé de ce signal prédit dans le canal opposé. Une caractéristique importante de cette implantation est que le signal d'annulation est injecté à l'entrée du canal opposé, plutôt qu'à la sortie, donc les trajets croisés d'ordre supérieurs sont automatiquement annulés. Finalement, chaque canal est égalisé par la fonction de transfert ipsilatérale correspondante.

Structure shuffler On peut aussi imaginer une structure shuffler. Ecrivons pour cela

$$\begin{cases} y_g = Ax_g - Bx_d = \frac{(A+B)}{2}(x_g - x_d) + \frac{(A-B)}{2}(x_g + x_d) \\ y_d = -Cx_g + Dx_d = -\frac{(C+D)}{2}(x_g - x_d) + \frac{(-C+D)}{2}(x_g + x_d) \end{cases}, \quad (\text{III.28})$$

avec

$$\begin{bmatrix} A = \frac{H_{dd}}{\Delta} & ; & B = \frac{H_{dg}}{\Delta} \\ C = \frac{H_{gd}}{\Delta} & ; & D = \frac{H_{gg}}{\Delta} \end{bmatrix};$$

on obtient donc, au facteur 2 près (qui n'est autre qu'un gain global)

$$\begin{bmatrix} A + B = \frac{H_{dd} + H_{dg}}{\Delta} = L_g & ; & A - B = \frac{H_{dd} - H_{dg}}{\Delta} = R_g \\ C + D = \frac{H_{gd} + H_{gg}}{\Delta} = L_d & ; & -C + D = \frac{-H_{gd} + H_{gg}}{\Delta} = R_d \end{bmatrix}, \quad (\text{III.29})$$

ce qui nous permet de dessiner le schéma d'une structure shuffler (cf. fig. III.25(b)). Comme pour l'équation (III.21), on peut écrire matriciellement

$$\mathbf{H}^{-1} = \mathbf{U}^{-1} \begin{pmatrix} R_g & L_g \\ R_d & L_d \end{pmatrix} \mathbf{U}. \quad (\text{III.30})$$

Le concept de phase minimale jointe, introduit par Cooper et Bauck dans le cas d'une structure shuffler pour un dispositif de restitution symétrique, ne peut plus s'appliquer ici. Il faut procéder à une

inversion à phase totale du déterminant Δ .

Notons que si le dispositif de restitution est symétrique, alors

$$\begin{bmatrix} H_{gg} & = & H_{dd} & = & S \\ H_{gd} & = & H_{dg} & = & A \end{bmatrix},$$

d'où

$$\begin{bmatrix} \Delta & = & S^2 - A^2 \\ L_g = L_d & = & N' \\ R_g = R_d & = & P' \end{bmatrix}, \quad (\text{III.31})$$

et l'équation (III.28) redevient l'équation (III.19).

III.3.3.4 Limitation de la largeur de bande

L'annulation des trajets croisés ne peut être effectuée sur toute la largeur de la bande audio, c.-à-d. de 20 Hz à 20 kHz. Plusieurs problèmes existent en hautes fréquences :

- le gain des fonctions de transfert interaurales ITF peut devenir supérieur à 1 à cause d'anti-résonances dans la fonction de transfert ipsilatérale, causant une instabilité des filtres transauraux (cf. § III.3.3.3) [91] ;
- plus la longueur d'onde diminue, plus le système transaural devient sensible à la position de la tête de l'auditeur ;
- l'inversion d'une fonction de transfert est toujours délicate à réaliser en hautes fréquences, et il y a fort à parier que le résultat soit finalement plus catastrophique que si on effectue aucun traitement.

Le calcul des filtres est donc mené jusque 6 kHz, ce qui correspond à une longueur d'onde de 50 mm environ. Cette limite a été vérifiée expérimentalement comme largement suffisante à la réalisation d'un système transaural convaincant.

Modèle général La méthode la plus efficace, proposée par Cooper et Bauck [47], consiste à spécifier différemment la matrice \mathbf{F} en basses fréquences et en hautes fréquences. On peut donc la récrire globalement

$$\mathbf{F} = \begin{pmatrix} H_{gg} & H_{PB}H_{dg} \\ H_{PB}H_{gd} & H_{dd} \end{pmatrix}^{-1}, \quad (\text{III.32})$$

avec H_{PB} la fonction de transfert d'un filtre passe-bas ; cela revient à remplacer, dans l'équation (III.22), les termes ITF par leur version filtrée par le filtre passe-bas. En basses fréquences, on a toujours $\mathbf{F} = \mathbf{H}^{-1}$, mais en hautes fréquences, on a maintenant

$$\mathbf{F}_{f>f_c} = \begin{pmatrix} \frac{1}{H_{gg}} & 0 \\ 0 & \frac{1}{H_{dd}} \end{pmatrix}, \quad (\text{III.33})$$

ce qui correspond à une simple égalisation ipsilatérale. Cette formulation est importante lorsqu'on désire réaliser un système qui a les bonnes propriétés de répartition de la puissance. Lorsque la direction de provenance de la source virtuelle est celle d'un haut-parleur, le signal provenant de ce haut-parleur est uniforme en fréquence ; mais l'autre haut-parleur émet une certaine puissance en hautes fréquences, puisque le système d'annulation des trajets croisés n'y est pas effectif. Nous décrivons plus tard une méthode pour rétablir les bonnes propriétés de répartition de la puissance.

Lorsque le dispositif de restitution est symétrique, la matrice \mathbf{F} devient, pour une implantation shuffler

$$\mathbf{F}_{f>f_c} = \frac{1}{S} \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \quad (\text{III.34})$$

avec $S = H_{gg} = H_{dd}$. C'est ce principe qui est implanté dans le Spatialisateur[®], à la différence que la fonction de transfert de normalisation n'est pas directement la fonction de transfert ipsilatérale, mais une moyenne quadratique des fonctions de transfert ipsi et contralatérales, c.-à-d. $\sqrt{S^2 + A^2}$; ce choix permet une transition plus homogène entre le domaine des basses et des hautes fréquences, et a été confirmé par des tests d'écoute.

Lorsque le dispositif de restitution n'est pas symétrique, la matrice \mathbf{F} devient, pour une implantation shuffler

$$\mathbf{F}_{f>f_c} = \begin{pmatrix} G & 0 \\ 0 & D \end{pmatrix}, \quad (\text{III.35})$$

avec

$$\begin{cases} G = \sqrt{\frac{1}{2}(L_g^2 + R_g^2)} \\ D = \sqrt{\frac{1}{2}(L_d^2 + R_d^2)} \end{cases}.$$

Modèle amélioré de transfert de puissance Pour une source w , les signaux alimentant les haut-parleurs sont donnés par

$$\mathbf{y} = \mathbf{F}\mathbf{h}w, \quad (\text{III.36})$$

avec \mathbf{h} les HRTF correspondant à la position de la source. En hautes fréquences, d'après l'équation (III.33), l'équation précédente devient

$$\mathbf{y}_{f>f_c} = \begin{pmatrix} H_g/H_{gg} \\ H_d/H_{dd} \end{pmatrix} w. \quad (\text{III.37})$$

Il s'agit d'une égalisation des HRTF en champ libre; et globalement, les haut-parleurs émettent des signaux binauraux. Lorsque la direction de provenance de la source est celle d'un haut-parleur, le signal provenant de ce haut-parleur est uniforme en fréquence, grâce à l'égalisation, mais l'autre haut-parleur émet la réponse binaurale contralatérale égalisée par rapport au champ libre, ce qui viole les propriétés de répartition de la puissance. Si l'annulation des trajets croisés s'étendait jusqu'aux plus hautes fréquences, la réponse contralatérale serait annulée et n'apparaîtrait pas sur l'autre haut-parleur; cependant, comme nous l'avons mentionné au début de ce paragraphe, ceci n'est pas opportun. Il convient donc de rechercher une technique permettant d'optimiser l'annulation des trajets croisés en hautes fréquences, tout en préservant les propriétés de répartition de la puissance.

Cette solution est décrite par Gardner [91]. Le déphasage des signaux parvenant aux oreilles en hautes fréquences étant important, on considère que ces signaux s'additionnent de manière incohérente, c.-à-d. en énergie. Les fonctions de transfert de puissance sont dérivées des mesures de HRTF. La matrice de transfert de puissance est inversée dans le but de calculer les puissances à injecter dans les haut-parleurs pour obtenir les bonnes puissances au niveau des oreilles. Souvent, il n'est pas possible d'obtenir les bonnes valeurs de puissances au niveau des oreilles, par exemple si on veut synthétiser une source plus à droite que le haut-parleur droit. En effet, dans ce cas la différence de niveau interaural doit être plus grande que si on envoie le signal uniquement dans le haut-parleur droit; or le signal émis par le haut-parleur gauche réduit la différence de niveau interaural. Dans les cas où il n'existe pas de solution exacte, l'idée de la méthode est d'envoyer le signal dans un seul haut-parleur, et d'effectuer une pondération de sa puissance pour que le transfert de puissance total aux deux oreilles soit égal à la puissance totale des HRTF considérées. Pratiquement, la puissance émise par chaque haut-parleur en hautes fréquences est contrôlée par un gain associé à chaque canal; et comme le système d'annulation des trajets croisés est diagonal en hautes fréquences, les gains peuvent être associés aux HRTF considérées. Ainsi, l'équation précédente devient

$$\mathbf{y}_{f>f_c} = \begin{pmatrix} g_g H_g/H_{gg} \\ g_d H_d/H_{dd} \end{pmatrix} w, \quad (\text{III.38})$$

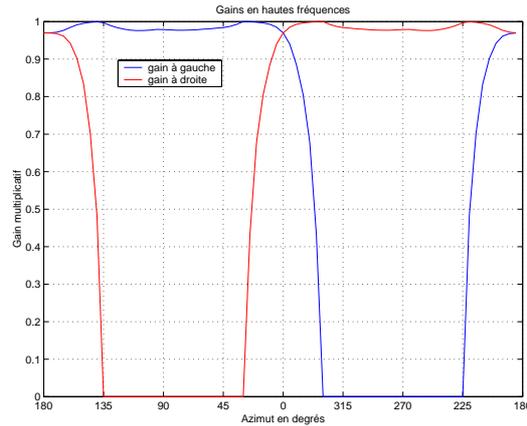


FIGURE III.26 – Gains appliqués en hautes fréquences sur les HRTF pour avoir une répartition plausible de la puissance entre les deux oreilles ; la géométrie du dispositif d'écoute est celle de la stéréophonie classique

et on peut ainsi créer un ensemble de HRTF « modifiées » qui ont leur réponse adaptées en hautes fréquences pour satisfaire les bonnes propriétés de répartition de la puissance entre les deux oreilles, après passage dans un système d'annulation des trajets croisés.

Nous ne détaillerons pas ici le détail du calcul des gains g_g et g_d ; le lecteur se reportera à [91]. Mais une attention particulière doit être apportée à la mise en œuvre du filtre passe-haut utilisé pour calculer la composante hautes fréquences des HRTF. Il est possible d'utiliser deux filtres shelving (un par canal), mais leur phase en basses fréquences doit être identique, et leur module indépendant du gain en hautes fréquences, sous peine d'introduire des effets parasites au niveau des différences interaurales de temps et d'intensité. Gardner conseille donc d'utiliser un filtre FIR à phase nulle. La figure III.26 présente les valeurs de gains en hautes fréquences pour un dispositif d'écoute symétrique classique (haut-parleurs à $\pm 30^\circ$, tête de l'auditeur à 0°). On remarque que lorsque la direction de provenance de la source est celle d'un haut-parleur, la solution est d'envoyer toute la puissance dans ce haut-parleur et d'éteindre l'autre. Lorsque la direction de provenance de la source est supérieure à 30° , il n'y a pas de solution exacte, et le principe consiste à éteindre le haut-parleur contralatéral tandis que le gain du haut-parleur ipsilatéral est choisi pour conserver l'énergie totale.

On peut enfin imaginer une méthode pour calculer les HRTF modifiées ; notons qu'il existe un ensemble de gains pour chaque géométrie du dispositif de restitution, et pour chaque position de la tête de l'auditeur. Il est nécessaire de séparer en deux le domaine fréquentiel, et on pourrait penser utiliser deux filtres shelving, voire un seul si les filtres ont une réponse complémentaire. Malheureusement, il n'est pas possible d'utiliser une paire de filtres shelving récursifs car leur réponse en phase en basses fréquences dépend de leur gain en hautes fréquences, et inversement. Il faut donc absolument utiliser des filtres RIF à phase nulle. Un seul filtre est nécessaire, par exemple un filtre passe-bas pour rester cohérent avec l'équation (III.32), puisqu'on peut écrire

$$\begin{aligned}\hat{w} &= H_{PB} w + g(1 - H_{PB})w \\ &= gw - H_{PB} w(1 - g),\end{aligned}\tag{III.39}$$

ce qui conduit au schéma de la figure III.27.

III.3.4 Validation en champ libre

Nous allons maintenant tester en champ libre les différentes implantations d'un système transaural conçu pour fonctionner sur un dispositif de restitution quelconque.

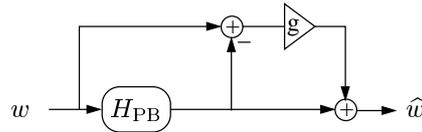


FIGURE III.27 – Synthèse binaurale permettant d'appliquer des gains spécifiques en hautes fréquences

Rappelons qu'il existe une matrice inverse par dispositif de restitution (c.-à-d. pour une position des haut-parleurs et de l'auditeur). De plus, il est *théoriquement* nécessaire que les filtres transauraux soient calculés à partir des fonctions de transfert de l'auditeur ; ceci étant impossible pour des raisons pratiques, on utilise, comme pour la synthèse binaurale, un jeu de HRIR mesurées sur un mannequin acoustique que l'on « adapte » afin qu'elles soient valides pour un grand nombre de sujets [142, 176–178]. Le mannequin utilisé est le KEMAR, dont les fonctions de transfert d'oreille ont été mesurées par Gardner [92]. Les réponses impulsionnelles d'oreille comportent 512 points.

Sauf dans le cas de l'implantation symétrique sous forme de structure shuffler, nous allons devoir effectuer l'inversion du déterminant Δ de la matrice \mathbf{H} , matrice de transfert électroacoustique directe. Δ est la différence de deux produits de fonctions de transfert d'oreille, à savoir $\Delta = H_{gg}H_{dd} - H_{gd}H_{dg}$. Il n'y a aucune raison pour que ce filtre soit à phase minimale. Comme nous devons inverser ce filtre, et qu'il n'existe pas de solution exacte, nous mettons en œuvre une technique d'optimisation au sens des moindres carrés comportant un filtre de régularisation ; le calcul est effectué dans le domaine temporel, afin de s'affranchir du problème de repliement inhérent à la spécification dans le domaine fréquentiel discret.

Nous nous proposons de comparer plusieurs implantations du système transaural, dans une configuration d'écoute symétrique, et dans une configuration d'écoute non symétrique.

Plusieurs critères objectifs ont été utilisés :

- la comparaison des signaux contenant l'information acoustique entre chaque haut-parleur et chaque oreille, par rapport au but recherché, c.-à-d. l'annulation des trajets croisés ;
- la comparaison du vecteur de signaux \mathbf{z} arrivant aux oreilles de l'auditeur, par rapport au signal binaural \mathbf{x} arrivant à l'entrée du système transaural ;
- le retard interaural, ou ITD (Interaural Time Delay), indice pertinent de la localisation auditive en basses fréquences.

D'autre part, nous avons implanté tous ces dispositifs dans le logiciel Max/MSP®, et nous avons effectué une comparaison auditive des résultats.

III.3.4.1 Dispositif de restitution symétrique

Plaçons-nous d'abord dans le cas d'un dispositif de restitution symétrique, c.-à-d. deux haut-parleurs situés à $\pm 30^\circ$ par rapport au plan médian de la tête de l'auditeur.

Nous nous proposons ensuite de comparer :

- l'implantation sous forme classique, conforme à la figure III.24, et mettant en œuvre quatre filtres RIF ;
- l'implantation sous forme shuffler général (avec un dispositif de restitution symétrique), conforme à la figure III.25(b), et sous forme shuffler symétrique, conforme à la figure III.23(b), mettant en œuvre quatre (ou seulement deux) filtres RIF calculés directement d'après le système d'équations correspondantes (III.29) ;
- l'implantation sous forme shuffler général (avec un dispositif de restitution symétrique), conforme à la figure III.25(b), et sous forme shuffler symétrique, conforme à la figure III.23(b), mettant en

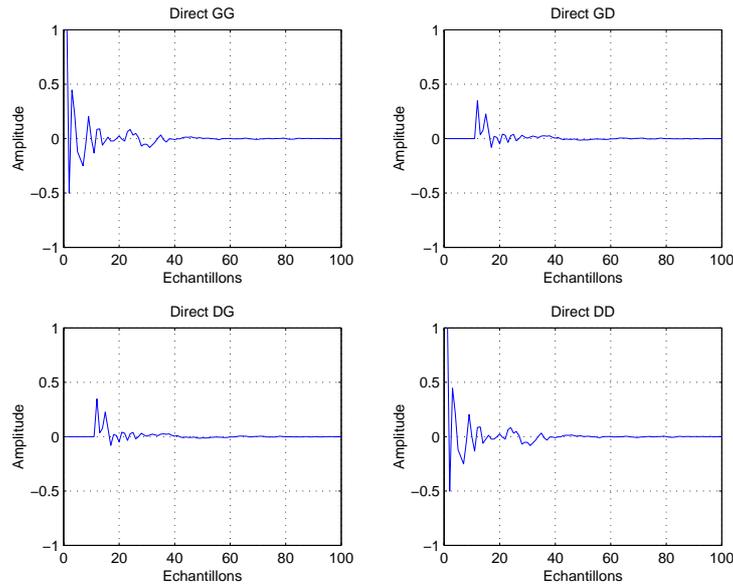


FIGURE III.28 – Fonctions de transfert entre haut-parleurs et oreilles, mesurées en chambre anéchoïque, pour un système de restitution symétrique composé de deux haut-parleurs à $\pm 30^\circ$

œuvre quatre (ou seulement deux) filtres RIF ou RII calculés à partir d'une décomposition en module et excès de phase, ce qui permet de limiter la largeur de bande du système transaural.

Sur les figures suivantes, les couleurs sont réparties comme suit :

- bleu → implantation classique
- bleu plus clair → implantation générale shuffler, avec quatre filtres RIF
- vert → implantation générale shuffler, avec quatre filtres RIF calculés à partir d'une décomposition en module et excès de phase, avec limitation de la largeur de bande du système transaural
- cyan → implantation générale shuffler, avec quatre filtres RII calculés à partir d'une décomposition en module et excès de phase, avec limitation de la largeur de bande du système transaural
- rose → implantation symétrique shuffler, avec deux filtres RIF
- rouge → implantation symétrique shuffler, avec deux filtres RIF calculés à partir d'une décomposition en module et excès de phase, avec limitation de la largeur de bande du système transaural
- magenta → implantation symétrique shuffler, avec deux filtres RII calculés à partir d'une décomposition en module et excès de phase, avec limitation de la largeur de bande du système transaural

Les éléments de la matrice de transfert électroacoustique directe \mathbf{H} sont présentés figure III.28 ; dans un souci de clarté, nous n'avons représenté que les 100 premiers points (sur les 512).

Nous devons effectuer l'inversion du déterminant de cette matrice ; chaque HRIR comportant 512 points, Δ en comporte 1023. Nous choisissons de calculer l'inversion à phase totale du déterminant sur 1024 points (cf. fig III.29). Le bruit résiduel est inférieur à -30 dB, et nous considérons la précision du calcul comme suffisante. Nous avons écouté l'influence de ce filtre sur un signal musical, et il n'est pas apparu de phénomènes désagréables.

Dans le cas d'une implantation shuffler symétrique, cette inversion n'est pas nécessaire, puisqu'on utilise la propriété de phase minimale jointe des filtres $N = S - A$ et $P = S + A$, avec S la fonction de transfert ipsilatérale et A la fonction de transfert contralatérale (cf. § III.3.3.2). Une inversion à phase minimale des filtres N et P est seulement nécessaire.

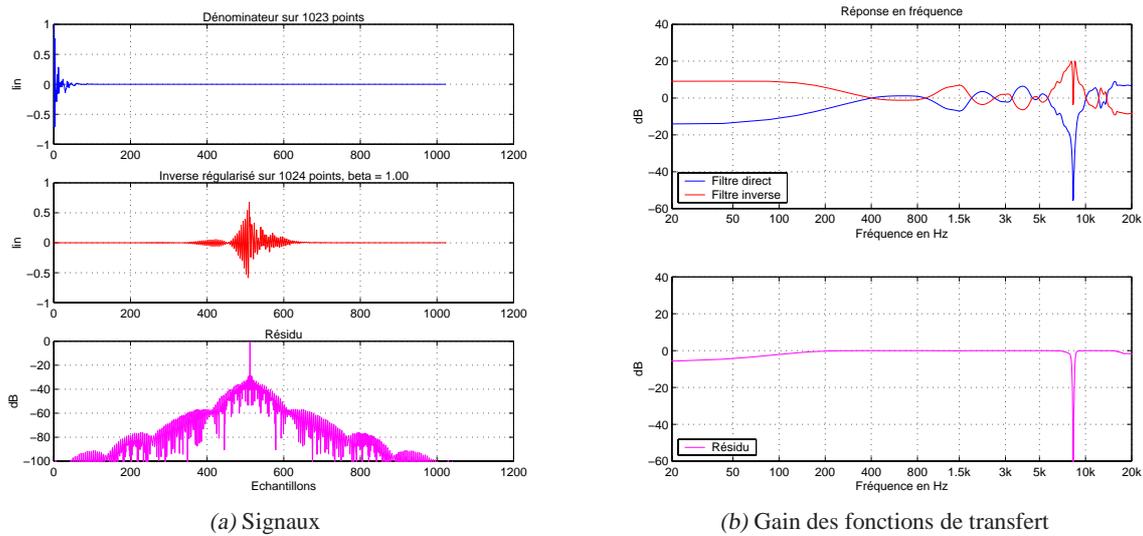


FIGURE III.29 – Inversion du déterminant de la matrice de transfert électroacoustique directe, dans le cas d'une écoute en chambre sourde, sur un dispositif de restitution symétrique

Nous pouvons maintenant procéder à la comparaison des différentes structures d'inversion des trajets croisés.

Annulation des trajets croisés Intéressons-nous aux fonctions de transfert entre haut-parleurs et oreilles, après passage dans les systèmes transauraux. La figure III.30 présente les gains des fonctions de transfert résiduelles entre chaque haut-parleur et chaque oreille, c.-à-d. le produit matriciel \mathbf{HF} qui devrait en toute rigueur être égal à la matrice identité.

On remarque que les trajets directs sont correctement restitués, mis à part une légère faiblesse en basses fréquences pour les courbes bleues ; cette faiblesse est due au filtre de régularisation utilisé pour inverser le déterminant de la matrice de transfert directe.

Pour les trajets croisés, on constate trois ensembles de courbes :

- les courbes bleues, correspondant à la mise en œuvre de quatre filtres RIF de 1535 points, pour lesquelles le taux de réjection est supérieur à 60 dB, mis à part une légère faiblesse en basses fréquences, toujours due à la régularisation ;
- la courbe rose, correspondant à la mise en œuvre de deux filtres RIF de 1535 points, pour laquelle le taux de réjection est variable, mais cependant assez faible en basses fréquences ;
- les autres courbes, correspondant à une technique de calcul de filtres par décomposition en module et excès de phase, et permettant une limitation du système transaural à 6 kHz, pour lesquelles le taux de réjection est le plus faible ; les filtres (deux ou quatre suivant la prise en compte de la symétrie) sont laissés sous forme RIF de 512 points (correspondant à la longueur des HRIR originales) ou modélisés sous forme RII d'ordre 12.

Les deux courbes bleues montrent que le filtrage RIF permet de réaliser un système d'annulation des trajets croisés d'une très grande précision, mais au prix d'un coût de calcul en temps réel plus élevé qu'avec des filtres RII (cf. § C.1.5).

On peut s'interroger sur la courbe rose, car le système correspondant devrait théoriquement présenter le même taux de réjection des trajets croisés que les systèmes correspondant aux courbes bleues ; nous n'avons pu expliquer ce phénomène.

Les quatre courbes présentant le taux de réjection le plus faible sont parfaitement groupées. Ceci montre que l'implantation shuffler symétrique et shuffler non symétrique sont ici équivalentes (en sup-

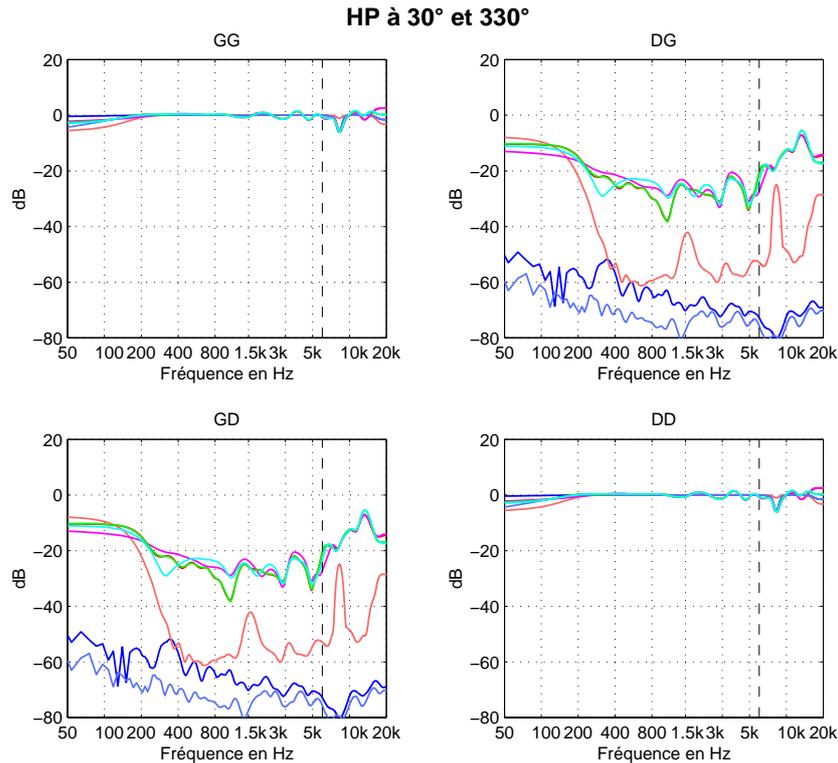


FIGURE III.30 – Gains des fonctions de transfert des trajets acoustiques entre chaque haut-parleur et chaque oreille, pour sept implantations différentes d'un système transaural, correspondant à une restitution sur deux haut-parleurs à $\pm 30^\circ$ à gauche et à droite

posant que les HRIR utilisées sont bien symétriques), et que la modélisation sous forme RII d'ordre 12 est parfaitement efficace (algorithme de Steiglitz-MacBride avec warping fréquentiel, cf. § C.1.7 et § C.1.9). Le taux de réjection bien plus faible (de l'ordre de 25 dB entre 400 Hz et 6 kHz) est à mettre sur le compte de la précision de modélisation du filtre inverse, à savoir 512 points au lieu de 1535 pour les autres configurations testées ; cependant, nous n'avons pas d'autre alternative, étant donné que tous les filtres sont décomposés en module et excès de phase, les HRIR initiales ayant une longueur de 512 points. Mais, un avantage de cette technique est la possibilité relativement simple de limiter la largeur de bande des filtres transauraux, alors qu'avec un calcul direct des filtres sous forme RIF, il faut utiliser une technique plus complexe de filtrage à phase nulle pour séparer les composantes basses et hautes fréquences, sur lesquelles on applique un gain pour rétablir la bonne distribution de puissance au niveau des oreilles (cf. § III.3.3.4). Un autre avantage de cette technique est l'implantation sous forme de filtres RII, nettement moins coûteux en temps réel que les filtres RIF.

Fonctions de transfert Nous avons également simulé la propagation d'un signal binaural depuis son admission à l'entrée du système transaural jusqu'aux oreilles du mannequin ; nous avons utilisé les signaux binauraux utilisés dans le Spatialisateur[®], c.-à-d. les fonctions de transfert d'oreille du mannequin KEMAR modélisées sous forme d'un filtre à phase minimale récursif d'ordre 12, et d'un retard pur. Pour un signal binaural constitué de la HRIR à 0° dans le plan horizontal, et injecté à l'entrée du système transaural, les deux signaux arrivant aux haut-parleurs sont identiques en module et en phase, quelque soit le système transaural utilisé (la figure III.31 représente uniquement les modules des fonctions de transfert).

Les signaux arrivant aux oreilles du mannequin, après passage dans le système transaural et dans le système électroacoustique, sont également identiques en module et en phase (cf. fig. III.32) ; la courbe noire représente le signal binaural injecté, c.-à-d. la HRIR théorique, et on peut constater que tous les

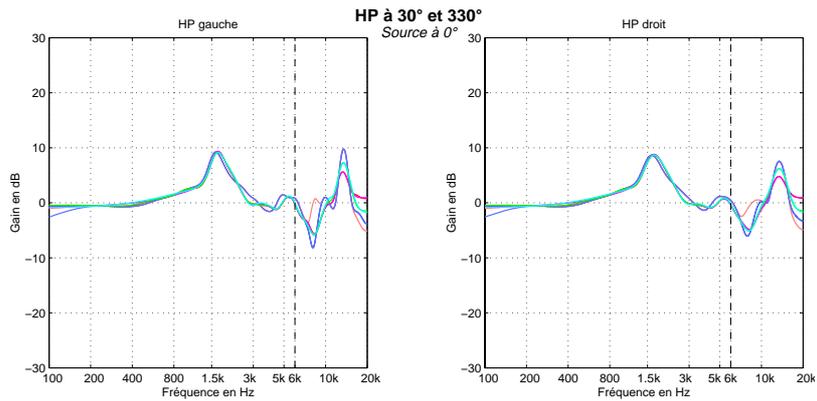
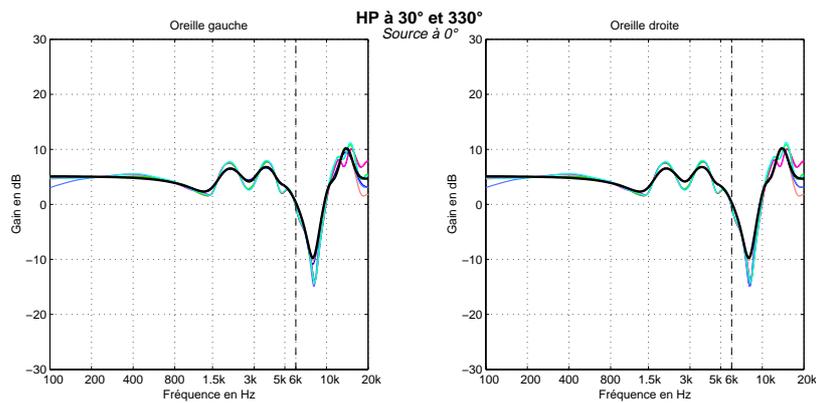
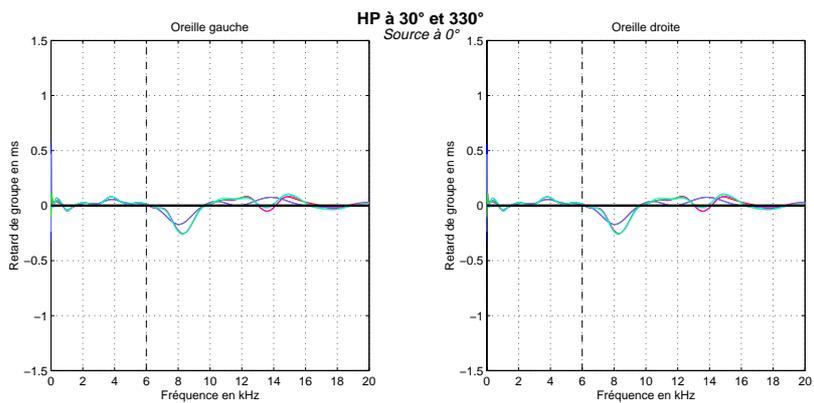


FIGURE III.31 – Module des signaux arrivant sur chacun des haut-parleurs, dans le cas d'un signal binaural constitué de la HRIR à 0° injecté à l'entrée du système transaural, correspondant à une restitution sur deux haut-parleurs à $\pm 30^\circ$ à gauche et à droite



(a) Gain



(b) Retard de groupe

FIGURE III.32 – Module et retard de groupe des signaux arrivant sur chacune des oreilles du mannequin, dans le cas d'un signal binaural constitué de la HRIR à 0° injecté à l'entrée du système transaural, correspondant à une restitution binaurale sur deux haut-parleurs à $\pm 30^\circ$ à gauche et à droite

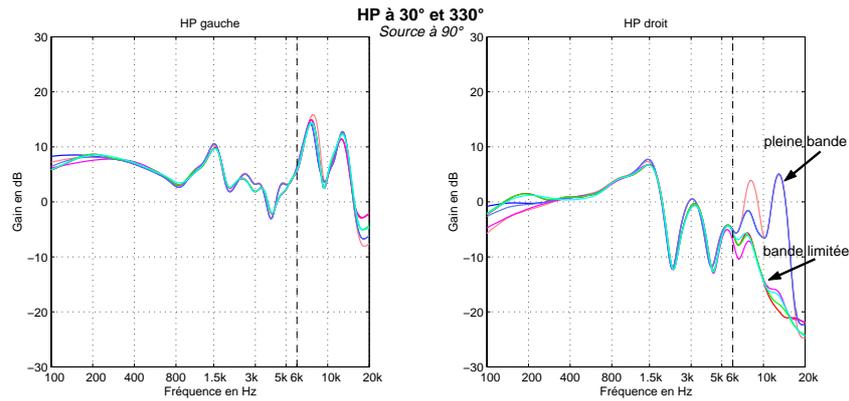


FIGURE III.33 – Module des signaux arrivant sur chacun des haut-parleurs, dans le cas d'un signal binaural constitué de la HRIR à 90° à gauche injecté à l'entrée du système transaural, correspondant à une restitution sur deux haut-parleurs à $\pm 30^\circ$ à gauche et à droite

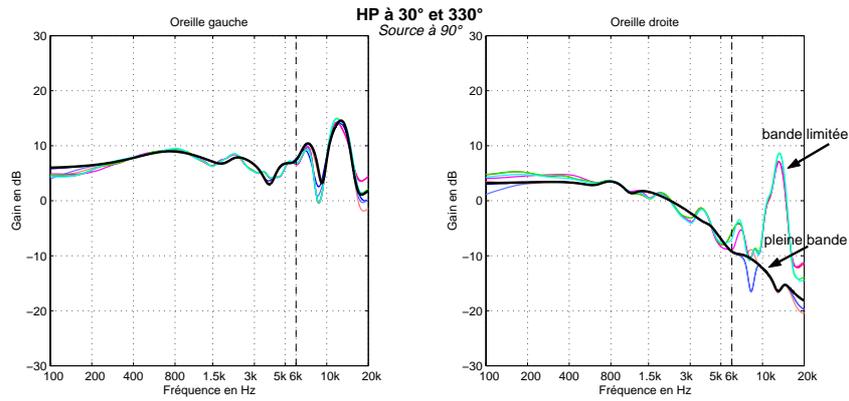
systèmes étudiés permettent de retrouver ce signal aux oreilles du mannequin.

Nous avons refait l'expérience avec un signal binaural constitué de la HRIR à 90° à gauche dans le plan horizontal. Jusqu'à 6 kHz, les signaux arrivant aux haut-parleurs sont identiques quelle que soit l'implantation choisie (cf. fig. III.33). Au-delà de cette limite, et uniquement sur le haut-parleur droit, tous les systèmes transauraux calculés pleine bande se détachent nettement des systèmes optimisés jusqu'à 6 kHz ; ces derniers ont tendance à diminuer fortement le niveau en hautes fréquences, ce qui est perceptivement souhaitable afin d'éviter les sifflements désagréables.

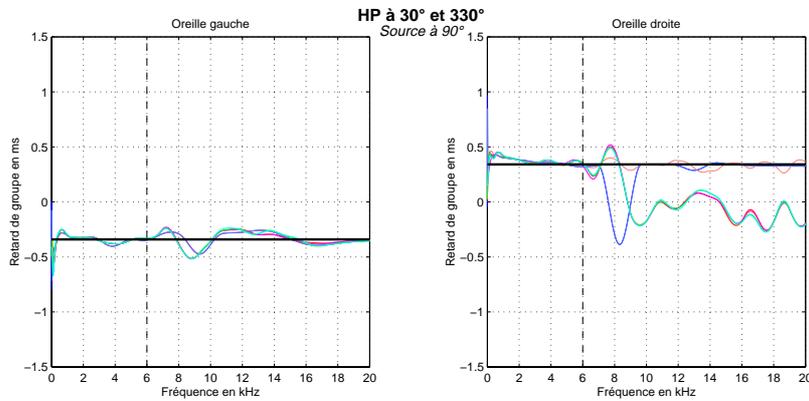
En observant la figure III.33, on peut craindre la perception de sifflements sur le haut-parleur contralatéral, dans le cas d'une inversion pleine bande. Mais en observant la figure III.34(a), il semblerait au contraire que ce soit la transauralisation limitée qui induise une perception de sifflements sur l'oreille contralatérale ; or une comparaison auditive des deux systèmes montre qu'il n'en est rien, sans doute parce que le système n'est pas parfait. En effet, en étudiant le module et la phase des signaux ipsi et contralatéraux arrivant aux oreilles, on s'aperçoit que, pour les systèmes calculés pleine bande, ces signaux devraient théoriquement être identiques en module mais opposés en phase, et donc s'annuler ; perceptivement, cette annulation n'a pas lieu. Pour s'en assurer, nous avons observé le comportement du système avec des HRTF décalées de $\pm 5^\circ$: on constate alors la présence de deux pics dont celui correspondant à l'inversion calculée pleine bande est bien plus important (cf. fig. III.35).

Sur la figure III.34(b), on remarque que le retard de groupe est correctement restitué sur l'oreille ipsilatérale. En revanche, sur l'oreille contralatérale, le retard de groupe restitué par les systèmes transauraux limités en fréquence s'écarte franchement de la valeur théorique à partir de 9 kHz ; le retard de groupe restitué par les systèmes transauraux calculés pleine bande s'écarte de la valeur idéale entre 7 et 10 kHz, zone correspondant à une régularisation importante (cf. fig. III.29).

Retard interaural Nous avons calculé, pour chaque système transaural étudié, le retard interaural des signaux arrivant à chaque oreille, et nous les avons comparé avec le retard interaural théorique, c.-à-d. celui du signal binaural injecté à l'entrée du système transaural. Pour cela nous avons injecté à l'entrée du système transaural une distribution complète de HRIR tous les 5° , et pour chaque HRIR nous avons calculé le retard interaural ; celui-ci est obtenu grâce à la procédure décrite par Larcher [176], à savoir par régression linéaire sur la plus grande portion linéaire de la différence gauche-droite des excès de phase des signaux binauraux, entre deux fréquences déterminées (200 et 5000 Hz, ici). Comme on peut le constater sur la figure III.36, tous les systèmes étudiés permettent de retrouver la bonne valeur du retard interaural.



(a) Gain



(b) Retard de groupe

FIGURE III.34 – Module et retard de groupe des signaux arrivant sur chacune des oreilles du mannequin, dans le cas d'un signal binaural constitué de la HRIR à 90° à gauche injecté à l'entrée du système transaural, correspondant à une restitution sur deux haut-parleurs à $\pm 30^\circ$ à gauche et à droite

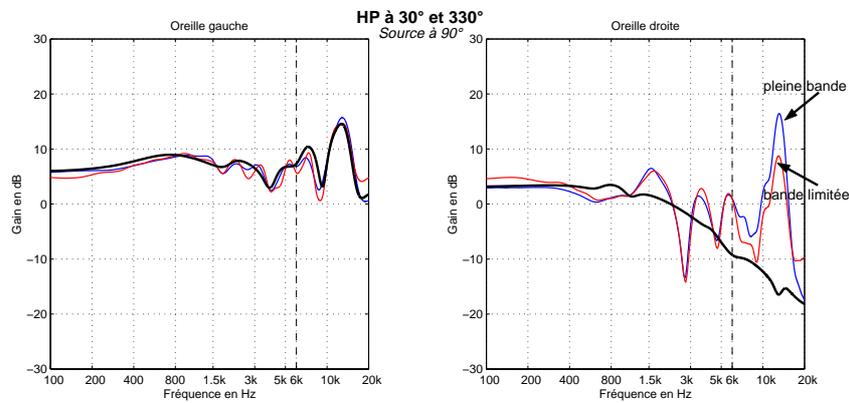


FIGURE III.35 – Idem figure III.34, pour une restitution sur deux haut-parleurs à $\{-25^\circ; +35^\circ\}$ à gauche et à droite

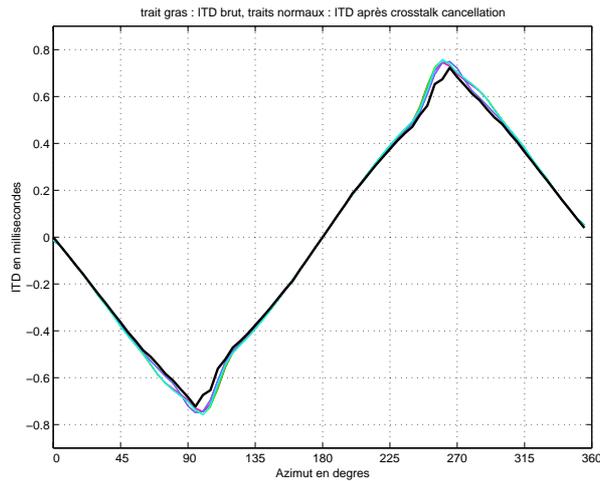


FIGURE III.36 – Retard interaural entre les signaux arrivant aux oreilles du mannequin, comparé avec le retard interaural théorique (courbe noire épaisse), pour un système de restitution symétrique composé de deux haut-parleurs à $\pm 30^\circ$

III.3.4.2 Dispositif de restitution non symétrique

Nous avons refait les mêmes tests en supposant un dispositif de restitution non symétrique, en l'occurrence un haut-parleur à 10° à gauche et un autre à 50° à droite (ces angles correspondent à peu près aux valeurs rencontrées en automobile). Dans ce cas, nous comparons :

- l'implantation sous forme classique, conforme à la figure III.24, et mettant en œuvre quatre filtres RIF ;
- l'implantation sous forme shuffler général, conforme à la figure III.25(b), mettant en œuvre quatre filtres RIF calculés directement d'après le système d'équations correspondantes (III.29) ;
- l'implantation sous forme shuffler général, conforme à la figure III.25(b), mettant en œuvre quatre filtres RIF ou RII calculés à partir d'une décomposition en module et excès de phase, ce qui permet de limiter la largeur de bande du système transaural.

Sur les figures suivantes, les couleurs sont réparties comme suit :

- bleu → implantation classique
- bleu plus clair → implantation générale shuffler, avec quatre filtres RIF
- vert → implantation générale shuffler, avec quatre filtres RIF calculés à partir d'une décomposition en module et excès de phase, avec limitation de la largeur de bande du système transaural
- cyan → implantation générale shuffler, avec quatre filtres RII calculés à partir d'une décomposition en module et excès de phase, avec limitation de la largeur de bande du système transaural

Les éléments de la matrice de transfert électroacoustique directe \mathbf{H} sont présentés figure III.37 ; dans un souci de clarté, nous n'avons représenté que les 100 premiers points (sur les 512). Il faut remarquer que :

- chaque réponse impulsionnelle a un retard différent, ce qui, à la différence du système symétrique, nécessite *obligatoirement* d'effectuer une inversion incluant la composante à phase non minimale (le concept de phase minimale jointe ne peut plus s'appliquer dans le cas de la structure shuffler) ;
- les différences entre les amplitudes de chaque élément, dues à l'atténuation sphérique et à la diffraction par la tête, requièrent une dynamique plus importante des filtres inverses, comparées au système symétrique.

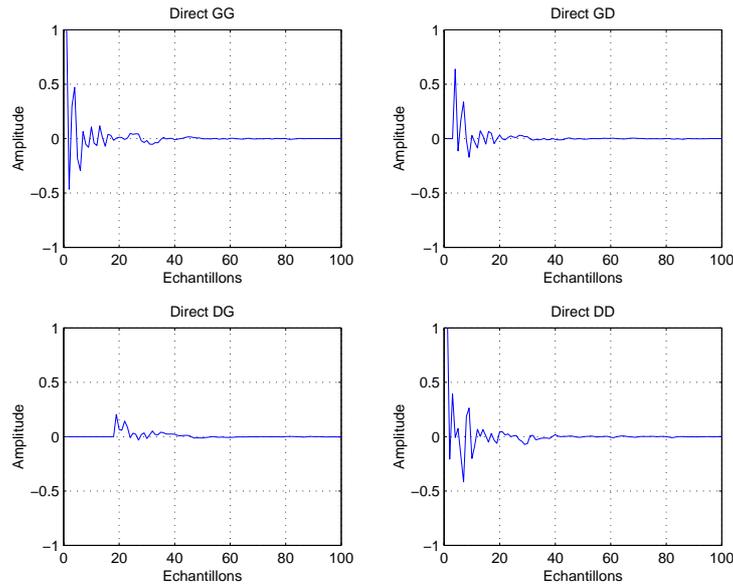


FIGURE III.37 – Fonctions de transfert entre haut-parleurs et oreilles, mesurées en chambre anéchoïque, pour un système de restitution non symétrique composé d'un haut-parleur à 10° à gauche et d'un haut-parleur à 50° à droite

Chaque HRIR comportant 512 points, Δ en comporte 1023, et nous choisissons de calculer le filtre inverse sur 1024 points (cf. fig III.38). La précision du résidu est de 30 dB environ, et nous la considérons comme suffisante. Nous avons écouté l'influence de ce filtre sur un signal musical, et il n'est pas apparu de phénomènes désagréables.

Nous pouvons maintenant procéder à la comparaison des différentes structures d'inversion des trajets croisés.

Annulation des trajets croisés Sur la figure III.39, on peut constater que les trajets directs sont correctement restitués, mis à part la légère faiblesse en basses fréquences, due au filtre de régularisation.

Pour les trajets croisés, on constate deux ensembles de courbes :

- les courbes bleues, correspondant à la mise en œuvre de quatre filtres RIF de 1535 points, pour lesquelles le taux de réjection est supérieur à 60 dB, mis à part une légère faiblesse en basses fréquences, toujours due à la régularisation ;
- la courbe verte et la courbe cyan, correspondant à une technique de calcul de filtres par décomposition en module et excès de phase, et permettant une limitation du système transaural à 6 kHz, pour lesquelles le taux de réjection est de 20 dB entre 200 et 6000 Hz ; les quatre filtres sont laissés sous forme RIF de 512 points (correspondant à la longueur des HRIR originales) ou modélisés sous forme RII d'ordre 12.

Les deux courbes bleues montrent, comme dans le cas symétrique, que le filtrage RIF permet de réaliser un système d'annulation des trajets croisés très efficace.

Les courbes vertes et cyan mettent en évidence une réjection bien plus faible, due essentiellement à la plus faible longueur des filtres transauraux ; ici encore, la modélisation RII proprement dite est parfaitement efficace, puisque ces deux courbes sont pratiquement semblables.

Fonctions de transfert Pour un signal binaural constitué de la HRIR à 0° dans le plan horizontal, et injecté à l'entrée du système transaural, les deux signaux arrivant aux haut-parleurs sont globalement identiques à la fois en module et en phase, quelque soit le système transaural utilisé (la figure III.40 ne

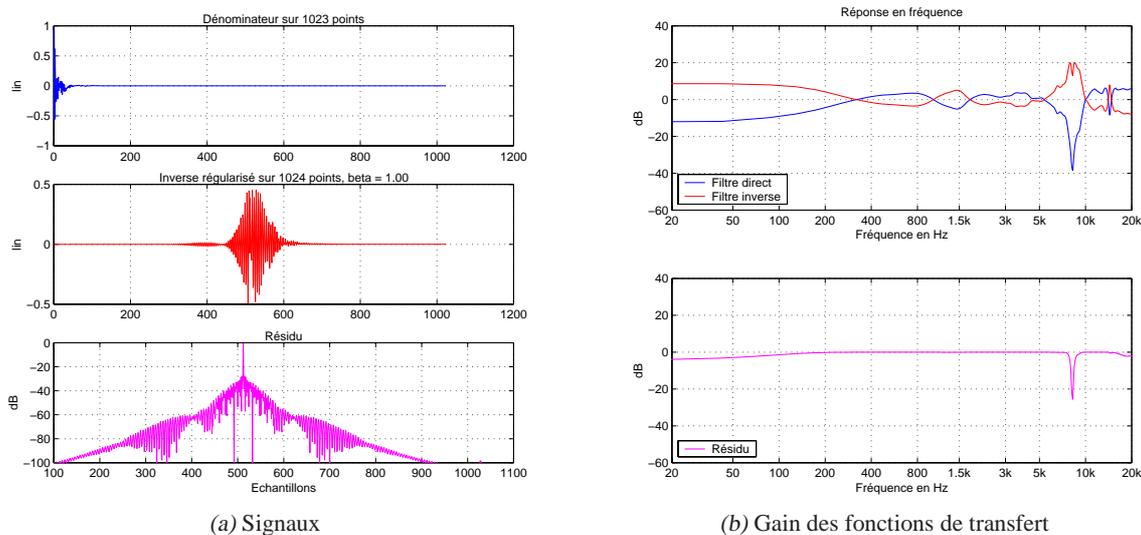


FIGURE III.38 – Inversion du déterminant de la matrice de transfert électroacoustique directe, dans le cas d'une écoute en chambre sourde sur un dispositif de restitution non symétrique

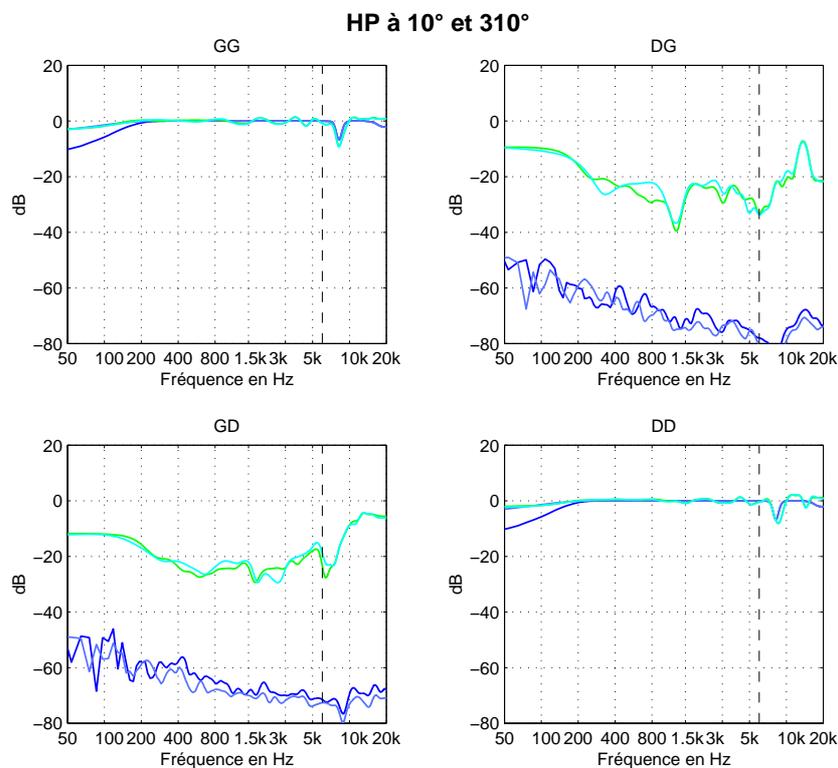


FIGURE III.39 – Gains des fonctions de transfert des trajets acoustiques entre chaque haut-parleur et chaque oreille, pour quatre implantations différentes du système transaural, correspondant à une restitution sur un haut-parleur à 10° à gauche et un haut-parleur à 50° à droite

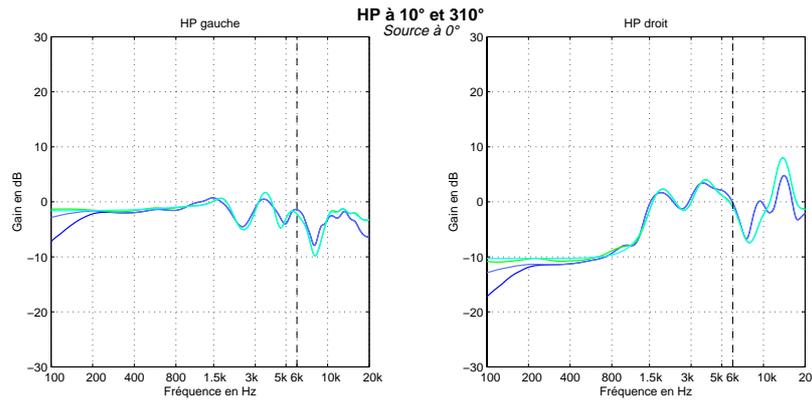


FIGURE III.40 – Module des signaux arrivant sur chacun des haut-parleurs, dans le cas d'un signal binaural constitué de la HRIR à 0° injecté à l'entrée du système transaural, correspondant à une restitution sur un haut-parleur à 10° à gauche et un haut-parleur à 50° à droite

représente que les modules) ; on constate cependant une différence au-delà de 6 kHz sur le haut-parleur de droite.

Les signaux arrivant aux oreilles du mannequin acoustique (cf. fig. III.41), après passage dans le système transaural et dans le système électroacoustique, sont également identiques en module et en phase ; la courbe noire représente le signal binaural injecté, et on peut constater que tous les systèmes étudiés permettent de retrouver ce signal aux oreilles du mannequin, avec cependant une amplification plus importante des résonances et antirésonances pour le système shuffler dont les filtres sont calculés par décomposition en module et excès de phase.

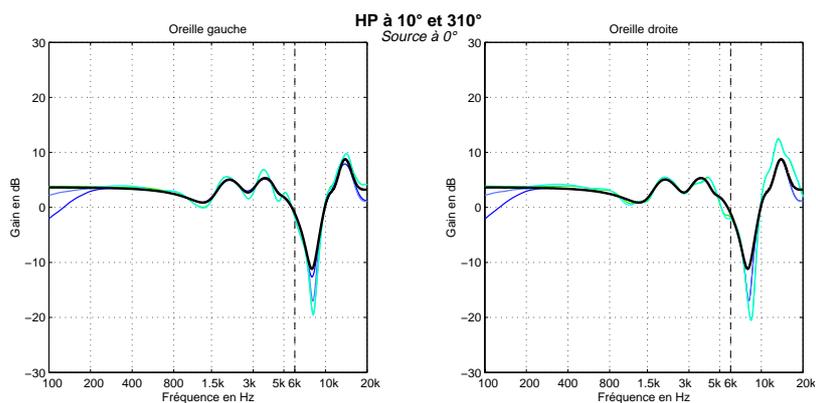
Nous avons refait l'expérience avec un signal binaural constitué de la HRIR correspondant à une source à 90° à gauche dans le plan horizontal, et nous pouvons faire les mêmes remarques qu'avec le dispositif de restitution symétrique. Les systèmes calculés pleine bande font clairement apparaître, au niveau des haut-parleurs, des résonances très marquées en hautes fréquences, et qui se traduisent perceptivement par des sifflements très désagréables pouvant nuire à la localisation de la source virtuelle. Mais la validation objective au niveau des oreilles du mannequin montre toujours une inversion du phénomène (cf. l'analyse du problème page 168). Les résultats sont rassemblés sur les figures III.42 et III.43.

Sur la figure III.43(b), on remarque que les systèmes transauraux calculés pleine bande donnent une valeur correcte du retard de groupe sur les deux oreilles. Comme dans le cas d'un dispositif de restitution symétrique, le retard de groupe observé sur l'oreille contralatérale, à l'aide des systèmes calculés sur une bande limitée, se détache nettement de la valeur idéale à partir de 8 kHz ; pour le système implanté sous forme récursive, on constate, comme sur le module, une erreur importante autour de 5 kHz.

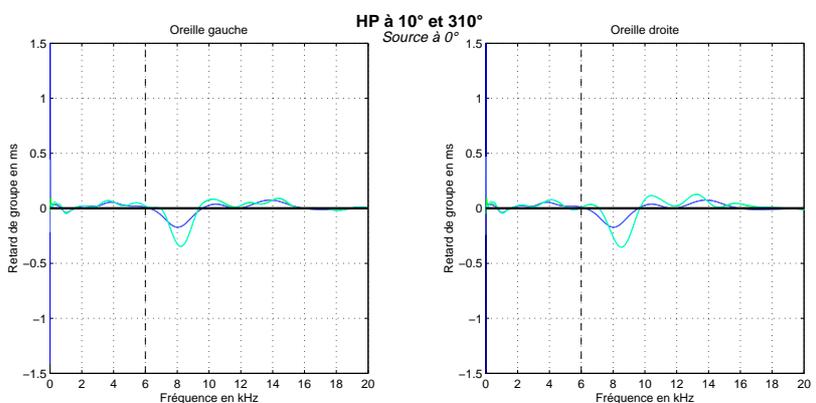
Retard interaural Nous avons calculé, pour chaque système transaural étudié, le retard interaural des signaux arrivant à chaque oreille, et nous les avons comparé avec le retard interaural théorique, c.-à-d. celui du signal binaural injecté à l'entrée du système transaural. Comme on peut le constater sur la figure III.44, les systèmes transauraux basés sur le calcul de filtres RIF spécifiés pleine bande permettent de retrouver la bonne valeur du retard interaural, pour tous les azimuts dans le plan horizontal ; par contre, les systèmes transauraux mettant en jeu des filtres RIF ou RII optimisés jusqu'à 6 kHz présentent une erreur entre les azimuts 95° et 125° (et de même pour le secteur angulaire symétrique) : l'extremum est atteint pour 110° environ (resp. 250°), et sa valeur est majorée de 20%.

III.3.4.3 Conclusion

Nous venons d'étudier deux structures de filtres transauraux, la structure classique mettant en œuvre quatre filtres RIF, et la structure shuffler, mettant en œuvre quatre (ou deux par symétrie) filtres RIF ou



(a) Gain



(b) Retard de groupe

FIGURE III.41 – Module et retard de groupe des signaux arrivant sur chacune des oreilles du mannequin, dans le cas d'un signal binaural constitué de la HRIR à 0° injecté à l'entrée du système transaural, correspondant à une restitution sur un haut-parleur à 10° à gauche et un haut-parleur à 50° à droite

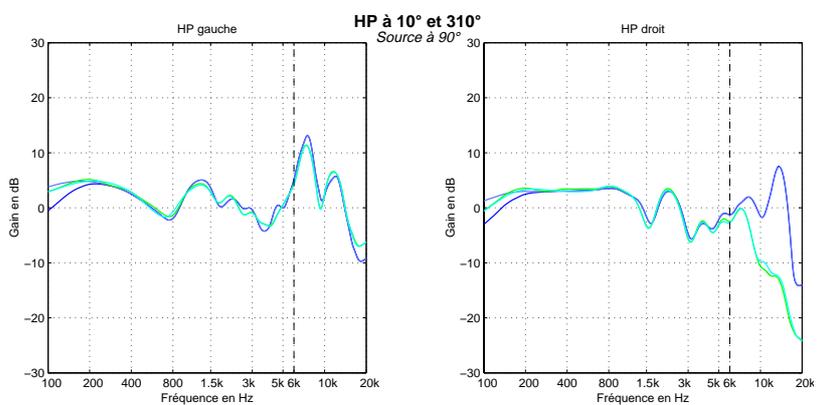


FIGURE III.42 – Module des signaux arrivant sur chacun des haut-parleurs, dans le cas d'un signal binaural constitué de la HRIR à 90° à gauche injecté à l'entrée du système transaural, correspondant à une restitution sur un haut-parleur à 10° à gauche et un haut-parleur à 50° à droite

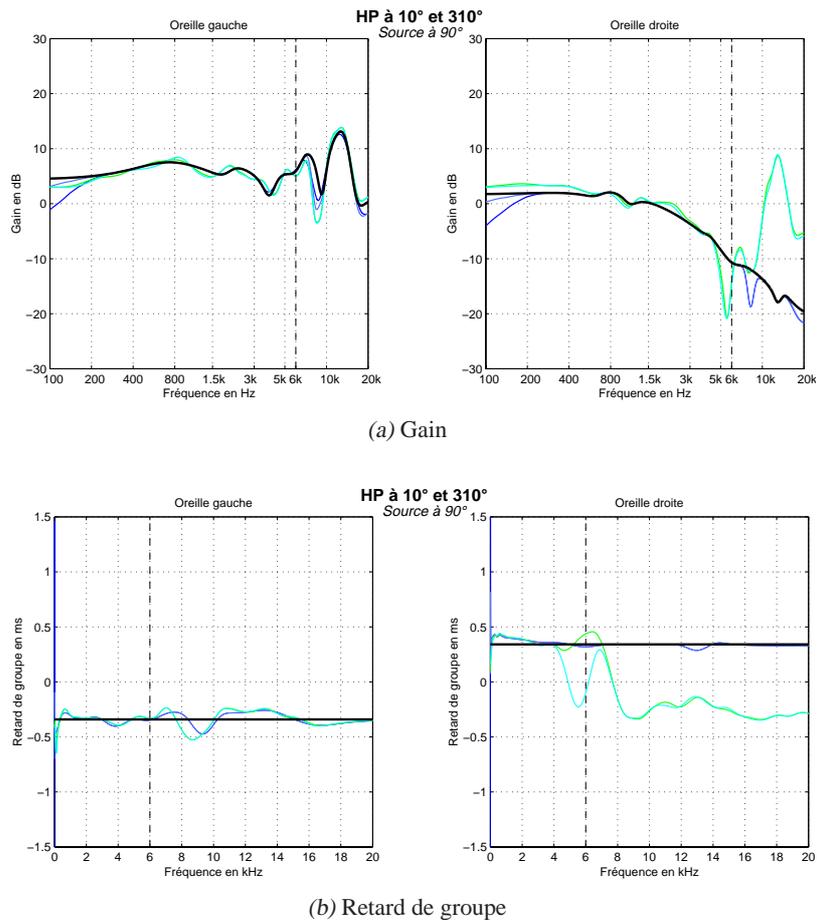


FIGURE III.43 – Module et retard de groupe des signaux arrivant sur chacune des oreilles du mannequin, dans le cas d'un signal binaural constitué de la HRIR à 90° injecté à l'entrée du système transaural, correspondant à une restitution sur un haut-parleur à 10° à gauche et un haut-parleur à 50° à droite

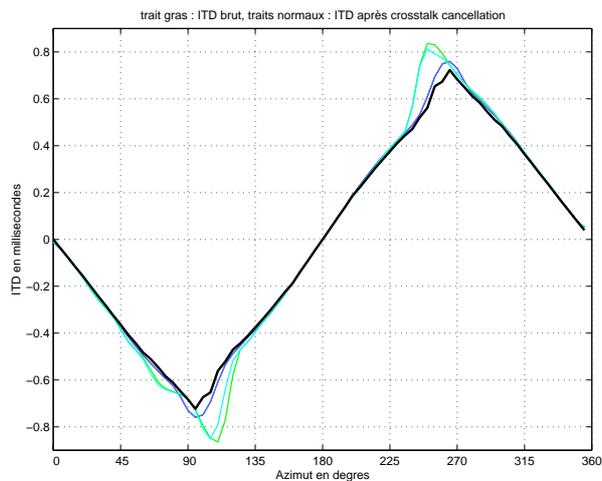


FIGURE III.44 – Retard interaural entre les signaux arrivant aux oreilles du mannequin, comparé avec le retard interaural théorique (courbe noire épaisse), pour un système de restitution non symétrique composé d'un haut-parleur à 10° à gauche et d'un haut-parleur à 50° à droite

RII ; dans cette dernière, les filtres peuvent être calculés par décomposition en module et excès de phase, ce qui permet de limiter facilement l'efficacité du système transaural à 6 kHz, largeur suffisante pour garantir la pertinence auditive des traitements effectués.

Nous avons étudié ces structures dans le cas d'un dispositif de restitution symétrique, mais aussi dans le cas d'un dispositif de restitution non symétrique. Dans ce cas, il est nécessaire d'effectuer une inversion complète (module et phase) du déterminant de la matrice de transfert électroacoustique ; comme il n'existe pas de solution exacte, nous avons choisi l'optimisation au sens des moindres carrés, avec régularisation par filtrage, calculée dans le domaine temporel.

Nous avons vérifié objectivement que le calcul des filtres effectué directement sous forme RIF (dans une structure classique ou une structure shuffler, symétrique ou non symétrique) permet d'atteindre des résultats plus ambitieux, en regard de l'annulation des trajets croisés, que le calcul des filtres effectué sous forme RIF après décomposition en module et excès de phase (dans une structure shuffler, symétrique ou non symétrique). Nous y voyons deux explications : d'abord la seconde méthode conduit à des filtres de longueur 512 points, puisque les fonctions de transfert directes sont stockées sur ce même nombre de points. Ensuite, la seconde méthode induit une régularisation par limitation de la dynamique du module, sans toucher à l'excès de phase ; cette méthode est facile à mettre en œuvre, mais cause nécessairement une erreur sur le filtre inverse (absence d'adéquation entre module et phase, au niveau des zones de régularisation) ; l'erreur est cependant assez faible ici. La deuxième méthode permet en revanche de limiter facilement la largeur de bande des filtres de la matrice inverse, et permet également de générer des filtres RII, peu coûteux en temps de calcul. Alors que la première méthode permet de tailler au mieux le filtre inverse du déterminant, à la fois en termes objectifs et perceptifs, mais génère des filtres beaucoup plus long dans la matrice inverse, d'où un coût de calcul beaucoup plus élevé.

Nous avons implanté toutes ces configurations dans le logiciel Max/MSP®, et nous avons effectué un test informel. Perceptivement, il faut admettre que les systèmes calculés pleine bande présentent des sifflements en hautes fréquences ; ce défaut nuit à la localisation correcte de la source virtuelle, et il est nécessaire de limiter cette bande passante, par l'une des deux méthodes présentée au § III.3.3.4.

Cependant, plus l'angle entre les haut-parleurs et la tête de l'auditeur est faible, plus cette limite doit être repoussée en hautes fréquences, car le masquage de la tête est de moins en moins efficace pour les réponses impulsionnelles contralatérales ; ainsi, pour une configuration en dipôle stéréophonique, les filtres doivent être calculés sur une bande très large [158, 160].

III.3.5 Validation en véhicule

L'inversion du déterminant de la matrice de transfert électroacoustique directe va ici se poser de manière plus radicale, puisque la composante passe-tout des fonctions de transfert supporte un effet acoustique composé de réflexions multiples (réverbération). La matrice inverse réalise donc l'annulation de cet effet d'habitacle.

Les réponses impulsionnelles binaurales (RIB) ont été mesurées avec le mannequin HMS II, de marque HEAD ACOUSTICS, appartenant à la Direction de l'Ingénierie Véhicule de RENAULT. Les filtres transauraux sont donc *théoriquement* calculés pour être écoutés par ce même mannequin. Comme dans le cas des mesures de réponses impulsionnelles d'oreille en chambre sourde, il est préférable « d'adapter » les mesures de réponses impulsionnelles afin qu'elles soient valides pour un grand nombre de sujets ; il est donc possible de normaliser les réponses impulsionnelles binaurales par la fonction de transfert en champ diffus du mannequin.

III.3.5.1 Calcul des filtres transauraux

Les RIB contenant l'effet acoustique de l'habitacle, il faut conserver un nombre d'échantillons garantissant que le maximum d'information est présent. Le temps de réverbération moyen étant de 50 ms,

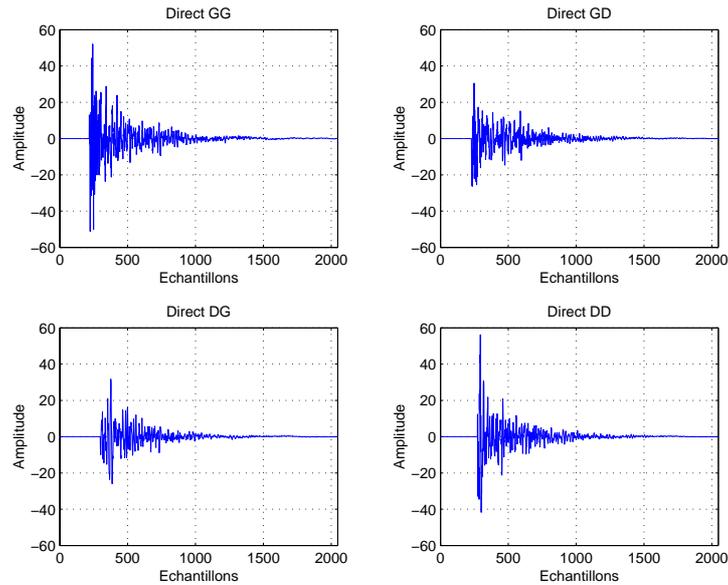


FIGURE III.45 – Fonctions de transfert entre haut-parleurs et oreilles, mesurées dans un véhicule de type Safrane

et la fréquence d'échantillonnage de 44,1 kHz, nous choisissons de conserver 2048 points. On pourrait choisir un nombre de points plus élevé, mais les filtres de la matrice inverse comporteraient plus de points, d'où un coût de calcul plus élevé. On pourrait aussi choisir un nombre de points plus faibles, mais les basses fréquences ne seraient pas totalement conservées. En tout état de cause, l'effet d'habitacle ne sera inversé que sur ces 2048 premiers points. Les éléments de la matrice de transfert électroacoustique directe \mathbf{H} sont présentés figure III.45.

L'inversion du déterminant Δ de la matrice \mathbf{H} est présenté figure III.46. Nous choisissons de conserver 2048 points du déterminant (sur 4095 au total), et de calculer l'inverse sur 2048 points également. La précision du résidu est de 30 dB environ, et nous la considérons comme suffisante. L'influence de la régularisation est clairement visible sur la figure III.46(b). Nous avons écouté l'influence de ce filtre sur un signal musical, et il n'est pas apparu de phénomènes désagréables. Nous pouvons maintenant procéder à la comparaison des différentes structures d'inversion des trajets croisés.

Nous nous proposons de comparer :

- l'implantation sous forme classique, conforme à la figure III.24, et mettant en œuvre quatre filtres RIF ;
- l'implantation sous forme shuffler général, conforme à la figure III.25(b), mettant en œuvre quatre filtres RIF calculés directement d'après le système d'équations correspondantes (III.29) ;
- l'implantation sous forme shuffler général, conforme à la figure III.25(b), mettant en œuvre quatre filtres RIF calculés à partir d'une décomposition en module et excès de phase, ce qui permet de limiter la largeur de bande du système transaural.

Dans les deux premiers cas, la longueur des filtres de la matrice inverse est donc de 4095 points, puisque les filtres directs comportent 2048 points et que l'inverse du déterminant de la matrice, dont on a conservé les 2048 premiers points, est également calculé sur 2048 points. Dans le dernier cas, la longueur des filtres de la matrice inverse est de 4096 points, puisque pour éviter le repliement temporel dans le calcul de l'inverse du déterminant, la décomposition en module et excès de phase est effectuée sur 4096 points.

La régularisation par limitation de la dynamique du module, effectuée lors de l'inversion du déterminant dans l'implantation shuffler général à bande limitée, génère ici des erreurs beaucoup plus impor-

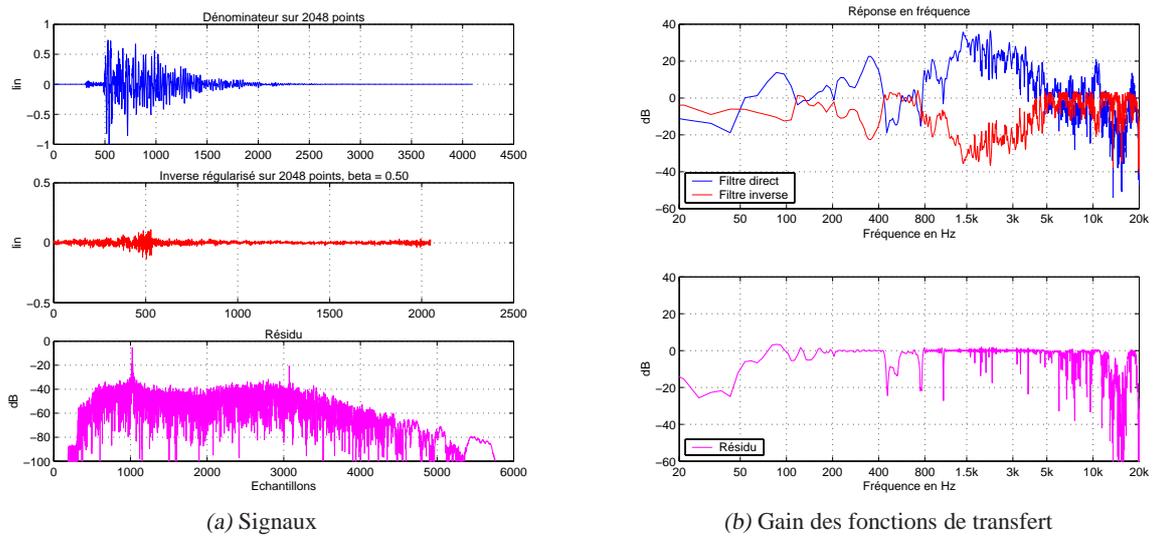


FIGURE III.46 – Inversion du déterminant de la matrice de transfert électroacoustique directe, dans le cas d'une écoute dans un véhicule de type Safrane

tantes que dans le cas où le système transaural est calculé pour être restitué en chambre anéchoïque. Cet effet est dû aux nombreuses zones du module devant subir l'opération de régularisation. Il faut donc à tout prix effectuer le calcul de l'approximation du filtre inverse en utilisant une technique de régularisation par filtrage, nettement plus onéreuse, mais nettement plus précise.

D'autre part, la modélisation des filtres de la matrice inverse sous forme RII est complètement illusoire.

Sur les figures suivantes, les couleurs sont réparties comme suit :

- bleu → implantation classique
- bleu plus clair → implantation générale shuffler, avec quatre filtres RIF
- vert → implantation générale shuffler, avec quatre filtres RIF calculés à partir d'une décomposition en module et excès de phase, avec limitation de la largeur de bande du système transaural

Annulation des trajets croisés Sur la figure III.47, on peut constater que les filtres calculés remplissent globalement leur rôle. Les trajets directs sont correctement restitués, mis à part une légère faiblesse en basses et hautes fréquences, due à la régularisation ; il en est de même autour de 500 Hz.

Pour les trajets croisés, on constate que ceux-ci subissent une réjection supérieure à 20 dB à partir de 200 Hz ; la limitation en hautes fréquences est parfaitement visible pour le système représenté par la courbe verte.

Fonctions de transfert Les signaux arrivant aux oreilles du mannequin acoustique après passage dans le système transaural et dans le système électroacoustique, sont présentés figure III.48 ; la courbe noire représente toujours la fonction de transfert du signal binaural injecté. Le système transaural ayant été calculé à partir des mesures de RIB non égalisées par la fonction de transfert en champ diffus du mannequin, nous avons aussi utilisé des signaux binauraux non égalisés (à la différence de la validation en chambre sourde). On peut constater que tous les systèmes étudiés permettent de retrouver globalement ce signal aux oreilles du mannequin, même si il apparaît clairement des antirésonances sur les modules des fonctions de transfert des signaux passés au travers du système transaural et du système électroacous-

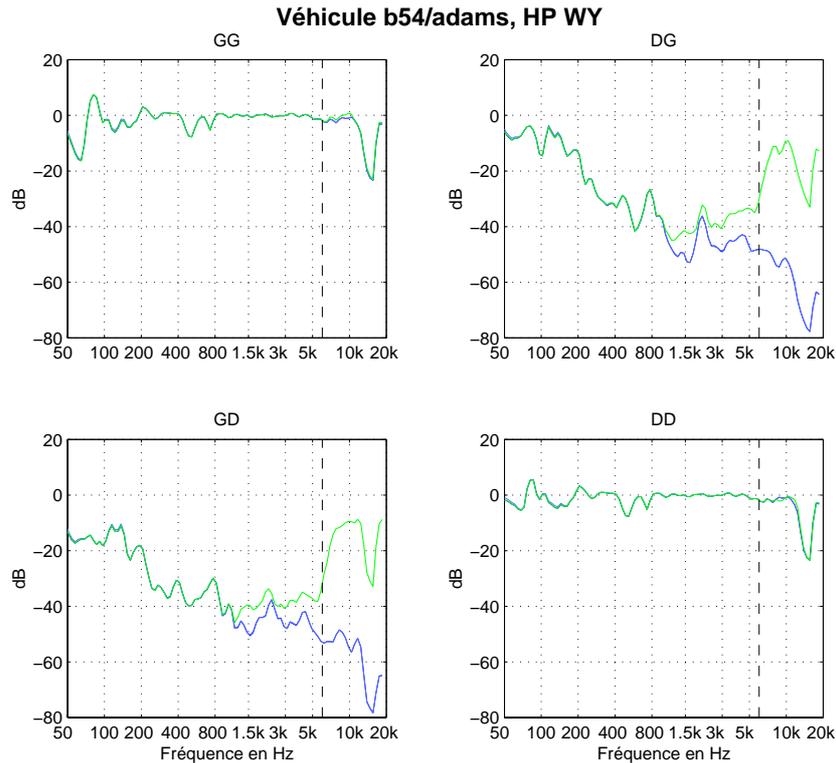


FIGURE III.47 – Gains des fonctions de transfert des trajets acoustiques entre chaque haut-parleur et chaque oreille, pour trois implantations différentes du système transaural, correspondant à une restitution binaurale sur les deux haut-parleurs avant d'un véhicule de type Safrane

tique. Ces antirésonances ont lieu aux alentours de 500 Hz, région qui correspond à une régularisation importante (cf. fig. III.46(b)) ; nous avons déjà noté cette faiblesse sur la figure III.47 représentant les gains des fonctions de transfert des trajets acoustiques directs. Mis à part ces imperfections, les modules des fonctions de transfert binaurales sont correctement reproduits. On notera aussi que le système transaural à bande limitée est moins précis en hautes fréquences, mais cela est justifié par des considérations perceptives dont nous nous sommes déjà fait l'écho.

Retard interaural Nous avons calculé, pour chaque système étudié, le retard interaural (ITD) des signaux arrivant sur chaque oreille, entre 200 Hz et 5000 Hz. La méthode de calcul utilisée (régression linéaire sur l'excès de phase) impose de considérer que le système transaural annule complètement l'acoustique de l'habitacle ; le calcul des filtres transauraux étant effectués à partir des 2048 premiers points de la matrice de transfert électroacoustique directe, le calcul des signaux arrivant aux oreilles doit être effectué en ne considérant que ces mêmes 2048 premiers points. Nous avons alors comparé ce retard interaural avec le retard interaural théorique, obtenu à partir de mesures de HRIR effectuées en chambre sourde (cf. fig. III.49). Les résultats sont très satisfaisants, et confirment que l'on est en mesure de percevoir une localisation correcte des sources sonores virtuelles.

Corrélation interaurale Nous avons également calculé l'intercorrélation des signaux arrivant sur chacune des oreilles du mannequin, en considérant cette fois la totalité des fonctions de transfert directes (l'effet d'habitacle n'est donc que partiellement annulé). En repérant l'instant d'arrivée du maximum de cette corrélation, on peut en déduire le retard interaural (ITD) et le coefficient IACC (cf. § III.4.1). Le calcul a été mené lorsqu'un signal binaural est injecté à l'entrée du système transaural, puis diffusé par les haut-parleurs du véhicule, et capté par les oreilles du mannequin placé en position conducteur. La corrélation interaurale reste très élevée, même pour une source virtuelle localisée à 90° à gauche, c.-à-d.

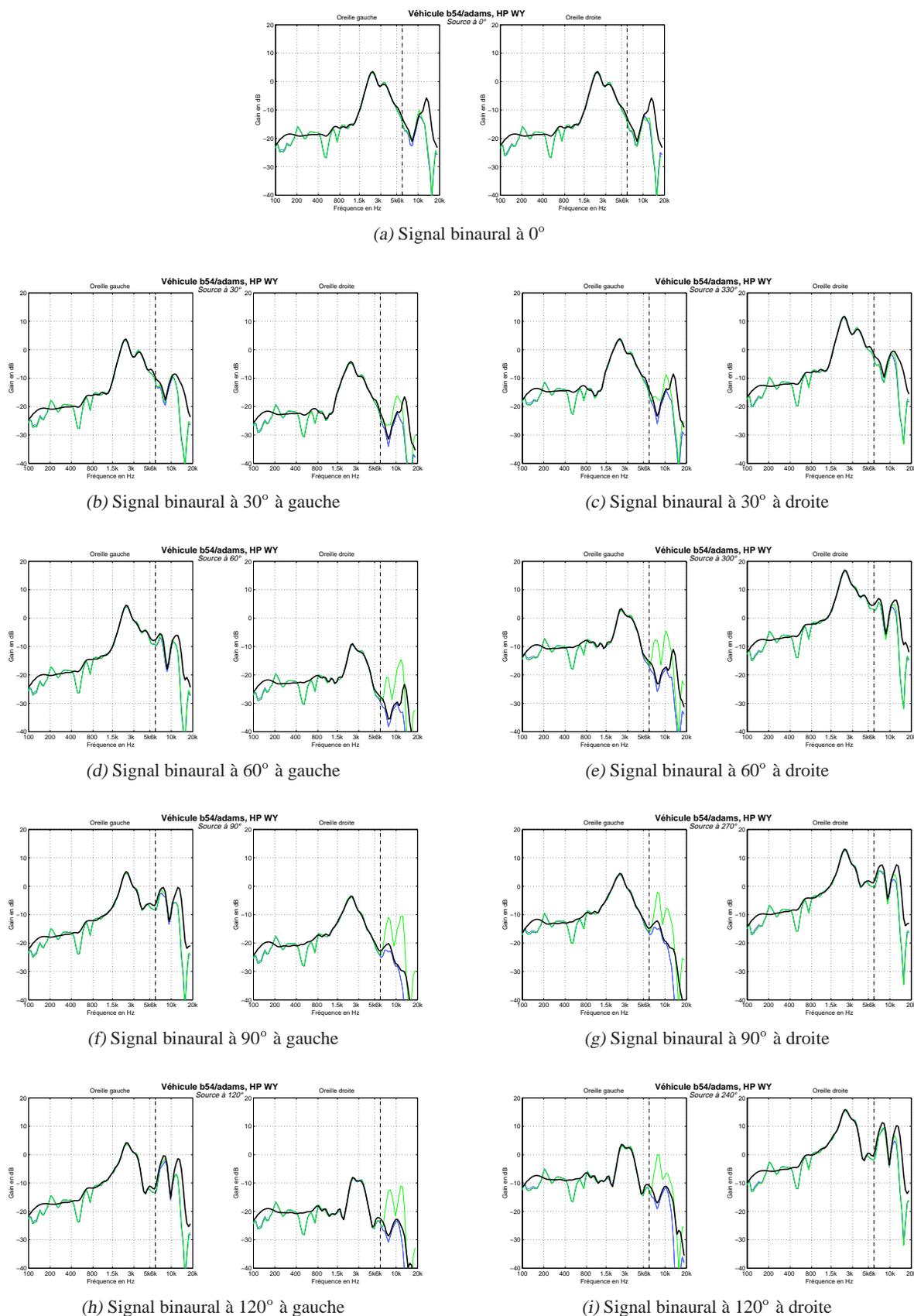


FIGURE III.48 – Module des signaux arrivant sur chacune des oreilles du mannequin, correspondant à une restitution binaurale sur les deux haut-parleurs avant d'un véhicule de type Safrane

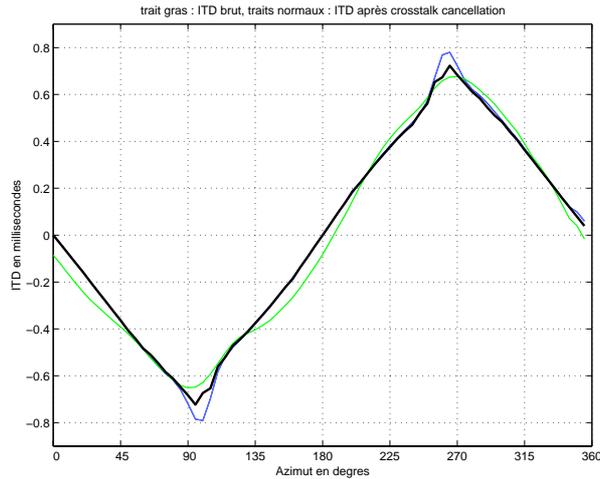


FIGURE III.49 – Retard interaural entre les signaux arrivant aux oreilles du mannequin, comparé avec le retard interaural théorique (courbe noire épaisse), pour une restitution binaurale sur les deux haut-parleurs avant d'un véhicule de type Safrane

à l'extérieur de l'habitacle, que nous sommes parfaitement en mesure de localiser avec précision.

III.3.5.2 Influence de la position de l'auditeur

Les systèmes transauraux que nous avons étudiés ont été implantés dans le logiciel Max/MSP[®], et nous avons effectué un test informel en véhicule. Le positionnement de sources virtuelles à droite du conducteur est correct, mais il est un peu moins satisfaisant à gauche, au fur et à mesure que l'on s'écarte de la direction frontale. Cependant, le positionnement de deux sources virtuelles à $\pm 30^\circ$ nous a paru convaincant ; l'objectif que nous nous étions fixé est donc atteint.

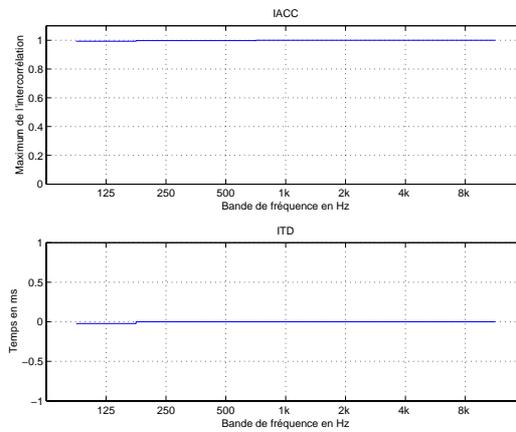
Dans le même temps, nous avons annulé de l'effet acoustique de l'habitacle (sur les 2048 premiers points), et il devient alors possible d'ajouter un effet de salle plus en accord avec le message musical choisi par l'auditeur.

De manière auditive, nous avons pu nous apercevoir que la zone d'écoute optimale est assez réduite, et le positionnement de l'auditeur assez sensible. L'extension de la zone d'écoute est de l'ordre de moins d'une dizaine de centimètres de gauche à droite ; elle est cependant plus importante d'avant en arrière. Ce résultat n'est pas étonnant, dans la mesure où le phénomène bien connu existe aussi lors d'une restitution en chambre sourde.

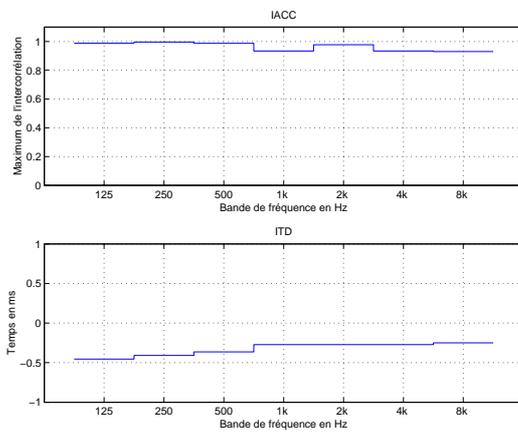
De manière objective, nous avons examiné la robustesse du système en utilisant les filtres transauraux calculés pour une écoute au point A avec les réponses impulsionnelles binaurales mesurées au point B. Pour cela, nous avons utilisé des mesures effectuées le même jour, mais avec une position différente du siège sur lequel était posé le mannequin (assise reculée et dossier penché) ; la position des oreilles a varié de plus de dix centimètres dans les trois directions de l'espace.

Sur la figure III.51, on peut constater que l'annulation des trajets croisés est de l'ordre de 10 dB jusque 2 kHz, et inexistante au-delà.

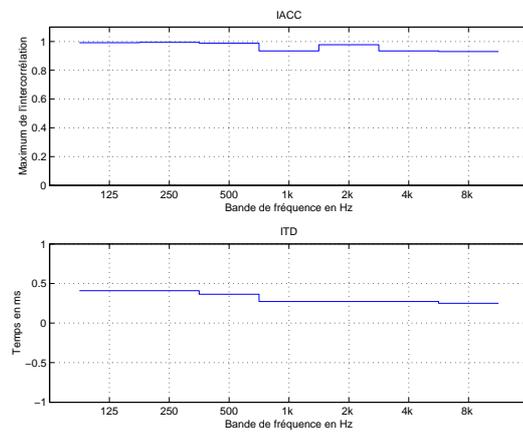
Par contre, l'ITD calculé en basses fréquences entre 200 et 500 Hz est assez bien reproduit, comme on peut le constater figure III.52. Globalement, les valeurs sont sous-évaluées. L'allure générale de la courbe est conservée pour les azimuts compris entre 0° et 180° , c.-à-d. à gauche de l'auditeur ; en revanche, il n'en est pas de même pour les azimuts compris entre 225° et 330° . On peut donc craindre une



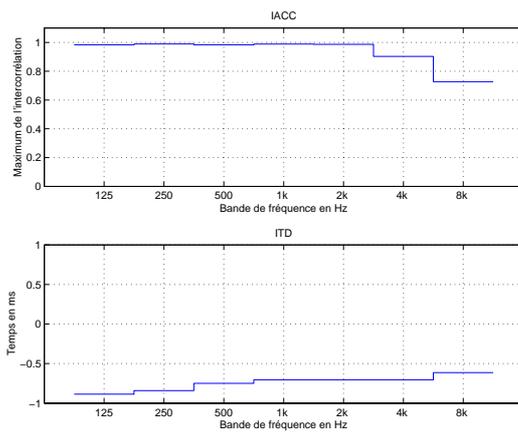
(a) Signal binaural à 0°



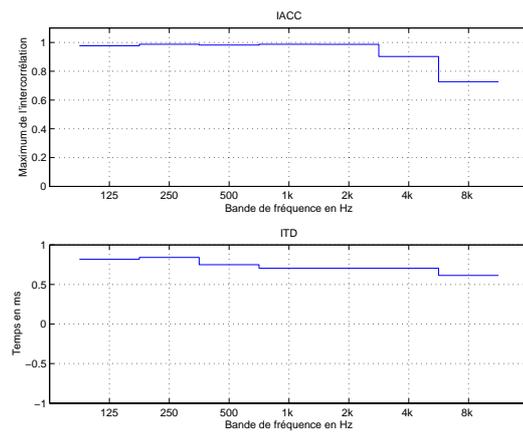
(b) Signal binaural à 30° à gauche



(c) Signal binaural à 30° à droite



(d) Signal binaural à 90° à gauche



(e) Signal binaural à 90° à droite

FIGURE III.50 – IACC et ITD calculés sur les signaux arrivant sur les oreilles du mannequin, correspondant à une restitution binaurale sur les deux haut-parleurs avant d'un véhicule de type Safrane

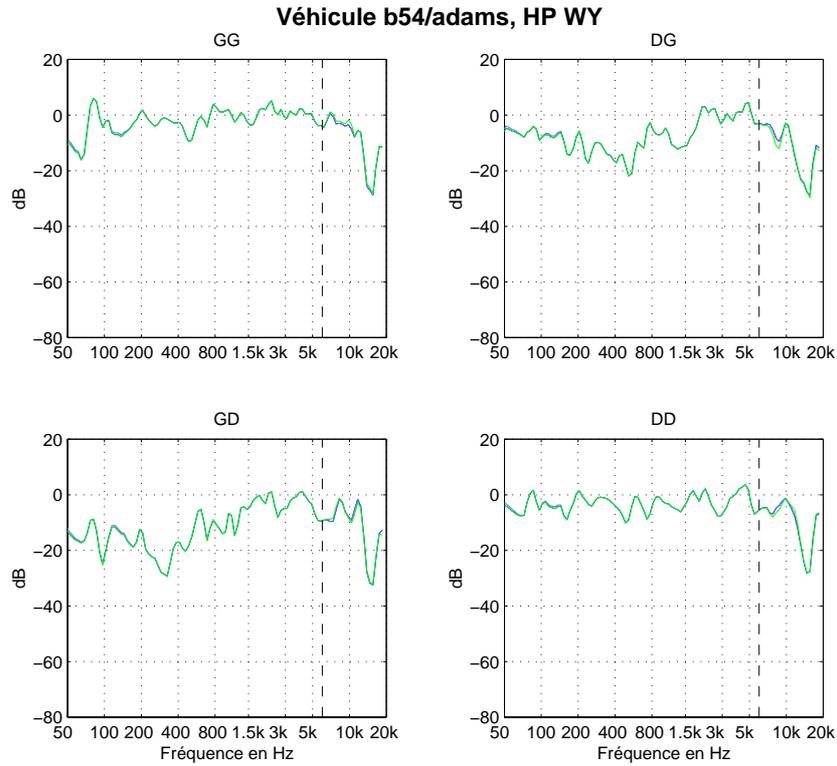


FIGURE III.51 – *Idem figure III.47, mais pour une position différente du mannequin acoustique*

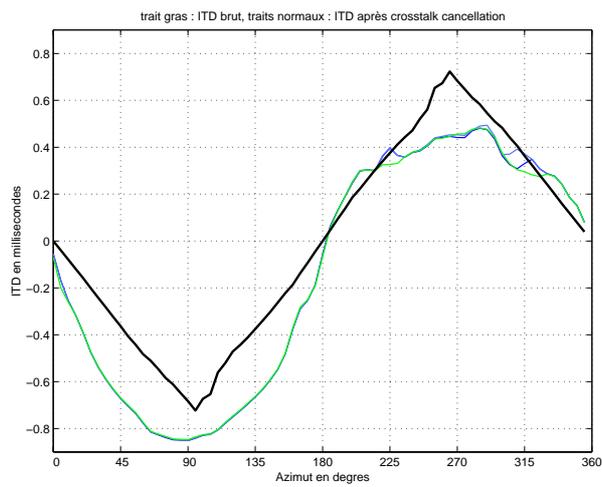


FIGURE III.52 – *Idem figure III.49, mais pour une position différente du mannequin acoustique*

erreur sur la localisation d'une source. Cette erreur aura des conséquences d'autant plus importantes que la source est fixe. Car dans le cas d'une source en mouvement, celui-ci sera correctement perçu, tout au moins si la source évolue entre -30° et 225° .

Il est nécessaire, à ce stade, d'envisager un test d'écoute plus rigoureux, à plus large échelle, ayant pour but de quantifier les erreurs de localisation, et donc d'affiner les paramètres de réglage du système transaural.

III.3.5.3 Propositions d'amélioration

Il existe plusieurs limitations à notre application : le coût de calcul est important, la zone d'écoute est réduite, et dans une plus large mesure le système n'est valable que pour un seul auditeur.

Diminution du coût de calcul L'implantation d'un des deux systèmes transauraux (sous forme classique ou sous forme shuffler) met en œuvre quatre filtres de longueur 4096 points. La convolution en temps réel sans retard d'un signal stéréophonique avec ces quatre filtres est assez coûteuse ; mais l'absence de retard n'étant pas une exigence, on peut utiliser un algorithme de convolution rapide basé sur la FFT (cf. § C.1.5).

On pourrait envisager de ne garder qu'un nombre plus restreint de points dans les réponses impulsionnelles binaurales, mais la performance du système irait en diminuant.

Il serait par contre intéressant de tester l'approche envisagée par Kahana [149], qui préconise de mesurer la matrice de transfert électroacoustique directe en chambre anéchoïque, avec les deux portières complètes et le siège dans une position identique à celle existant dans le véhicule. Les performances se rapprochent de celles atteintes en chambre sourde : une source virtuelle apparaît dans une direction unique. Même si les directions latérales restent problématiques du côté de la vitre, les directions à 30° à gauche et à droite sont tout-à-fait satisfaisantes. De plus, ces images ne semblent pas disparaître lorsque l'auditeur bouge légèrement la tête (translation ou rotation).

Les réponses impulsionnelles directes ne contenant pas la réverbération, l'inversion du déterminant de la matrice de fonctions de transfert directe est moins problématique ; une faible régularisation est suffisante. Mais il s'agit d'un compromis : ne sont inversées que les caractéristiques les plus importantes des réponses impulsionnelles ; le système fonctionne correctement grâce principalement aux temps d'arrivée et aux amplitudes des ondes sonores qui se rapprochent très fortement de celles obtenues en chambre sourde. Mais des erreurs sont inévitables, car la réverbération naturelle de l'habitacle est laissé inchangé, et les nombreuses réflexions sur les vitres et le toit dégradent la précision de la localisation de la source virtuelle.

Le coût de calcul est inévitablement réduit, puisqu'on peut se contenter de mesurer les réponses impulsionnelles directes sur 256 points, et de calculer les filtres inverses sur 512 points.

Une autre approche pour diminuer le nombre de points d'un filtre inverse, en dégradant de manière limitée la performance de l'inversion, consiste à utiliser la technique du warping temporel (cf. § C.1.9) sur le filtre RIF dont on veut calculer l'inverse [75, 165], en l'occurrence ici le dénominateur de la matrice de transfert directe. En effet, on peut montrer que l'information basses fréquences est conservée dans les premiers échantillons du filtre warpé, tandis que l'information hautes fréquences est rejetée dans les derniers échantillons. Le principe est alors de tronquer le filter warpé, ce qui permet de conserver la majeure partie de l'information en basses fréquences (une troncature à 10% est typiquement envisageable), avant d'effectuer l'inversion (cf. fig. III.53). Si l'on veut filtrer un signal par ce filtre warpé inverse, il est nécessaire d'utiliser une convolution avec warping inverse intégré, c.-à-d. une convolution dans laquelle le retard conventionnel z^{-1} est remplacé par un filtre passe-tout $(z^{-1} - r)/(1 - rz^{-1})$ (cf. fig. C.7), avec r le facteur de warping prenant la valeur opposée de celle ayant servi à calculer le filtre warpé.

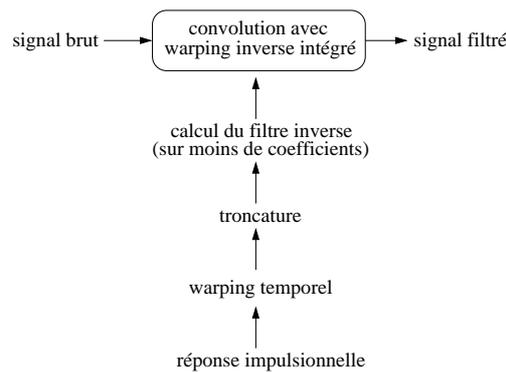


FIGURE III.53 – Principe de calcul et d'utilisation d'un filtre inverse dans le domaine warpé

Mais il faut noter que le warping inverse du retard de modélisation (cf. procédure de calcul d'un filtre inverse) entraîne une erreur sur la phase. Un simple retard de p échantillons devient un filtre passe-tout ayant p pôles à $z = r$ et p zéros à $p = 1/r$; en conséquence, le warping d'un retard pur devient une séquence infiniment longue, dont le retard de groupe a une forme exponentielle. La distorsion de la réponse en phase est importante en hautes fréquences lorsque r est positif (les hautes fréquences sont plus retardées que les basses), elle est importante en basses fréquences lorsque r est négatif (les basses fréquences sont plus retardées que les hautes). Cette distorsion de la réponse en phase des filtres peut être compensée par filtrage FIR dans le domaine non warpé avec la réponse impulsionnelle retournée du filtre correspondant au retard pur dans le domaine warpé.

Cependant, cette distorsion en phase étant la même pour chaque filtre du système transaural, elle ne trouble pas l'information temporelle relative contenue dans un signal binaural. D'après Kirkeby [165], la distorsion de la phase induite par le retour dans le domaine non warpé d'un retard de 16 ou 32 échantillons (correspondant à des filtres de longueur 32 ou 64 points) est parfaitement inaudible en pratique ; cela est en accord avec les études menées par Karjalainen [154] qui suggère qu'un retard de groupe inférieur à 5 ms est parfaitement acceptable. Dans notre cas, ce seuil devrait encore être plus élevé car le retard de groupe du filtre dans le domaine non warpé est une fonction très lisse.

Élargissement de la zone d'écoute Les performances de localisation se dégradent substantiellement lorsque l'auditeur n'est plus dans la zone d'écoute idéale, plus connue sous l'appellation *sweet spot* ; les dimensions de cette zone se mesurent en centimètres. Ce phénomène est bien connu dans le cas d'une restitution en chambre anéchoïque [15, 187, 235, 313], et il est encore plus critique lors de la présence de réflexions parasites [88, 186].

En chambre anéchoïque, on peut montrer que l'étendue de cette zone d'écoute optimale varie avec la fréquence et la géométrie du dispositif de restitution ; pour résumer (dans le cas d'un dispositif de restitution symétrique) :

- une réduction de l'angle entre les haut-parleurs et la tête de l'auditeur diminue le volume de la boule d'écoute pour les octaves graves (63, 125 et 250 Hz) ;
- pour les octaves moyennes (500 et 1000 Hz), l'angle n'affecte pas de manière significative le volume de la boule d'écoute optimale ;
- pour les octaves en hautes fréquences (2 et 4 kHz), de faibles angles (inférieurs à 15°) produisent une zone d'écoute plus large que celle obtenue avec des angles plus élevés (supérieurs à 30°) ;
- cependant, de faibles angles ne permettent pas de tirer profit de la séparation naturelle effectuée par le masquage de la tête en hautes fréquences.

Les expériences faites dans le cas d'un dispositif de restitution non symétrique, mais toujours en chambre anéchoïque, montrent que la robustesse aux mouvements latéraux est identique au cas d'un dispositif de restitution symétrique [273].

Et d'après les investigations effectuées par Kahana en véhicule [149], un déplacement latéral de deux centimètres suffit à provoquer des colorations dans les fonctions de transfert calculées au niveau des oreilles de l'auditeur ; ces colorations sont dues au décalage entre les RIB ayant servi à calculer le système transaural et les RIB effectives. Mais ceci n'a pas été vérifié par un test d'écoute, pas plus que la dégradation de la localisation (il est en effet possible que les colorations parasites constatées objectivement soient interprétées par l'oreille en terme de modifications de la localisation).

CONFIGURATION PARTICULIÈRE

Farina [16] a évalué l'implantation en véhicule d'un système transaural en configuration de dipôle stéréophonique [160] ; en effet, il est bien connu que cette configuration permet d'avoir une zone d'écoute d'autant plus large que les haut-parleurs de reproduction sont proches l'un de l'autre. Il a pour cela placé deux petits haut-parleurs face à la tête du conducteur, formant un angle de $\pm 5^\circ$ par rapport au plan médian de la tête. Il obtient bien évidemment des performances très importantes en terme de localisation, de qualité sonore, et de taille de la zone d'écoute. Mais une telle configuration est difficilement envisageable dans un véhicule de série, sauf si l'on utilise des haut-parleurs plats insérés dans la planche de bord. De telles structures ont d'ailleurs récemment fait leur apparition [7, 271].

AUGMENTATION DU NOMBRE DE CANAUX

Le système est surdéterminé si il y a plus d'équations que d'inconnues, c.-à-d. si il y a plus de points de restitution que de haut-parleurs. Dans ce cas, l'approximation au sens des moindres carrés peut seulement trouver l'erreur minimale où la solution exacte n'est pas atteignable. D'après Kahana [148], un tel système pourrait donner des résultats satisfaisants dans un cas parfaitement utopique, c.-à-d. avec les mêmes HRIR à la restitution, un environnement parfaitement anéchoïque, et des amplificateurs et haut-parleurs idéaux, et un dispositif de restitution symétrique. Mais pratiquement, un tel système ne peut donner de bons résultats.

Le système est pleinement déterminé si il y a autant de canaux de diffusion que de points de restitution. L'utilisation de plus de deux canaux de diffusion dans le véhicule permet de réaliser la simulation de sources virtuelles pour plus d'une position de l'auditeur, ou pour plusieurs auditeurs [147, 148, 184]. Le principe reste le même, sachant qu'il faut N haut-parleurs pour reproduire correctement une source virtuelle en N points de l'espace (ou $N/2$ positions de la tête d'un auditeur, celle-ci comportant deux oreilles). La matrice de transfert électroacoustique directe comporte alors $N \times N$ réponses impulsionnelles ; le calcul de l'inverse de son déterminant est plus complexe, et la matrice inverse comporte aussi $N \times N$ réponses impulsionnelles, d'où un coût de calcul très élevé pour une implantation en temps réel. Par exemple, avec quatre canaux on peut restituer correctement une source virtuelle pour deux positions de l'auditeur, typiquement le conducteur et le passager. Mais d'après les auteurs cités en début de paragraphe, le champ sonore est alors beaucoup plus complexe autour de chaque auditeur, et l'image de la source virtuelle est moins précise.

Toujours en utilisant des haut-parleurs plats, il est parfaitement envisageable d'implanter deux haut-parleurs supplémentaires au milieu de la planche de bord.

En poussant le raisonnement à l'extrême, il serait possible d'avoir un nombre de points d'écoute infinis en tapissant la surface intérieure du véhicule de haut-parleurs, c'est le principe de l'holophonie [237]. Pour des raisons pratiques et économiques, une telle approche est difficilement envisageable.

Le système est indéterminé si il y a moins d'inconnues que de variables, c.-à-d. si il y a moins de points de restitution que de haut-parleurs. Une solution unique peut être obtenue par l'ajout de contraintes supplémentaires [231]. Par exemple, il est tout-à-fait possible d'utiliser cinq haut-parleurs (plus exactement cinq canaux de reproduction) et de créer des sources virtuelles pour deux auditeurs, de manière

plus précise qu'avec quatre canaux.

SÉPARATION EN DEUX BANDES DE FRÉQUENCES

Une approche intéressante consiste à séparer le système transaural en deux bandes de fréquences [14]. Bauck part du principe que le volume de la boule idéale de restitution varie avec la fréquence (les hautes fréquences sont plus affectées que les basses fréquence, car le déplacement de l'auditeur correspond à une plus grande fraction de la longueur d'onde). L'idée consiste à séparer la voie gauche et la voie droite en deux bandes de fréquences (basse et haute), et à les diffuser par des haut-parleurs séparés (woofer et tweeter). Les tweeters sont placés proches l'un de l'autre (typiquement $\pm 3^\circ$), et comme ils reçoivent un signal pauvre en basses fréquences, ils ne sont pas sujets aux niveaux importants qui peuvent survenir dans cette configuration en basses fréquences ; et comme ils sont proches, ils génèrent une boule d'écoute idéale étendue en hautes fréquences. Les woofers sont au contraire placés loin l'un de l'autre (typiquement $\pm 20^\circ$), et comme ils reçoivent un signal pauvre en hautes fréquences, ils n'imposent pas de boule d'écoute réduite en hautes fréquences. La fréquence de séparation a été choisie par Bauck égale à 1 kHz. Tous les haut-parleurs sont supposés être alignés (physiquement ou électroniquement par ajout de retards).

Le système a été testé pour un dispositif de restitution symétrique. Sur la figure III.23(a), les fonctions de transfert S et A deviennent

$$\begin{cases} S &= C_w S_w + C_t S_t \\ A &= C_w A_t + C_t A_t \end{cases}, \quad (\text{III.40})$$

avec l'indice w correspondant au woofer, l'indice t correspondant au tweeter, et C un filtre de séparation en deux bandes de fréquences adjacentes.

Les résultats montrent que ce système permet de cumuler les avantages des systèmes basés soit sur deux haut-parleurs proches soit sur deux haut-parleurs éloignés, sans les inconvénients de chacun ; le volume de la boule d'écoute est très important.

Une version plus sophistiquée de ce concept pourrait être tentée en extrapolant à un réseau de haut-parleurs multibandes pour approximer des motifs de rayonnement [59].

Ce système pourrait être testé en véhicule, car le signal acoustique diffusé par les haut-parleurs avant est en fait scindé en une partie basses fréquences diffusée par un haut-parleur situé traditionnellement en portière, et une partie hautes fréquences diffusée par un haut-parleur situé en planche de bord ou en bandeau de portières. On pourrait même être tenté de spécifier deux systèmes transauraux, à savoir un par bande de fréquences. Si on pose comme postulat que tout signal peut être décomposé exactement en une partie basses fréquences et une partie hautes fréquences, les signaux parvenant aux oreilles se recomposent après passage dans chacun des deux systèmes transaural ; en pratique, cela n'est pas vrai car dans le cas d'une écoute en véhicule, on ne réalise pas une inversion *exacte* du déterminant de la matrice de transfert électroacoustique directe, toujours à cause des composantes à phase non minimale.

APPROCHE ADAPTATIVE

Ce principe a surtout été utilisé pour l'égalisation de la fonction de transfert d'une salle en de multiples points [71]. Il consiste à ajuster les coefficients du filtre afin de minimiser la somme des carrés des erreurs entre la réponse égalisée en de nombreux points dans la salle et une version retardée du signal original. Le filtre d'égalisation permet d'obtenir une réponse en fréquence plus uniforme dans un plus grand volume que si le filtre avait été spécifié en un seul point. Elliott présente une généralisation multi-canaux de l'algorithme LMS filtré [318] (pour un filtre inverse RIF), et de l'algorithme de Eriksson [73] (pour un filtre inverse RII).

Une application de ce principe à l'élargissement de la zone d'écoute d'un système transaural peut

être trouvé dans [235].

MODÉLISATION ARMA

La modélisation ARMA d'une réponse impulsionnelle de salle revêt deux avantages : d'une part la réduction de l'ordre du filtre inverse, et d'autre part l'insensibilité aux mesures effectuées pour différentes positions de la source et du récepteur [23, 115, 116, 222, 224, 226]. Les pôles correspondent aux fréquences de résonances de la salle, et sont par conséquent indépendants des positions de la source et du récepteur.

Mourjopoulos a montré que le modèle tout-pôles (AR) est le plus adapté lorsque l'on cherche à satisfaire les deux conditions précédentes. De plus, pour une salle donnée, il existe un ordre optimal qui approxime au mieux la réponse impulsionnelle réelle. Mourjopoulos propose ensuite d'utiliser une quantification vectorielle de ces réponses impulsionnelles modélisées [227]. Cette technique permet d'atteindre une classification optimale d'un grand ensemble de réponses impulsionnelles en un petit nombre de groupes ayant chacun des propriétés représentatives ; un dictionnaire de fonctions de transfert peut être estimé pour chaque salle, et utilisé pour calculer les filtres d'égalisation. Ces filtres d'égalisation suppriment les pics dûs aux résonances (caractéristiques de la salle elle-même), et sont donc moins sensibles aux changements de position de la source ou du récepteur.

III.4 Augmentation de la sensation d'espace

III.4.1 Position du problème

L'impression d'espace a été décrite par Blauert [20]. Mais plusieurs termes ont été utilisés depuis, sans être réellement définis de manière distincte. L'analyse multidimensionnelle de tests d'écoute entrepris pour clarifier la situation a montré que l'impression d'espace comporte au moins deux composantes [215]. La première est la largeur apparente de la source, ou ASW (Auditory Source Width), qui est définie comme la largeur de l'image sonore fusionnée temporellement et spatialement avec l'image sonore du son direct. La seconde est l'enveloppement, ou LEV (Listener EnVelopment), qui rend compte du degré de plénitude de l'image sonore autour de l'auditeur, hormis l'image sonore composant la largeur apparente de la source. La figure III.54 illustre ces considérations.

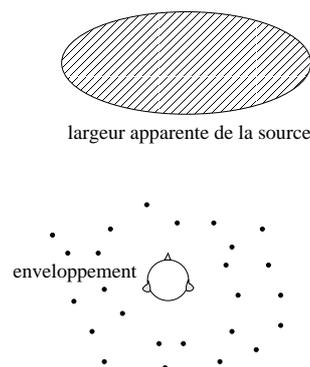


FIGURE III.54 – Concepts de largeur apparente de la source et d'enveloppement

III.4.1.1 Approche classique

L'approche classique ne prend en considération que les phénomènes acoustiques des lieux d'écoute.

Largeur apparente de la source Il est bien connu qu'un des critères objectifs reliés à la largeur apparente de la source (ASW) est la corrélation interaurale, ou IACC (Inter Aural Cross Correlation). Ce coefficient mesure l'intercorrélation des signaux parvenant aux deux oreilles ; en effet, les signaux parvenant aux deux oreilles ne sont toujours que partiellement corrélés entre eux, voire totalement incohérents, du fait des processus de distorsion subits au cours de leur trajet. Pour un signal en champ libre, les deux signaux parvenant à chaque oreille diffèrent par leur intensité et leur date d'arrivée, en fonction de l'angle d'incidence de la source (sauf si elle est dans le plan médian). Ce critère est basé sur la fonction de corrélation interaurale normalisée $\phi_{gd}(t)$ mesurée avec une tête artificielle, entre les temps d'intégration t_1 et t_2 , c.-à-d.

$$\phi_{gd}(\tau) = \frac{\int_{t_1}^{t_2} h_g(t)h_d(t + \tau) dt}{\sqrt{\int_{t_1}^{t_2} h_g^2(t) dt \int_{t_1}^{t_2} h_d^2(t) dt}}, \quad (\text{III.41})$$

avec $h_g(t)$ et $h_d(t)$ les signaux de pression arrivant aux oreilles gauche et droite. La valeur de IACC correspond au maximum de cette fonction de corrélation, c.-à-d.

$$\text{IACC} = \max_{|\tau| \leq 1 \text{ ms}} |\phi_{gd}(\tau)|. \quad (\text{III.42})$$

Compte tenu des dimensions moyennes de la tête d'un auditeur, la position du maximum n'exécède jamais 1 ms. Les maxima de la fonction $\phi_{gd}(t)$ traduisent l'existence de directions d'incidence privilégiées dans le champ sonore. On peut donc utiliser la mesure de la corrélation interaurale pour décrire la répartition directionnelle de l'énergie. Ce critère est toujours compris entre 0 et 1, et sa valeur est minimale lorsqu'il n'existe aucune direction prépondérante. Il permet donc d'évaluer le caractère diffus du champ. Il est traditionnellement évalué dans plusieurs bandes d'octaves. Pour une revue détaillée de ce critère, et pour son utilisation dans la caractérisation acoustique des salles de concert, le lecteur se reportera à [27, 52, 124, 125, 150, 289].

ASW est corrélé négativement avec IACC, c.-à-d. qu'une faible corrélation interaurale rend compte d'une largeur apparente de la source importante. Des études ont été menées pour vérifier si il existait une relation entre la direction de provenance des réflexions et la largeur apparente de la source [217] ; il ressort que la largeur apparente de la source perçue dans des champs sonores différents, mais avec le même taux de corrélation interaurale, est identique, quels que soient le nombre et la direction de provenance des réflexions. Cependant, dans le cas où le retard des premières réflexions est très faible, en l'occurrence inférieur à 4 ms (ce qui est le cas dans un habitacle de véhicule automobile), Morimoto montre que la largeur apparente de la source décroît lorsque ces réflexions proviennent d'une direction de plus en plus frontale [214].

D'autres auteurs ont étudié l'influence des premières réflexions latérales sur l'impression d'espace [11]. Cet effet est linéairement dépendant d'une mesure objective, la fraction d'énergie latérale précoce, ou LEF (Lateral Energy Fraction), correspondant au rapport entre l'énergie latérale et l'énergie totale, dans les 80 premières millisecondes, c.-à-d.

$$\text{LEF} = 10 \log_{10} \frac{\int_5^{80} p^2(t) \cos(\theta) dt}{\int_0^{80} p^2(t) dt}, \quad (\text{III.43})$$

où θ est l'angle entre la direction d'arrivée de la réflexion et l'axe imaginaire traversant les deux oreilles de l'auditeur. La limite de 80 ms après l'arrivée du son direct est fixée par Barron comme étant la frontière entre les événements précoces et les événements tardifs. Dans la pratique, on mesure l'efficacité latérale à l'aide de deux microphones, l'un ayant une directivité dipolaire, l'autre omnidirectionnel, c.-à-d.

$$\text{LEF} = 10 \log_{10} \frac{\int_5^{80} p_L^2(t) dt}{\int_0^{80} p^2(t) dt}, \quad (\text{III.44})$$

où $p_L(t)$ représente la pression mesurée avec le microphone dipolaire, et $p(t)$ représente la pression mesurée avec le microphone omnidirectionnel. L'inconvénient de cette représentation est qu'en procédant ainsi, on obtient une énergie latérale qui est pondérée par le carré d'un cosinus. Kleiner [168] propose donc une nouvelle méthode de calcul qui s'approche d'avantage du calcul théorique, c.-à-d.

$$\text{LEFC} = 10 \log_{10} \frac{\int_5^{80} p_L(t) \times p(t) dt}{\int_0^{80} p^2(t) dt}, \quad (\text{III.45})$$

où $p_L(t)$ représente la pression mesurée avec le microphone dipolaire, et $p(t)$ représente la pression mesurée avec le microphone omnidirectionnel.

Théoriquement, ce rapport vaut 0,5 lorsque les directions de provenance des réflexions sont uniformément réparties. Une valeur élevée de ce critère objectif dans les bandes d'octaves centrées entre 125 et 1000 Hz est considéré comme important pour l'enveloppement (LEV), les basses fréquences étant prépondérantes ; alors qu'une valeur élevée de ce critère dans les bandes d'octaves supérieures a une influence sur la largeur apparente de la source (ASW). De plus, l'impression d'espace est aussi fonction du niveau sonore global, et plus particulièrement du niveau des premières réflexions latérales.

Plusieurs études ont été menées pour synthétiser l'influence de la corrélation interaurale et la fraction d'énergie latérale précoce sur la largeur apparente de la source [27, 240]. Le facteur $[1 - \text{IACC}_{E3}]$, où IACC_{E3} représente la corrélation interaurale moyennée dans les trois bandes d'octaves centrées sur 500, 1000 et 2000 Hz, associé au facteur de force moyenné dans les octaves centrées sur 125 et 250 Hz semble être une bonne mesure de la largeur apparente de la source ; l'influence des réflexions en basses fréquences est rapporté par le facteur de force, tandis que l'influence des réflexions en hautes fréquences est rapporté par le facteur $[1 - \text{IACC}_{E3}]$. La fraction d'énergie latérale moyennée dans les quatre bandes d'octaves centrées sur 125, 250, 500 et 1000 Hz, semble être une bonne mesure du facteur $[1 - \text{IACC}_{E3}]$.

D'autre part, d'après les études de Blauert et Lindemann [21], les composantes hautes fréquences (supérieures à 3 kHz) des réflexions latérales produisent une augmentation de la largeur apparente de la source, tandis que les composantes basses fréquences (inférieures à 3 kHz) ont tendance à produire une augmentation de la profondeur apparente de la source.

Notons que les critères IACC et LEF peuvent présenter des variations non négligeables en fonction de la position dans la salle, alors que la largeur apparente de la source reste identique [61] ; ces variations sont d'autant plus fortes que les réflexions latérales sont importantes. La cause de ces variations est de nature interférentielle.

Enveloppement sonore D'après Morimoto [213], le rapport d'énergie avant/arrière affecte de manière significative l'enveloppement (LEV) : celui-ci augmente quand le rapport diminue. De plus, ce rapport énergétique n'a aucune influence sur la largeur apparente de la source (ASW) si la valeur du critère IACC est maintenue constante. Morimoto va même plus loin en examinant la contribution relative des premières réflexions et des réflexions tardives du rapport énergétique avant/arrière [218] ; le résultat le plus important de cette étude est que la contribution des réflexions précoces et tardives sur l'enveloppement est identique.

D'autre part, on constate que l'enveloppement augmente lorsque le rapport entre l'énergie du son direct et l'énergie du champ réverbéré augmente, de même lorsque le temps de réverbération augmente [10, 98, 118, 216]. Plus précisément, l'enveloppement (LEV) est relié aux niveaux, aux directions et aux temps d'arrivée des réflexions tardives ; il est même montré que lorsqu'on augmente la sensation d'enveloppement, la sensibilité aux changements de la largeur apparente de la source décroît [28]. Bradley [29] propose un critère objectif de mesure de l'enveloppement, appelé gain latéral, ou LG (Lateral Gain), correspondant au rapport entre le niveau de pression du champ sonore tardif capté par un microphone dipolaire $p_L(t)$ et le niveau de pression capté du champ sonore global capté par un microphone

omnidirectionnel en champ libre à une distance de 10 m avec la même source $p_A(t)$, c.-à-d.

$$LG = 10 \log_{10} \frac{\int_{80}^{\infty} p_L^2(t) dt}{\int_0^{\infty} p_A^2(t) dt}. \quad (\text{III.46})$$

Ce critère prend en compte à la fois le niveau global et la distribution spatio-temporelle de l'énergie tardive. Il est particulièrement efficace pour les octaves centrées sur 125, 250, 500 et 1000 Hz.

Une autre mesure de l'enveloppement peut être effectuée grâce au temps de décroissance précoce, ou EDT (Early Decay Time). Originellement, ce critère est défini comme le temps qu'il faut pour que l'énergie du son décroisse de 10 dB ; il est calculé par régression linéaire sur les 10 premiers dB de la décroissance énergétique. Griesinger [112] suggère de remplacer cette limite énergétique de 10 dB par une limite temporelle de 350 ms, qui serait plus précise pour des temps de réverbération inférieurs à 2 s. De plus, il suggère également de ne plus effectuer l'intégration rétrograde sur le signal de pression mesuré avec un microphone omnidirectionnel, mais sur la différence interaurale IAD ; il appelle ce nouveau critère décroissance latérale précoce, ou LEDT (Lateral Early Decay Time). Sa valeur est en général inférieure à celle de EDT, et plus la réverbération est frontale, plus elle est faible.

Pour résumer, on peut donc dire que :

- la largeur apparente de la source peut être contrôlée par des haut-parleurs frontaux (en d'autres termes, des haut-parleurs placés derrière l'auditeur n'ont aucune influence sur ce facteur perceptif) ;
- des haut-parleurs arrières sont par contre nécessaires pour produire une sensation d'enveloppement.

Griesinger [112] propose un autre critère permettant d'avoir une idée de la localisation de l'énergie réfléchiée dans une salle, en fonction du temps et de la fréquence : il s'agit de la différence interaurale, ou IAD (Inter Aural Difference). Ce critère est le rapport entre le carré de la différence égalisée des signaux gauche et droit mesurés avec une tête artificielle, et la somme des carrés des deux signaux individuels, c.-à-d.

$$IAD = 10 \log_{10} \frac{(G(t) - D(t))_{\text{ég}}^2}{G(t)^2 + D(t)^2}; \quad (\text{III.47})$$

l'égalisation de la différence des deux signaux gauche et droit est effectuée en dessous de 300 Hz et consiste en une rampe de 6 dB par octave (elle sert à compenser la réduction de la différence interaurale en champ diffus). Ce critère peut être calculé à partir de réponses impulsionnelles binaurales, mais aussi à partir de signaux binauraux enregistrés.

Tous ces critères ne sont pas applicables tels quels dans les petites salles en raison des bornes temporelles bien trop élevées ; il conviendrait, dans le cas de l'écoute en véhicule, de déterminer expérimentalement la limite temporelle entre les évènements précoces et les évènements tardifs.

III.4.1.2 Approche originale appliquée aux petites salles

Une autre approche, développée par Griesinger [107, 108, 110, 112], met l'accent sur la reproduction du son dans un environnement clos, par l'intermédiaire d'un dispositif électroacoustique. Griesinger ne prend pas seulement en compte les informations que l'on peut extraire de la réponse impulsionnelle acoustique, mais il étudie également le contenu du message sonore diffusé et met l'accent sur la séparation du message sonore en flux. Il s'intéresse de plus à la géométrie du dispositif de reproduction électroacoustique. Il a en outre mené des études dans les petites salles [109].

Les grandes salles sont capables de générer une sensation d'espace et d'enveloppement sur des signaux de parole ou de musique. Mais dans une petite salle, telle un studio d'écoute ou un habitacle de

véhicule automobile, cette perception est absente. La tâche de création d'une impression d'espace dans un environnement inhospitalier incombe donc à l'ingénieur du son (au moment de l'enregistrement) ou au concepteur du système audio (au moment de la reproduction). Bien que les phénomènes physiques et psychacoustiques qui sous-tendent ce processus ne fassent pas encore l'unanimité, on peut dire que la sensation d'espace est gouvernée par :

1. La détection de l'énergie des réflexions à travers les variations rapides de la différence de temps interaural, ou ITD (Interaural Time Delay), et de la différence de niveau interaural, ou ILD (Interaural Level Difference) ; ces fluctuations sont créées par les interférences entre le son direct et les multiples réflexions qui proviennent de directions différentes.
2. Un processus de ségrégation auditive qui sépare un son en plusieurs flux d'avant-plan et un flux d'arrière-plan. Les flux d'avant-plan sont composés d'événements sonores qui ont le même timbre, la même localisation, ou le même « sens ». Les sons qui ne sont pas identifiables en tant qu'événements précis, ou qui ne peuvent être groupés en un flux d'avant-plan, forment le flux d'arrière-plan. Ce sont les propriétés spatiales de ce flux d'arrière-plan qui donnent naissance à une perception d'espace et d'enveloppement.

Le processus de séparation décrit ci-dessus n'est pas toujours possible, par exemple avec un ensemble d'instruments à corde jouant lentement et *legato*. Dans ce cas, la perception d'espace est différente (et même plus faible) que si un flux d'arrière-plan avait pu être formé. Donc la perception d'espace dépend du type de musique, et de la manière dont elle a été jouée et enregistrée. Pour apprécier pleinement la perception d'espace, il est nécessaire qu'il y ait du vide entre les événements sonores (ce qui est le cas avec la parole). De plus, on peut montrer que même avec un enregistrement sonore comportant beaucoup d'intervalles blancs, la géométrie du réseau de haut-parleurs influence fortement ce que nous percevons. L'objectif est donc ici de maximiser les fluctuations de l'ITD et de l'ILD au niveau des oreilles de l'auditeur, particulièrement durant les portions de signal réverbéré présent sur l'enregistrement. Nous désirons maximiser ces fluctuations pour une zone d'écoute aussi large que possible.

Les multiples réflexions latérales interfèrent avec le son direct, et sont à l'origine des fluctuations de l'ITD et de l'ILD. L'amplitude des fluctuations est proportionnelle à l'énergie des réflexions. La fréquence des fluctuations dépend de plusieurs facteurs, notamment le vibrato du signal musical. Lorsque la fréquence des fluctuations est inférieure à 3 Hz, celles-ci sont perçues comme une modification de la position de la source ; des fluctuations plus rapides causent un élargissement apparent de la source, ou une perception d'une source plus étroite en présence d'un effet « surround ».

Lorsque le son direct est continu, et ne peut pas être séparé en différents événements sonores, l'énergie réfléchie et les fluctuations interaurales correspondantes donnent naissance à une sensation d'espace qui est relié au son direct. Cette sensation peut être complètement enveloppante. Cependant, lorsque le niveau sonore global de la source augmente, l'impression d'espace ne varie pas. C'est en variant le rapport entre le son direct et le champ réverbéré que l'on peut faire varier cette impression. Griesinger appelle cette sensation l'impression continue d'espace, ou CSI (Continuous Spatial Impression).

Lorsque le son direct peut être séparé en différents événements sonores, de sorte que l'on puisse former un flux d'avant-plan, la perception d'espace se sépare en deux, suivant le retard des réflexions. L'énergie réverbérée qui arrive pendant un événement sonore, et dans les 50 ms qui suivent la fin de cet événement, est perçue comme faisant partie de l'événement lui-même. Griesinger appelle cette sensation l'impression précoce d'espace, ou ESI (Early Spatial Impression). Comme CSI, ESI ne varie pas si le niveau global du son est modifié. Cependant, à la différence de CSI, ESI n'est d'ordinaire pas totalement enveloppant. ESI est perçue comme provenant de la même direction que la source sonore.

ESI est l'impression d'espace des petites salles. Dans ce cas, l'effet de salle est limité à la source, et il n'y a ni impression d'espace ni sensation d'enveloppement.

Quand les dimensions de la salle sont suffisamment larges pour qu'une quantité substantielle d'énergie arrive plus de 50 ms après l'arrêt du son direct, on perçoit une impression d'espace d'arrière plan, ou BSI (Background Spatial Impression). Si cette énergie est spatialement diffuse, alors les fluctuations de l'ILD et de l'ITD sont maximales, et résulte une forte sensation d'enveloppement. À la différence de ESI et CSI, la perception de BSI n'est pas limitée à la direction de provenance du son direct, et est indépendante du rapport entre l'énergie du son direct et l'énergie du champ réverbéré. BSI dépend fortement du niveau global de la source, et du niveau de l'énergie réverbérée autant que du temps de réverbération. Pour produire cette sensation d'espace, il est absolument nécessaire que le message sonore puisse être scindé en flux de plusieurs événements ; en effet, BSI ne se produira jamais à l'arrêt d'un seul événement sonore (ou à la fin d'un morceau de musique). BSI peut seulement être perçue dans les intervalles entre les événements sonores d'avant-plan.

Dans l'acoustique des petits espaces, comme dans les grandes salles, on se réfère principalement à la perception de BSI lorsque l'on parle de l'enveloppement. Nous sommes donc assez sensibles aux propriétés spatiales du son qui arrive 50 ms après la fin du son direct. En fait, notre sensibilité à BSI augmente pour atteindre un maximum aux alentours de 160 ms après la fin d'un événement d'avant-plan. Malheureusement, il y a très peu d'énergie dans une petite salle 160 ms après la fin d'un événement sonore ; dans un habitacle automobile, il n'y en a même plus après 50 ms ! Donc, si nous voulons recréer cette sensation d'espace, il faut l'ajouter sur l'enregistrement, ou par des traitements électroniques lors de la restitution.

En hautes fréquences, d'après les études menées par Griesinger, on peut dire que :

1. Pour avoir une sensation maximale d'espace, les composantes réverbérées d'un enregistrement doivent être totalement décorrélées.
2. Dans le cas d'une reproduction stéréophonique, la bande de fréquence correspondant à la sensation maximale d'espace varie avec l'angle entre les deux haut-parleurs ; plus l'angle est élevé, plus basse est la première bande de fréquences générant une sensation d'espace, et plus élevée est la sensation d'espace.
3. Idéalement, les composantes réverbérées d'un enregistrement devraient être reproduites par un réseau non corrélé de haut-parleurs entourant l'auditeur.

Une paire de haut-parleurs dans une configuration stéréophonique classique est loin d'être idéale en hautes fréquences. Idéalement, le champ réverbéré doit être diffus. Même si les réflexions latérales peuvent diffuser l'énergie réverbérée d'un enregistrement, il est préférable d'utiliser un dispositif de restitution comportant de nombreux haut-parleurs.

En basses fréquences, les phénomènes sont plus compliqués. L'impression d'espace est toujours produite par les fluctuations de l'ITD, mais ces fluctuations ont une origine différente. En dessous de 500 Hz, le son est véhiculé par des ondes stationnaires, les modes propres. Or les modes propres d'une salle ont une phase constante. Supposons un seul mode propre : on ne peut pas mesurer un quelconque retard, mais juste une variation d'amplitude. Or le processus de localisation auditive en basses fréquences est presque exclusivement gouverné par le retard interaural. Si la phase est constante, il n'y a pas de retard, et l'onde stationnaire n'est pas localisable ; elle est perçue dans la tête.

Heureusement, dans une salle il y a de nombreux modes propres. C'est justement l'interaction entre les modes latéraux et les autres modes (verticaux, avant-arrières et transverses) qui donne naissance à un retard interaural. Mais comme chaque mode a une phase constante, il n'est pas évident qu'un décalage de phase puisse se produire. Cependant, lorsque le signal d'excitation et les modes propres de la salle ne sont pas exactement à la même fréquence, il résulte un décalage de phase, et celui-ci est constant dans la salle ; sa valeur dépend de la géométrie de la salle. C'est grâce à ce décalage que peut s'effectuer la localisation.

La sensation d'espace principalement ressentie est celle du signal audio lui-même : si la réverbération est diffusée par des canaux non corrélés, les retards interauraux produits par les modes propres de la salle

vont fluctuer, et une impression d'espace sera perçue. Cependant, dans la plupart des signaux musicaux, l'attaque et l'extinction sont plutôt lentes : l'impression d'espace correspond à CSI, et non à BSI (pour CSI, ce sont les fluctuations de l'ITD au sein du flux d'avant-plan qui ont une importance).

En général, ce que l'auditeur perçoit dépend des relations de phase du matériau musical, du type et de la position des haut-parleurs, et de l'acoustique de la salle d'écoute. La raison est toujours la même : tous ces facteurs déterminent la proportion de modes latéraux par rapport aux autres modes. L'impression d'espace en basses fréquences augmente lorsque la proportion de modes latéraux asymétriques augmente par rapport aux autres modes de la salle d'écoute. Des dispositifs acoustiques peuvent être mis en œuvre pour satisfaire cette condition [109]. Pour résumer, la configuration optimale correspond à un système stéréophonique sur deux haut-parleurs pleine bande formant un dipôle, dans une salle dont l'arrière et le plafond sont absorbants, et avec un dispositif électronique permettant d'élargir la largeur de l'image en basses fréquences [105].

Globalement, la sensation d'enveloppement sonore est convaincante et naturelle aussi bien en basses qu'en hautes fréquences si le dispositif de reproduction électroacoustique comporte plusieurs haut-parleurs diffusant des signaux non corrélés.

Enfin, Griesinger introduit un critère objectif de mesure de l'enveloppement dans les petites salles [112] en basses fréquences : il s'agit de la fonction de transfert diffuse, ou DFTF (Diffuse Field Transfer Function). Elle est basée sur le fait que si une petite salle est capable de reproduire une sensation d'enveloppement sonore, les fluctuations de l'ITD et de l'ILD doivent être transmises du matériau musical aux oreilles de l'auditeur. L'idée est donc de savoir comment la salle transmet ces informations, d'où l'origine du nom. Elle est calculée à partir de réponses impulsionnelles binaurales. Nous ne détaillerons pas ici le processus de calcul assez complexe. Une difficulté principale consiste à construire un détecteur robuste du retard interaural. D'autre part, le critère doit être calibré pour que :

1. Sa valeur soit faible pour une source monophonique.
2. Sa valeur soit faible pour un système de reproduction avec deux canaux en phase.
3. Sa valeur augmente quand au moins deux sources sonores reproduisent un signal qui comporte des fluctuations en amplitude et en phase.
4. Sa valeur calculée en chambre anéchoïque soit maximale lorsque les deux haut-parleurs sont disposés de chaque côté de l'auditeur, et décroisse vers zéro proportionnellement au sinus de l'angle entre les haut-parleurs et la direction frontale.

Ce critère a été testé avec un bruit rose filtré par bandes dans le cas d'une salle rectangulaire. Il permet de dresser des constatations importantes quant au retard et à la distribution temporelle et spatiale des réflexions précoces en basses fréquences.

En conclusion, l'impression d'enveloppement sonore (LEV) doit être transférée de l'enregistrement à l'auditeur, ou ajoutée par traitements électroniques. Pour cela, le signal doit contenir des fluctuations en amplitude et en phase, conditions qui permettent au système auditif humain de créer cette sensation d'enveloppement sonore. Il est également nécessaire de créer une bonne séparation latérale en basses fréquences (il a été montré que cette séparation devait, dans certaines conditions, être effective au moins à partir de 60 Hz). En hautes fréquences, le système électroacoustique doit pouvoir développer une image jusqu'à des angles de $\pm 150^\circ$ derrière l'auditeur ; un tel système, qui maintiendrait en outre une bonne séparation latérale entre les sources arrières gauche et droite, est capable de fournir une sensation d'enveloppement importante, pour peu que le matériau musical contienne des hautes fréquences non corrélées dans ces deux canaux arrières.

III.4.2 Implantation en véhicule

Forts des considérations établies précédemment, nous avons testé plusieurs méthodes d'augmentation de la sensation d'espace en véhicule.

III.4.2.1 Ajout de premières réflexions latérales

Grâce au Spatialisateur® (cf. § E.1.1), nous avons pu facilement tester en véhicule l'effet perceptif obtenu en ajoutant des premières réflexions latérales en véhicule. Il est en effet possible de régler la distribution, le niveau et la direction de provenance de huit réflexions précoces.

Le paramétrage peut être effectué de manière perceptive, grâce au paramètre *envelopment* de l'interface de haut niveau (cf. fig. E.1) ; ce paramètre contrôle en fait le niveau relatif des sections temporelles *direct* et *early*. Il est explicitement relié à la largeur apparente de la source. Afin de se focaliser uniquement sur l'influence des premières réflexions, le gain des sections *cluster* et *reverb* peut être mis à zéro.

Le paramétrage des premières réflexions peut aussi être effectué par le biais d'une interface de bas niveau, beaucoup plus proche du traitement du signal. Il est alors possible de définir précisément les niveaux de chaque section temporelle *direct* et *early*, ou même plus en détail les temps d'arrivée, directions et niveaux des réflexions ajoutées. Mais le réglage s'effectue de manière déconnectée de toute considération perceptive.

Nous avons pu constater, à l'aide d'une écoute informelle en véhicule, que l'ajout de premières réflexions latérales à la direction de provenance de la source permet effectivement d'augmenter la largeur apparente de la source.

III.4.2.2 Retard et atténuation sur les haut-parleurs arrières

Suivant les expériences menées par Parizet [250], nous avons appliqué un retard et un gain uniquement sur les haut-parleurs arrières. Grâce au logiciel Max/MSP®, nous avons pu faire varier en temps réel la valeur du retard et du gain, et écouter en même temps le résultat. Une écoute informelle a été effectuée dans plusieurs véhicules, avec différents types de musique. Une sensation de profondeur apparaît effectivement lorsque la valeur du retard est très importante (de l'ordre de 50 ms) ; mais l'effet ne nous est pas apparu très convaincant. Il mériterait cependant d'être quantifié par un test d'écoute approfondi, car Parizet avait constaté une augmentation radicale de la sensation d'enveloppement (mais les auditeurs étaient placés à l'arrière).

Nous avons calculé la corrélation interaurale à court terme lorsqu'une atténuation de 6 dB et un retard de 50 ms sont appliqués sur les haut-parleurs arrières (cf. § I.5.6). Le mannequin est placé en position conducteur, et les quatre haut-parleurs sont alimentés par le même signal. La comparaison des figures III.55(a) et III.55(b) montre que ces traitements permettent de diminuer la corrélation tardive principalement pour la bande d'octave centrée sur 250 Hz ; mais la chute de la corrélation dans les bandes supérieures est plus nuancée.

III.4.2.3 Augmentation du niveau en basses fréquences

Concernant l'installation d'un caisson de basses les avis divergent ; et spécialement pour la musique classique, il semblerait que les propriétés du champ sonore en dessous de 60 Hz soient sans rapport avec le jugement de la qualité sonore globale. Dans le cas d'une écoute en véhicule, ceci mériterait d'être quantifié par un test d'écoute effectué lorsque le véhicule roule, car il se produit un phénomène important de masquage en basses-fréquences. Mais en tout état de cause, plus l'extension en basses fréquences est importante, plus la sensation d'enveloppement sonore (LEV) est importante [21].

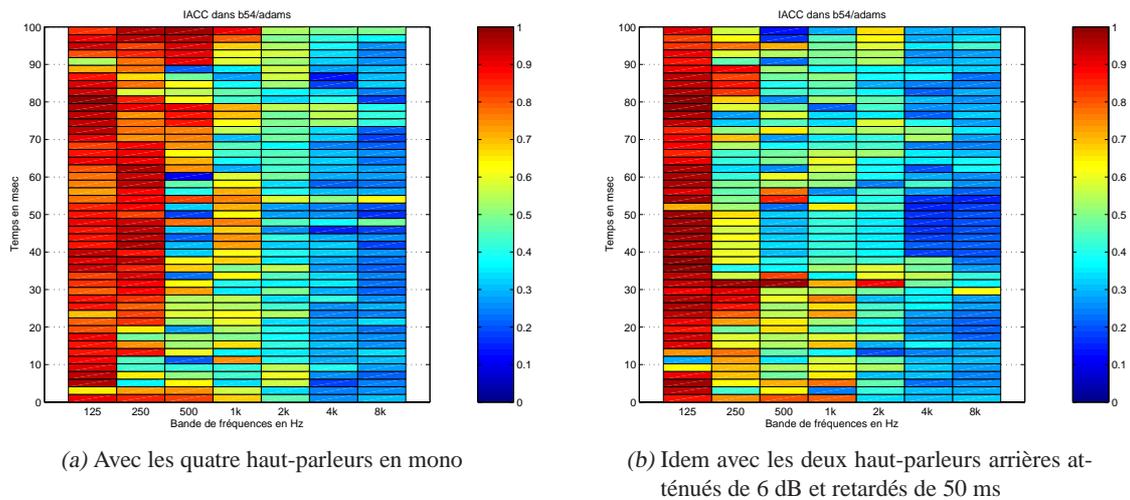


FIGURE III.55 – IACC à court terme calculé à partir de mesures réalisées dans un véhicule de type Safrane pour un signal monophonique

D'après Zacharov [331], le positionnement d'un caisson de basses n'est pas critique dans les petites salles. Mais en accord avec Griesinger [111], il est recommandé d'installer deux caissons de basses latéraux ; et un an plus tôt, Griesinger [109] déclarait que son positionnement dans un coin d'une petite salle rectangulaire produit plus d'enveloppement en basses fréquences que si il était placé au centre de la salle ; il en serait de même pour deux caissons de basses alimentés en stéréo ou même en mono. L'influence particulière d'un deuxième caisson de basses, permettant la présentation de signaux basses fréquences non corrélés, est étudiée par Martens [190].

Pour résumer, d'après Salava [276], il peut être établi que, en-dessous de la fréquence de Schroeder [287] (cf. § I.4.2.2) :

1. La fonction de transfert entre la source et l'auditeur dépend, outre les propriétés de la salle, toujours de la position de la source et du récepteur.
2. L'influence sonore des différents modes peut être positive ou négative, mais dans une petite salle il est primordial que le maximum de modes soient pleinement excités.
3. Dans une petite salle rectangulaire, tous les modes sont excités si une source monopolaire est située dans un coin.
4. Dans une petite salle, la position la moins avantageuse de l'auditeur pour les basses fréquences est le centre de la salle.

Nous n'avons pas effectué de tests avec un caisson de basses, mais simplement utilisé un filtre paramétrique pour amplifier le niveau des basses fréquences. Cet effet est assez inadapté avec un signal de musique classique, sauf éventuellement pour compenser le bruit du moteur et le bruit de roulement ; mais à ce sujet, il est nécessaire de mener une évaluation perceptuelle. Cet effet doit cependant rester assez discret, même avec de la musique pop.

III.4.2.4 Décorrélation des signaux alimentant les haut-parleurs

Diminuer voir annuler la corrélation de signaux audio peut produire des images sonores qui ont la largeur, la profondeur, et la sensation d'espace typiques des environnements naturels, sans le coût de calcul important d'un environnement de simulation acoustique, c.-à-d. d'une réverbération artificielle. L'absence de corrélation a au moins cinq effets sur la perception de l'image spatiale [156] :

1. Les modifications du timbre (coloration et filtrage en peigne) associées aux interférences constructives et destructives des multiples réflexions est perceptivement éliminée.

2. Des canaux non corrélés produisent un champ sonore diffus (semblable au champ réverbéré tardif d'une salle de concert), mais sans réverbération proprement dite.
3. Des canaux non corrélés produisent une excellente sensation d'extériorisation lors d'une écoute au casque.
4. La position de l'image sonore est identique même si la position de l'auditeur par rapport au dispositif de reproduction est modifiée.
5. L'effet de précédence (ou loi du premier front d'onde, appelé *effet de Haas*), qui cause un effondrement de l'image sonore dans le haut-parleur le plus proche, est annulé.

Les deux effets principaux qui nous concernent dans notre étude sont la production d'un champ sonore diffus, exempt de réverbération, et la disparition de l'effet de précédence. Le champ diffus sera perçu comme entourant l'auditeur, aucune direction de provenance n'étant définissable ; et la prédominance du haut-parleur le plus proche de l'auditeur sera annulée.

Mesure de la corrélation La corrélation entre deux signaux $x_1(t)$ et $x_2(t)$ peut être déterminée en calculant la fonction de corrélation croisée

$$R(\tau) = \lim_{T \rightarrow \infty} \frac{1}{2T} \int_{-T}^{+T} x_1(t)x_2(t + \tau) dt, \quad (\text{III.48})$$

où τ représente le décalage temporel entre $x_1(t)$ et $x_2(t)$. Pratiquement, c'est l'extremum de cette fonction qui donne la mesure de la corrélation, et en fait $-1 \leq R \leq +1$. On peut montrer que :

- si les deux signaux $x_1(t)$ et $x_2(t)$ sont identiques, alors $R = 1$;
- si les deux signaux $x_1(t)$ et $x_2(t)$ sont en opposition de phase, alors $R = -1$;
- si les deux signaux $x_1(t)$ et $x_2(t)$ sont très différents, ils sont dits non corrélés, et dans ce cas $R = 0$.

Il est même possible de contrôler l'impression de distance à la source sonore. D'après les études menées par Kurozumi [170], et confirmées par Wilde [320, 321], la distance apparente à la source est fortement liée à la valeur de la corrélation entre les deux canaux (ces études ont néanmoins été effectuées avec du bruit diffusé par deux haut-parleurs frontaux). La distance la plus faible est obtenue lorsque $R = -1$ (mais la largeur apparente de la source est plutôt faible), et la distance la plus importante est obtenue lorsque $R = +1$ (la largeur apparente de la source est également faible). Une distance moyenne est obtenue pour $R = 0$ (et la largeur apparente de la source est alors maximale).

Filtre passe-tout Le principe de l'annulation de la corrélation entre plusieurs canaux sonores consiste à appliquer sur chaque canal un filtre passe-tout. De cette manière, le timbre du signal de sortie est préservé sur chaque canal, car l'oreille n'est pas sensible à la phase [256] ; par contre, l'oreille est sensible aux variations de phase, et donc brouiller la phase entre les canaux a une influence très forte sur l'impression d'espace : l'oreille devient incapable de définir une direction de provenance de la source sonore.

Kendall [156] propose de définir un jeu de filtres RIF construits par TFD inverse, ce qui permet de contrôler la corrélation R des signaux de sortie. Les filtres sont ainsi définis dans le domaine spectral : leur module est uniforme sur toute la bande de fréquences (pour satisfaire le caractère passe-tout), et leur phase est aléatoire entre $-\pi$ et π . La corrélation résultante des signaux de sortie est directement liée à la corrélation de la phase de ces filtres. Leur réponse impulsionnelle est obtenue par TFD inverse.

La corrélation R des signaux de sortie est contrôlée par la spécification de la réponse en phase des filtres passe-tout. Par exemple, pour deux canaux :

- si on désire une corrélation $R = 0$, les réponses en phase des filtres passe-tout doivent être différentes ;
- si on désire une corrélation $R = 1$, les réponses en phase des filtres passe-tout doivent être identiques ;

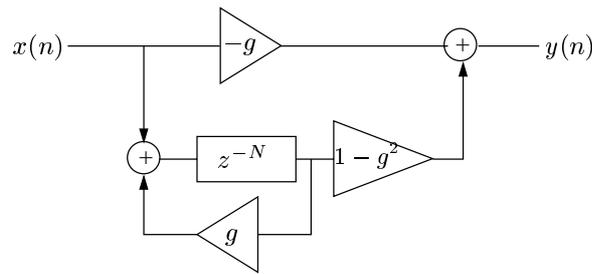


FIGURE III.56 – Structure récursive d'un filtre passe-tout

- si on désire une corrélation $0 < R < 1$, il faut pondérer la réponse en phase du premier filtre par un facteur k tel que $0 < k < 1$ et la sommer avec la réponse en phase du second filtre.

Si on désire une corrélation négative, le principe est le même, mais il faut ajouter π à la réponse en phase du second filtre.

Ce principe connaît néanmoins quelques limitations. En effet, si on effectue la TFD de la réponse impulsionnelle du filtre ainsi créé, on s'aperçoit que le module de la fonction de transfert, bien que globalement plat, comporte des variations entre les points de la TFD ; la neutralité du timbre peut donc être mise à mal. La solution consiste à augmenter le nombre de points de la TFD, mais cela augmente la longueur de la réponse impulsionnelle. D'une part, cela augmente le coût de calcul de l'opération de convolution, d'autre part, la durée de la réponse impulsionnelle ne doit pas excéder 20 ms, afin d'éviter l'étalement temporel qui perturbe les propriétés transitoires du signal source ; de plus, l'amplitude du décalage de phase sur les basses fréquences du signal d'entrée diminue avec le nombre de coefficients du filtre.

Une autre limitation liée au nombre de coefficients des filtres d'annulation de la corrélation concerne la précision de la corrélation calculée en sortie, par rapport à la corrélation spécifiée, spécialement pour les valeurs comprises entre -0,4 et +0,4. La solution préconisée par Kendall consiste simplement à recommencer le processus avec des réponses en phase calculées à partir d'autres séquences aléatoires.

Une approche alternative consiste à utiliser non plus des filtres passe-tout sous forme RIF, mais sous forme RII. Dans le plan complexe, les distances des pôles et des zéros au cercle unité sont contrôlées par une séquence aléatoire. L'ordre des filtres doit être élevé pour produire une corrélation proche de 0.

Il y a deux avantages à spécifier un filtre passe-tout sous forme récursive : d'une part, le coût de calcul est bien plus faible, d'autre part, ils permettent de réaliser facilement des variations dynamiques (les coefficients peuvent être continûment modifiés, ce qui change les distances des pôles et des zéros au cercle unité). Ces variations dynamiques produisent un effet semblable à celui d'un environnement sonore comportant des sources sonores ou des surfaces réfléchissantes en mouvement.

L'équation aux différences d'un filtre passe-tout est, dans le domaine temporel discret (cf. § C.1.3) [243]

$$y(n) = gy(n - N) - gx(n) + x(n - N), \quad (\text{III.49})$$

avec $x(n)$ le signal d'entrée, $y(n)$ le signal de sortie, N un retard en échantillons, et g un gain tel que $0 \leq g < 1$. Sa réponse impulsionnelle est donnée par

$$\begin{cases} h_{\text{pt}}(0) &= -g \\ h_{\text{pt}}(kN) &= g(1 - g^2) \end{cases} \quad \text{avec } k \in \mathbb{N}, \quad (\text{III.50})$$

ce qui correspond à la structure représentée figure III.56. La transformée en z de la réponse impulsionnelle est donnée par

$$H_{\text{pt}}(z) = \frac{-g + z^{-N}}{1 - gz^{-N}}, \quad (\text{III.51})$$

et sa réponse en fréquence s'écrit

$$H_{\text{pt}}(e^{2j\pi f}) = \frac{-g + e^{-2j\pi Nf}}{1 - ge^{-2j\pi Nf}}. \quad (\text{III.52})$$

On remarque que le numérateur et le dénominateur sont conjugués l'un de l'autre, et donc $|H_{\text{pt}}(e^{2j\pi f})| = 1$; de plus, les polynômes correspondant sont réciproques l'un de l'autre, ce qui explique que les racines soient les inverses de pôles.

Le retard correspond au délai qu'il faut attendre pour qu'un échantillon de sortie soit de nouveau injecté dans la sortie du filtre. Le gain permet de régler la proportion du signal d'entrée et du signal de sortie qui sera envoyée sur la sortie. Il est amusant de tester ces effets à l'oreille. Un retard court (quelques millisecondes) associé à un gain important (proche de 1) permet de « brouiller » la localisation du signal source ; un retard important (plusieurs centaines de millisecondes) associé à un gain faible (proche de 0) génère des échos.

Le coût de calcul d'une telle structure est de 3 opérations par échantillons, plus des opérations de stockage et de lecture en mémoire.

Comme avec la structure RIF, il est possible de garantir une certaine valeur de la corrélation. A partir de deux signaux $y_1(t)$ et $y_2(t)$ dont la corrélation vaut ρ , il faut générer les signaux

$$\begin{cases} y'_1(t) = (\alpha + \beta)y_1(t) + (\alpha - \beta)y_2(t) \\ y'_2(t) = (\alpha - \beta)y_1(t) + (\alpha + \beta)y_2(t) \end{cases} ; \quad (\text{III.53})$$

et l'on a

$$\begin{aligned} \begin{cases} \alpha = \frac{1}{2} \\ \beta = \frac{1}{2} \end{cases} &\Rightarrow \begin{cases} y'_1(t) = +y_1(t) \\ y'_2(t) = +y_2(t) \end{cases} && \text{et IACC} = \rho \\ \begin{cases} \alpha = \frac{\sqrt{2}}{2} \\ \beta = 0 \end{cases} &\Rightarrow \begin{cases} y'_1(t) = +\frac{\sqrt{2}}{2}y_1(t) + \frac{\sqrt{2}}{2}y_2(t) \\ y'_2(t) = +\frac{\sqrt{2}}{2}y_1(t) + \frac{\sqrt{2}}{2}y_2(t) \end{cases} && \text{et IACC} = 1 \\ \begin{cases} \alpha = 0 \\ \beta = -\frac{\sqrt{2}}{2} \end{cases} &\Rightarrow \begin{cases} y'_1(t) = -\frac{\sqrt{2}}{2}y_1(t) + \frac{\sqrt{2}}{2}y_2(t) \\ y'_2(t) = \frac{\sqrt{2}}{2}y_1(t) - \frac{\sqrt{2}}{2}y_2(t) \end{cases} && \text{et IACC} = -1 \end{aligned} \quad (\text{III.54})$$

Nous avons implanté ce système dans le logiciel Max/MSP®, à raison d'un filtre par haut-parleur. Un test informel a été effectué directement en véhicule, pour comparer trois paramétrages différents des quatre filtres passe-tout. La sensation d'enveloppement la plus importante est générée lorsque les paramètres de réglage des quatre filtres (c.-à-d. le gain et le retard) sont différents.

Remarquons que les valeurs optimales du retard et du gain de chaque filtre passe-tout doivent être réglées à l'oreille pour obtenir une impression d'espace maximale, sans toutefois rendre le son nasillard.

Nous avons également testé l'ajout de deux filtres uniquement sur les haut-parleurs arrières ; l'effet d'enveloppement est plus faible, mais cette configuration permet de garder une localisation du son vers l'avant.

Nous avons alors calculé la corrélation interaurale à court terme lorsque deux filtres passe-tout sont appliqués uniquement sur les haut-parleurs arrières (cf. fig. III.57). Le mannequin est placé en position conducteur, et les quatre haut-parleurs sont alimentés par le même signal. Le caractère diffus du champ tardif est cette fois beaucoup plus prononcé qu'avec l'adjonction d'un retard et d'une atténuation sur les haut-parleurs arrières (cf. fig. III.55).

L'utilisation de filtres passe-tout est en fait le premier pas vers l'ajout de réverbération. En effet, les filtres utilisés ont une réponse impulsionnelle dont la longueur, quoique faible, n'est pas réduite à un seul

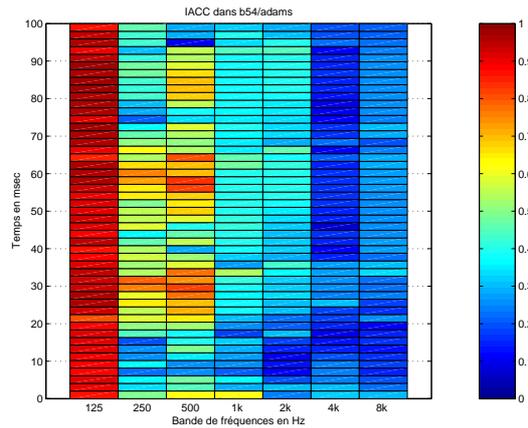


FIGURE III.57 – IACC à court terme calculé à partir de mesures réalisées dans un véhicule de type Safrane pour un signal monophonique, avec les deux haut-parleurs arrières décorrélés

échantillon. Et même si la convergence des coefficients vers zéro est rapide (cf. éq. (III.50)), il y a bien, au sens strict, ajout de réverbération. Ce principe est d'ailleurs mentionné par Jot [138]. Comportant de sérieuses limitations en terme de coût de calcul nécessaire à l'augmentation significative du temps de réverbération, mais aussi en terme de colorations parasites, il est préférable de faire appel à des réseaux de retards récursifs.

III.4.2.5 Ajout de réverbération

Toujours grâce au Spatialisateur[®], cet effet a pu facilement être testé en véhicule. L'avantage de ce logiciel est qu'il dispose d'une interface de commande reliée à des critères perceptifs (cf. fig. E.1). L'espace créé est totalement virtuel, mais le contrôle de l'effet perceptif produit est immédiat, car à la différence de dispositifs fonctionnant par convolution avec une réponse impulsionnelle mesurée, il n'est pas nécessaire de mettre à jour les coefficients d'un filtre.

Il est également possible de régler le processeur pour que notre perception de l'effet de salle créé soit celui d'une salle réelle, de laquelle nous possédons une réponse impulsionnelle. Le principe consiste à effectuer une analyse temps-fréquence de cette réponse impulsionnelle, et à calculer les paramètres de réglage temporels et fréquentiels du processeur [139, 310].

Même si l'ajout de réverbération est toujours très spectaculaire de prime abord, il ne nous semble pas opportun d'en exagérer les effets, car une certaine lassitude pourrait se produire.

III.5 Synthèse des différents traitements

Dans la première partie de cette étude sur l'amélioration des conditions d'écoute en véhicule, nous nous sommes intéressés à la correction fréquentielle. Ce traitement est destiné à corriger les résonances ou antirésonances assez marquées des fonctions de transfert de l'habitacle, il s'agit donc d'une correction du timbre. Nous avons montré que, bien que le support temporel de ces fonctions de transfert soit très ramassé, nous étions en mesure de distinguer le contenu spectral du son direct et celui de l'effet d'habitacle ; la perception de cette distinction s'appuie sur la distribution spatiale de l'effet d'habitacle. Nous avons donc choisi d'effectuer une correction différente pour le son direct (appliquée sur les haut-parleurs avant) et pour l'effet d'habitacle (appliquée sur tous les haut-parleurs). La mise en œuvre de ces corrections est effectuée à l'aide de filtres récursifs, peu coûteux et efficaces, à raison d'un filtre par haut-parleur : pour des filtres d'ordre 16, le nombre total d'opérations est de 132 opérations par échantillon, pour une erreur de modélisation inférieure à 2 dB.

Il est souhaitable que la correction soit permanente, quel que soit le type de signal diffusé : musique classique, jazz, pop ou parole.

Il convient néanmoins de définir une fonction de transfert « cible » pour le son direct et l'effet d'habitacle, grâce à des tests d'écoute auprès de la clientèle.

Cette correction est cependant calculée et optimisée pour *une* seule place dans *un* véhicule particulier ; c'est ce qui en fait sa principale limitation. Il semble réaliste de spécifier un ensemble de filtres par véhicule. Par contre il peut être plus délicat de spécifier un ensemble de filtres par position dans un véhicule (quelle place choisir au détriment de quelle autre ?) ; des techniques d'optimisation doivent alors être mises en œuvre.

Dans la deuxième partie de cette étude, nous avons voulu étudier les possibilités d'amélioration de l'image sonore. En effet, la position décentrée des passagers par rapport au système électroacoustique est source d'une dégradation importante de la perception de la scène sonore, principalement pour la musique classique et la parole.

Dans un premier temps, nous avons montré que l'ajout de retard et de gains sur les différents haut-parleurs permettait de recentrer l'image sonore, pour un coût de calcul dérisoire. Les valeurs de ces retards et gains sont cependant délicates à optimiser. Le principe est réellement efficace, spécialement pour un signal de radio ou de musique pop. Dans un deuxième temps, nous avons développé une technique permettant de créer deux sources sonores virtuelles de part et d'autre du plan médian de la tête d'un auditeur. Nous avons montré que cette technique doit être limitée en fréquence, afin de ne pas engendrer de phénomènes désagréables en hautes fréquences ; pour cette zone de fréquences, il est préférable de travailler sur l'énergie du signal. Cette deuxième approche est cependant beaucoup plus coûteuse en calcul, puisqu'il faut effectuer :

- quatre opérations de filtrage récursif avec des filtres d'ordre 12 (en fait deux filtres binauraux pour chaque source virtuelle), soit 100 opérations par échantillon ;
- quatre opérations de convolution avec des filtres dont le nombre de points N dépend de la précision de l'annulation des trajets croisés (typiquement $N = 4096$) ; si on accepte un certain retard entre l'entrée et la sortie, on peut effectuer la convolution par addition-recouvrement, ce qui nécessite trois opérations de FFT de taille $L = M + N - 1$ sur des blocs de signal de longueur $M \geq N$, soit $4 \times (4 \log_2(L) + 4) = 224$ opérations par échantillon pour les quatre filtres.

Comme pour la correction de timbre, ce système est optimisé pour une seule place dans un seul véhicule. Les mêmes remarques s'imposent donc.

La technique de création de sources virtuelles est limitée aux haut-parleurs avant. Pour un signal de parole, on pourra se contenter de générer une seule source sonore virtuelle face à l'auditeur (puisque un tel signal est toujours diffusé pour être perçu de manière frontale). Pour un signal de musique classique, la génération de deux sources virtuelles à $\pm 30^\circ$ par rapport au plan médian de la tête de l'auditeur permet de restaurer les conditions d'écoutes idéale de la stéréophonie, en accord avec le dispositif de prise de son ; il peut alors être opportun d'ajouter un certain effet de salle sur les haut-parleurs arrières, ou tout au moins de décorréler les signaux alimentant ces haut-parleurs.

Nous avons également montré qu'il était possible d'annuler l'effet d'habitacle. Cette annulation est théoriquement nécessaire lors de l'ajout de réverbération, bien qu'en pratique si l'effet de salle ajouté est largement supérieur à l'effet d'habitacle, le premier masque le second.

Dans la troisième partie de notre étude, nous nous sommes intéressés à l'impression d'espace. En effet, étant donné le petit volume d'un habitacle automobile, il est impossible de percevoir cette sensation. Nous avons montré qu'il fallait distinguer la largeur apparente de la source de l'enveloppement sonore.

La largeur apparente de la source peut être augmentée en ajoutant des premières réflexions latérales. L'enveloppement peut être augmenté en amplifiant le niveau des basses fréquences, en retardant les signaux alimentant les haut-parleurs arrière ou mieux en décorrélant les signaux envoyés aux différents haut-parleurs ; un effet de salle peut aussi être ajouté.

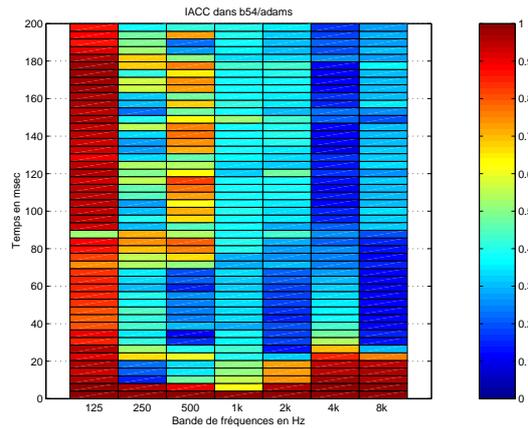


FIGURE III.58 – IACC à court terme calculé à partir de mesures réalisées dans un véhicule de type Safrane, comportant un système transaural sur les haut-parleurs avant, ainsi que deux filtres passe-tout et deux retards sur les haut-parleurs arrière

Ces techniques ne s'appliquent pas à un signal de radio, car il est perceptivement souhaitable que l'on ait la sensation d'une source parfaitement localisée et frontale.

L'augmentation de la largeur apparente de la source s'applique à tous les styles de musique, mais elle est plus efficace pour la musique classique ou le jazz. L'augmentation de l'enveloppement s'applique également à tous les styles de musique, mais la technique employée n'est pas la même. Avec de la musique pop, une amplification des basses fréquences semble communément appréciée ; la décorrélation des signaux alimentant les quatre haut-parleurs est envisageable, mais l'ajout d'un effet salle est catastrophique. Pour le jazz, on se limitera à la décorrélation. Pour la musique classique, l'amplification des basses fréquences est à rejeter (sauf de manière légère pour compenser le bruit du moteur et le bruit de roulement) ; la décorrélation des signaux alimentant les quatre haut-parleurs est également spectaculaire, mais en association avec la création de sources virtuelles à $\pm 30^\circ$ à l'aide des haut-parleurs avant, elle doit être limitée aux haut-parleurs arrière ; il en est de même pour la réverbération, qui doit, en outre, rester discrète.

Le coût de calcul nécessaire à la mise en œuvre de l'ajout de premières réflexions latérales est négligeable, puisqu'il ne fait intervenir qu'une opération de stockage/lecture en mémoire et une multiplication et addition par réflexion et par échantillon. L'implantation de filtres passe-tout récurrents ne nécessite que trois multiplications et additions par filtre et par échantillon. L'ajout de réverbération grâce à des réseaux de retards rebouclés est plus coûteux (au minimum huit cellules d'ordre deux pour les retards absorbants, soit 40 opérations par échantillon, et huit rebouclages, soit 24 opérations par échantillon).

Nous avons calculé la corrélation interaurale à court terme lorsqu'un système transaural est appliqué sur les deux haut-parleurs avant, et deux filtres passe-tout et deux retards sont appliqués sur les haut-parleurs arrière (cf. fig. III.58). Le mannequin est placé en position conducteur, un signal binaural correspondant à la HRIR à 0° est injecté à l'entrée du système transaural, et les haut-parleurs arrière sont alimentés en mono. Nous retrouvons ici une corrélation interaurale élevée sur les premières fenêtres (traduisant une localisation précise de la source), et une corrélation interaurale plutôt faible en champ tardif (traduisant une sensation d'espace) ; la corrélation reste cependant importante autour de 500 Hz.

L'objectif que nous nous étions fixé est donc atteint, puisqu'un tel système permet à un auditeur de localiser avec précision une source sonore virtuelle ; la synthèse de deux sources virtuelles à $\pm 30^\circ$ permet alors de recréer les conditions idéales d'une prise de son stéréophonique. D'autre part, ce système apporte une sensation d'enveloppement sonore importante, alors qu'elle est naturellement absente dans un petit volume comportant de nombreuses parois absorbantes.

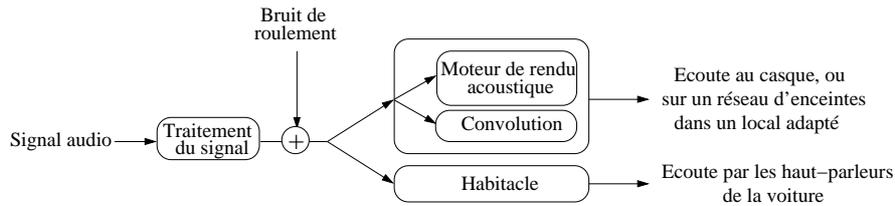


FIGURE III.59 – Schéma synoptique du dispositif complet de traitements et de restitution d'un signal audio dans un habitacle de véhicule automobile

III.6 Réalisation d'une plate-forme de tests

Toutes les opérations proposées dans cette étude ont été développées pour être testées en temps réel. Une plate-forme audio numérique a donc été développée à cette occasion. Elle est bâtie sur un ordinateur de type Macintosh® G4 exécutant le logiciel Max/MSP³ fonctionnant sous MacOS. La plate-forme effectue les traitements en temps réel et de manière purement logicielle, c.-à-d. sans carte électronique spécialement dédiée ; elle est donc complètement dépendante de la puissance du processeur hôte. Une fois les traitements effectués, une carte son délivre quatre signaux alimentant :

- directement en numérique, les quatre entrées du système de simulation ;
- après conversion numérique → analogique et amplification, les quatre sources sonores du véhicule (deux doubles à l'avant et deux à l'arrière).

Les signaux modifiés sont donc soit écoutés directement dans un véhicule, soit envoyés dans un dispositif de restitution fidèle de l'environnement acoustique d'un véhicule, pour être écoutés en laboratoire au moyen d'un casque d'écoute ou à travers un réseau d'enceintes.

III.6.1 Dispositif de traitements

Le schéma de principe du dispositif de traitements est exposé figure III.59, et l'interface de commande correspondante est présentée figure III.60.

Grâce à cette interface, il est possible :

- d'activer le système ;
- de choisir la source sonore et le bruit de roulement ;
- de rajouter le bruit de roulement en sortie du module de traitements, afin de donner un caractère plus réaliste à l'écoute d'un véhicule à l'arrêt ;
- de commander l'aiguillage du signal source vers l'habitacle du véhicule ou vers un dispositif de simulation (convolveur ou moteur de rendu acoustique) ;
- de mettre instantanément en (ou hors) service le module de traitements, ce qui donne la possibilité de basculer entre l'écoute du système audio brut et l'écoute du système modifié.

Outre un réglage de niveau global, il est également possible de choisir la source sonore et le bruit de roulement, et de régler indépendamment leur niveau.

La source sonore peut être composée d'un fichier son stocké sur le disque dur de l'ordinateur, mais il est également possible de connecter une source sonore externe ; tous les traitements étant effectués en numérique, il est nécessaire que la source externe le soit également, ou alors il faut prévoir un module de conversion analogique → numérique.

Le module de traitements est complètement paramétrable en temps réel. Les différents traitements peuvent être mis en fonctions les uns après les autres, comme le montre la figure III.61. Il comprend plusieurs étages représentés par des interrupteurs. La validation de ces interrupteurs peut donner lieu à

3. <http://www.cycling74.com/products/maxmsp.html>

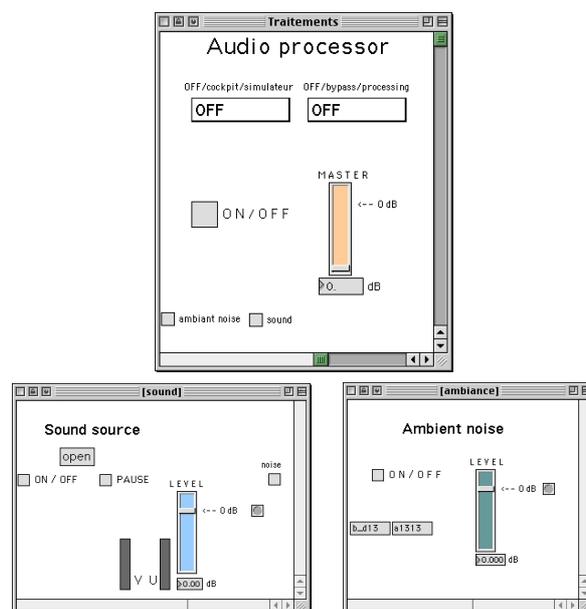


FIGURE III.60 – Interface générale de commande de la plate-forme de traitements et de restitution réalisée avec le logiciel Max/MSP®

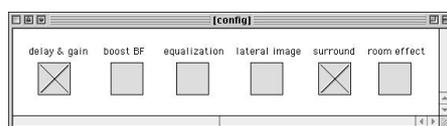


FIGURE III.61 – Interface de commande permettant de mettre en (ou hors) service les différents traitements proposés

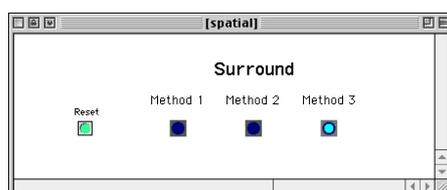


FIGURE III.62 – Interface permettant de tester différentes méthodes destinées à augmenter la sensation d'espace à l'intérieur du véhicule

l'ouverture d'une autre fenêtre permettant de choisir différentes méthodes de réalisation de l'effet recherché (sauf pour le recentrage de l'image sonore à l'aide de retards et de gains, et l'augmentation du niveau des basses fréquences, car les réglages sont figés). Le choix des différentes méthodes est toujours effectué de manière exclusive à l'aide de boutons radios (par exemple pour l'impression d'espace, cf. fig. III.62).

Pour la correction fréquentielle, il est possible de choisir entre :

- une égalisation réalisée à l'aide de filtres paramétriques réglables par l'utilisateur ;
- une égalisation conforme à celle décrite dans la première partie de ce chapitre, c.-à-d. séparée pour le son direct (provenant des haut-parleurs avant) et pour l'effet d'habitacle (diffusé par l'ensemble des haut-parleurs du véhicule).

Pour la correction de l'image sonore, il est possible de choisir entre :

- la classique balance gauche/droite et avant/arrière ;
- le PanPot d'intensité réalisé sur les quatre haut-parleurs ;
- la création de sources virtuelles à l'aide des techniques décrites dans la deuxième partie de ce chapitre.

Pour l'impression d'espace, il est possible de choisir entre :

- l'ajout de retards sur les haut-parleurs arrières ;
- l'ajout de premières réflexions latérales ;
- l'ajout de filtres passe-tout sur les haut-parleurs arrières.

Enfin l'ajout d'un effet de salle permet de choisir entre plusieurs configurations de réglage du Spatialisateur[®] correspondant à différentes salles de concert.

Pour chaque famille de traitements, nous avons privilégié une approche de complexité et d'efficacité croissante dans les différentes méthodes.

Cette plate-forme a fait l'objet d'une démonstration embarquée dans un véhicule à l'arrêt à l'occasion de la présentation des activités de la Direction de la Recherche de RENAULT. Il s'agissait de présenter l'activité de recherche en audio chez RENAULT, et de faire écouter l'amélioration apportée par les traitements électroniques. La démonstration est apparue satisfaisante, la commutation instantanée sans/avec traitements étant très démonstrative.

Mais il est nécessaire d'envisager la réalisation d'un test d'écoute plus rigoureux, afin de mieux cerner les attentes de la clientèle.

D'autre part, il faut réfléchir à l'implantation de ces traitements directement au sein de l'autoradio. L'ensemble des traitements, y compris l'ajout de réverbération, représente un coût de calcul relativement modique, de l'ordre de la centaine de MIPS, et par conséquent est à la portée de n'importe quel DSP actuellement en vente sur le marché.

Il est également souhaitable que les traitements proposés soient intégrés à l'ensemble des fonctions audio du véhicule. On pourrait par exemple diffuser le signal de l'autoradio de manière non corrélée sur les haut-parleurs arrière lorsque l'on établit une communication téléphonique avec un système « mains libres ». On pourrait aussi envisager de donner une localisation précise à certains messages d'alertes, afin de faciliter la prise de décision du conducteur.

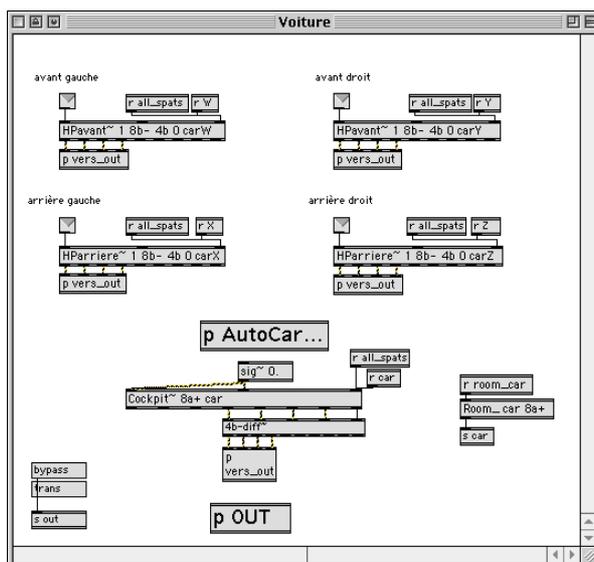


FIGURE III.63 – Implantation d'un système de modélisation de la restitution sonore à l'intérieur d'un véhicule à l'aide de la bibliothèque d'outils du Spatialisateur®

III.6.2 Dispositif de restitution

En accord avec le schéma présenté figure III.59, le dispositif de restitution sonore peut être :

- le véhicule lui-même ;
- un simulateur des conditions d'écoute en véhicule, c.-à-d. soit :
 - un dispositif permettant de réaliser une convolution en temps réel avec des réponses impulsionnelles mesurées dans l'habitacle ;
 - un moteur de rendu acoustique permettant de simuler la propagation acoustique à l'intérieur de l'habitacle.

Le système basé sur la convolution renvoie directement au deuxième chapitre de cette thèse, qui a conduit à la mise au point d'un dispositif de restitution sur casque fondé sur la synthèse binaurale.

À la suite de cette étude, des écoutes informelles sur enceintes, via un décodage transaural, ont montré des perspectives intéressantes pour mener des écoutes comparatives. Les principaux avantages observés, par rapport à l'écoute sur casque, sont une sensation plus naturelle de l'image spatiale (meilleure restitution frontale, tout en préservant la présence d'événements arrière) et, plus généralement, l'amélioration du réalisme de la reproduction (meilleures extériorisation et restitution du timbre en basses fréquences).

Le système basé sur un moteur de rendu acoustique est réalisé dans l'environnement Max/MSP® avec la bibliothèque d'objets du Spatialisateur® (cf. fig. III.63) ; à la différence du système basé sur la convolution, il est ici possible de manipuler en temps réel la localisation des sources sonores et l'effet de salle qui leur est associé. L'intérêt réside dans la possibilité de comparer une approche de simulation par convolution avec des réponses réelles mesurées, et une approche partant d'une modélisation de l'effet d'habitacle réalisées par des modules dérivés du Spatialisateur®. Dans ce cas, les paramètres de réglage sont issus d'une analyse temps-fréquence des réponses de l'habitacle. Certains modules ont été modifiés pour tenir compte des spécificités de la diffusion en habitacle, c'est le cas pour les sources sonores situées à l'avant de la cabine. Chaque source est dédoublée puisque, en général, l'habitacle est doté de haut-parleurs spécifiques et très éloignés pour les fréquences graves et aiguës. Le module gérant le son direct est ainsi divisé en deux sources ayant chacune une réponse fréquentielle et une direction de provenance particulière (cf. fig. III.64). La réverbération est commune à l'ensemble des sources.

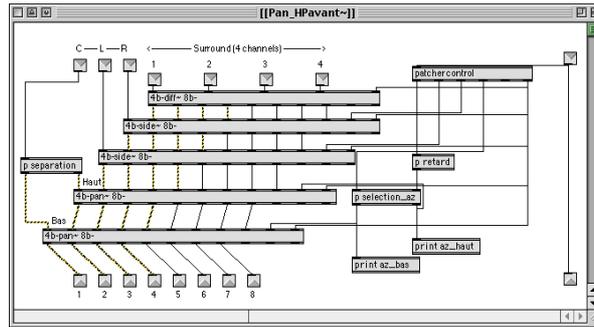


FIGURE III.64 – Module gérant le son direct d'un haut-parleur avant, permettant de synthétiser deux sources virtuelles dans deux bandes de fréquences et deux directions différentes

Plusieurs configurations de réverbérations ont été mises en œuvre (variation du nombre de canaux de rebouclage, et du nombre de sections temporelles).

L'architecture du simulateur se compose de quatre blocs gérant la localisation et la réponse en fréquence des sources sonores à créer, d'un bloc gérant la simulation de l'effet d'habitable, et d'un bloc gérant le format de sortie.

Deux dispositifs de restitution ont été testés : double transaural sur quatre haut-parleurs, et Ambisonic sur huit haut-parleurs. Plusieurs campagnes de réglage ont été effectuées, l'écoute se faisant dans un studio ; la référence était constituée par des fichiers sons convolués avec les réponses impulsionnelles de l'habitable, mesurées soit par un mannequin acoustique, soit par un micro SOUNDFIELD. Il ressort de ces essais que la diffusion à l'aide du système Ambisonic est assez loin de la référence ; en revanche, l'utilisation du double transaural est convaincante.

La restitution avec ce deuxième dispositif est moins fidèle que l'approche par convolution, car beaucoup plus difficile à configurer avec précision : d'une part, le dispositif comporte un certain nombre de simplifications quant à la description temps-fréquence-énergie (cf. § E.1.3), d'autre part, nous avons mis en évidence certaines limitations concernant la description de petits volumes (réglage de la distribution modale). Cependant, l'utilisation d'un tel outil offre la possibilité d'un contrôle plus étendu en terme de nombre, position, et réponse en fréquence des sources, ainsi qu'en terme de contrôle de l'effet d'habitable (variation du temps de réverbération et de la réponse en fréquence).

III.6.3 Couplage

Le couplage des deux dispositifs, tel que schématisé figure III.59, peut poser des problèmes de charge de calcul. En l'état actuel de la puissance de calcul disponible sur les ordinateurs personnels, nous conseillons de séparer la charge de calcul en deux : un ordinateur pour le dispositif de traitements, et un autre ordinateur pour le dispositif de restitution.

Le signal audio peut être chaîné sans problème sous format numérique (ADAT ou AES).

Quant au contrôle, il peut lui aussi être transmis d'une machine à l'autre grâce à des objets spécialement dédiées dans le programme Max/MSP® ou *jMax*®, la communication s'effectuant sous forme de signaux MIDI.

Si l'on choisit un dispositif de restitution par convolution, il est possible d'utiliser les cartes électroniques spécialement dédiées (système HURON CP4 de marque LAKE DSP) disponibles à la Direction de la Recherche de RENAULT.

Page blanche

Conclusion

La reproduction sonore à l'intérieur d'un habitacle de véhicule automobile comporte un certain nombre de distorsions, concernant principalement la modification de la réponse fréquentielle du système stéréophonique et la modification de sa réponse spatiale. Dans cette étude, nous avons abordé à la fois la caractérisation et l'optimisation de la restitution audio à l'intérieur d'un habitacle de véhicule automobile ; nous avons également développé une plate-forme d'auralisation de la qualité acoustique en habitacle, ainsi qu'un dispositif temps réel permettant de juger de la pertinence des traitements proposés.

La première partie du travail a été consacrée à différents aspects métrologiques. La mesure d'une réponse impulsionnelle acoustique en véhicule revêt des difficultés pratiques spécifiques. D'une part, nous désirions intégrer la fonction de transfert de l'amplificateur de l'autoradio dans la chaîne de mesure et, d'autre part, l'implantation de l'autoradio dans le tableau de bord ne permet pas d'accéder facilement aux connexions électriques. Nous avons donc envisagé de stocker les signaux de mesure sur un disque compact destiné à être inséré dans le lecteur de l'autoradio, et d'enregistrer la réponse du système sur un ordinateur autonome. Le découplage du système d'acquisition et de diffusion entraîne un décalage d'horloge entre le signal enregistré et le signal émis. Une étude a été consacrée à l'évaluation de la sensibilité à l'asynchronisme des horloges d'échantillonnage pour les principales techniques de mesures de réponses impulsionnelles, et une procédure de synchronisation automatique a été développée.

Les réponses impulsionnelles ont ensuite été analysées afin de mettre en évidence un ensemble d'indices objectifs permettant de caractériser objectivement le champ acoustique régnant à l'intérieur du véhicule. L'analyse objective a été principalement menée à partir de l'analyse temps-fréquence à partir de laquelle il est possible d'extraire des indices caractéristiques dont nous avons vérifié la pertinence perceptive. D'autre part, nous avons tenté de définir un protocole d'évaluation fiable de la répartition spatiale de l'énergie. Ces travaux de caractérisation nous ont permis de dégager des pistes intéressantes concernant les traitements à mettre en œuvre pour améliorer la qualité de la reproduction du signal audio à l'intérieur d'un véhicule.

Dans une seconde partie, nous avons mis en place une plate-forme de restitution en laboratoire ; un tel système permet, grâce à une écoute au casque, de reproduire fidèlement les conditions réelles d'écoute en véhicule, en étant libéré de son influence visuelle et en autorisant des comparaisons simultanées de différents véhicules ou traitements correctifs. Pour cela, nous avons travaillé sur une technique permettant d'utiliser les mesures de réponses impulsionnelles effectuées sur un mannequin acoustique, car il s'agit d'un dispositif qui permet d'obtenir des mesures reproductibles. Les caractéristiques acoustiques d'un mannequin étant assez éloignées de celles d'un individu, une restitution individuelle parfaite est impossible : celle-ci ne peut s'obtenir qu'en mesurant le sujet auditeur dans chacun des véhicules. Nous avons cependant travaillé sur la spécification de différents filtres de correction visant à minimiser ces

différences. Grâce à une étude objective et une analyse perceptive multidimensionnelle, nous avons pu déterminer la correction qui permet d'obtenir la restitution la plus fidèle possible. La principale conclusion est que la conduite optimale des tests d'écoute nécessite l'inversion numérique du casque d'écoute mesuré sur le mannequin ayant servi à relever les mesures de réponses impulsionnelles dans l'habitacle. En revanche, il ressort de l'étude qu'une égalisation individuelle par rapport au champ diffus n'apporte pas d'amélioration substantielle de la fidélité de restitution. L'égalisation proposée assure une garantie suffisante de fidélité pour l'évaluation sur écouteurs de la qualité de restitution dans l'habitacle et constitue une amélioration très nette par rapport à celle proposée par le fabricant, tant du point de vue objectif que perceptif.

La troisième partie a été consacrée au développement de différents axes de traitements électroniques permettant d'améliorer sensiblement la qualité de l'écoute embarquée. En effet, due à des caractéristiques spécifiques (petit volume, présence de parois absorbantes et réfléchissantes) et à de fortes contraintes (position de l'auditeur par rapport au système électroacoustique), la restitution d'un message audio dans un véhicule est loin d'être idéale ! Non seulement il est nécessaire de mettre en œuvre des traitements fréquentiels pour corriger les fortes aberrations spectrales, mais il faut également opérer dans le domaine temporel afin de corriger l'image sonore.

Compte tenu du support temporel très ramassé et du faible temps de réverbération d'une réponse impulsionnelle acoustique mesurée dans un habitacle, nous nous sommes interrogés sur la nécessité de distinguer le contenu fréquentiel du son direct de celui de l'effet d'habitacle. Nous avons pu répondre à cette question à l'aide d'un test d'écoute réalisé en laboratoire avec des réponses impulsionnelles synthétiques permettant de contrôler de manière fine la distribution temps-fréquence en énergie. Nous avons pu montrer que notre perception se fait suivant deux facteurs perceptifs indépendants liés à des indices objectifs se rapportant au son direct et à l'effet d'habitacle ; cette distinction s'appuie sur l'aspect spatial de l'effet d'habitacle. La correction de la réponse en fréquence des réponses impulsionnelles de l'habitacle peut alors être effectuée à l'aide de quatre filtres (un par haut-parleur) ; grâce à ces quatre filtres, il est possible d'agir à la fois sur le contenu spectral du son direct, mais aussi sur celui de l'effet d'habitacle. Cette correction est cependant spécifique à un véhicule donné, et à une position donnée de l'auditeur, puisque la correction de l'effet d'habitacle est calculée à partir de celle du son direct.

Concernant l'amélioration de l'image spatiale, nous avons principalement travaillé sur la réalisation d'un système transaural en véhicule. Cette méthode permet de créer une ou plusieurs sources sonores virtuelles autour de l'auditeur ; en outre, elle permet d'inverser le début de la réponse acoustique de l'habitacle. La difficulté est double : d'une part les haut-parleurs sont dans une position non symétrique par rapport à l'auditeur, d'autre part les réponses impulsionnelles binaurales ne sont pas mesurées en conditions anéchoïques. Dans cet esprit, nous avons travaillé sur l'inversion à phase totale de réponses impulsionnelles. Comme la résolution de l'équation intégrale de première espèce est un problème sans solution, il faut se contenter d'une approximation au sens des moindres carrés ; d'autre part, il faut également mettre en œuvre un principe de régularisation permettant de contrôler le processus d'inversion proprement dit, afin d'éviter l'apparition de résonances trop marquées dans le filtre inverse. Plusieurs techniques d'inversion et de régularisation ont été étudiées. La technique d'inversion la plus précise est l'inversion dans le domaine temporel, qui nécessite la résolution d'un système linéaire basé sur une matrice de Toeplitz symétrique pour lequel il existe un algorithme performant ; la technique de régularisation la plus efficace est l'utilisation d'un filtre défini dans le domaine fréquentiel qui permet de limiter la portée de l'inversion dans des zones spectrales bien définies. L'application d'un système transaural « étendu » en véhicule, adapté pour une configuration non symétrique du dispositif de reproduction et comprenant un effet d'habitacle, permet de recréer virtuellement les conditions d'écoute idéales de la stéréophonie classique, à savoir deux sources sonores à $\pm 30^\circ$ par rapport au plan médian de la tête de l'auditeur. Un résultat très satisfaisant peut être obtenu avec quatre filtres non récursifs de 4096 points ; cette correction est également spécifique à un véhicule donné et à une position donnée de l'auditeur.

Plusieurs propositions ont été formulées pour élargir la zone d'écoute (quantification vectorielle, filtrage adaptatif).

Nous avons enfin cherché à améliorer l'impression d'espace sonore. Nous avons pour cela exploré deux directions, à savoir l'augmentation de la largeur apparente de la source sonore, et l'augmentation de la sensation d'enveloppement de l'auditeur. La largeur apparente de la source peut être augmentée à l'aide de premières réflexions latérales. Quant à la sensation d'enveloppement sonore, elle peut être améliorée de manière spectaculaire en décorrélant les signaux alimentant les haut-parleurs ; elle n'est pas spécifique à une position donnée de l'auditeur. Cette décorrélation peut être réalisée de manière globale pour l'ensemble du message audio diffusé dans le véhicule, à l'aide de quatre filtres passe-tout (un par haut-parleur) ; mais on peut la limiter aux signaux diffusés par les haut-parleurs arrière, en l'associant à un système transaural sur les haut-parleurs avant. Cette deuxième solution permet d'obtenir la qualité d'écoute la plus satisfaisante, car elle combine une localisation précise de l'image sonore avec une sensation d'espace importante. Se dessine alors à l'horizon l'ajout d'une réverbération plus conséquente, afin d'immerger l'auditeur dans une ambiance de salle radicalement différente comme une salle de concert, par exemple.

Sur ces différents domaines, on peut résumer les principales contributions de cette thèse :

1. La mise en place d'une nouvelle méthode de mesure de réponses impulsionnelles, qui a pris corps par la création d'une procédure de mesure automatique ; cette procédure permet la détection de la dérive d'horloge d'échantillonnage, la conversion de fréquence d'échantillonnage, et la récupération des temps de propagation relatifs entre plusieurs sources mesurées de manière séquentielle.
2. La constitution d'une base de données de mesure de réponses impulsionnelles sur une dizaine de véhicules, avec trois capteurs différents, et pour quatre positions de ces capteurs dans chaque véhicule.
3. La mise en place d'une plate-forme de restitution en laboratoire, permettant de restituer au casque la même sensation d'écoute que dans un véhicule réel.
4. Une correction spectrale basée sur la répartition temps-fréquence-énergie des réponses impulsionnelles.
5. La réalisation d'un système transaural en véhicule permettant de restaurer les conditions d'écoute idéales de la stéréophonie.
6. L'augmentation de la sensation d'enveloppement sonore de l'auditeur grâce à une technique simple et peu coûteuse en calcul.
7. La livraison d'une plate-forme logicielle permettant de tester en temps réel dans un véhicule tous les traitements proposés.

Il est maintenant nécessaire de procéder à une validation perceptive à grande échelle des traitements proposés. Cette étude se doit d'être menée avec des spécialistes du marketing en vue de satisfaire les attentes des clients. Une fois les spécifications décidées, les traitements d'optimisation pourront alors être implantés directement dans l'autoradio. A l'exception du transaural, tous les traitements proposés sont peu coûteux en calcul, et peuvent être aujourd'hui même réalisés en temps réel et sans retard par n'importe quel processeur de signal ; la création de sources sonores virtuelles est plus coûteuse, mais le temps réel (avec retard) peut être maintenu grâce à une technique de filtrage rapide. Cela permet d'augmenter les possibilités de spatialisation de messages audio spécifiques, tels des signaux d'alerte, afin de leur conférer une identité particulière, et de faciliter les capacités de discrimination ; ces techniques ont déjà fait leur preuve dans le contexte aéronautique.

Certains points restent cependant à améliorer.

Le comportement de l'habitacle en basses fréquences mériterait d'être exploré plus en détail, afin de parfaire l'égalisation.

Nous n'avons pas proposé de nouveaux critères objectifs adaptés à la description de la qualité acoustique des habitacles de voiture. Considérant que la réalisation de tests d'écoute visant à caractériser de manière globale la qualité de la restitution sonore actuelle à l'intérieur des véhicules serait une opération longue et délicate, nous avons préféré nous contenter des déformations déjà recensées dans la littérature, ou connues de tous, et nous focaliser sur les traitements d'optimisation. Nous suggérons que la qualité sonore soit maintenant évaluée par le biais de tests d'écoute visant à juger de la pertinence auditive des différents traitements.

L'extension de la zone d'écoute optimale, dans le cas d'une correction spectrale ou d'une simulation de sources virtuelles, nous semble un axe de recherche également très important. Même si la majorité des véhicules en circulation ne comporte que le conducteur à bord, une optimisation de ces deux traitements pour un ensemble de places dans le véhicule pourrait être un argument de vente supplémentaire.

Enfin, nous n'avons pas évalué l'impact du bruit de roulement sur les différents traitements. En l'état actuel de nos investigations, il existe actuellement sur le marché des autoradios réalisant une correction automatique du volume en fonction de la vitesse du véhicule ; mais cette correction ne s'adapte pas au contenu spectral du bruit perçu dans l'habitacle. Il serait donc intéressant de voir dans quelle mesure une correction spectrale adaptative pourrait être mise en œuvre, afin de compenser le masquage fréquentiel en basses fréquences. En outre, il serait intéressant d'étudier l'influence du bruit de roulement sur la robustesse de la transauralisation, ainsi que sur la sensation d'enveloppement.

Nous ne saurions que trop insister sur la finalité d'une telle étude : améliorer les conditions de restitution du signal audio pour les occupants d'un véhicule. A ce titre, plusieurs éléments sont à prendre en considération.

D'abord, l'interface de commande se doit d'être aussi simple que possible ; un système trop complexe à utiliser n'est en réalité presque jamais utilisé. L'interface homme-machine pourrait se résumer à une simple distinction suivant le type de signal écouté ; ceci nous semble à la fois nécessaire, mais aussi largement suffisant.

Ensuite, si il est nécessaire de définir le degré d'optimisation recherché par la clientèle, rien n'empêche de proposer un système à la carte nettement plus performant, soit sur certains modèles haut de gamme soit pour des conducteurs plus exigeants.

Enfin, il faut tenir compte du fait que le conducteur d'un véhicule doit traiter un grand nombre d'informations nécessaires à la bonne marche du véhicule alors qu'il est en train d'écouter un signal audio. Le danger existe bel et bien qu'une sensation de confort extrême détourne sa vigilance de tâches nettement plus importantes. Il nous semble opportun d'étudier ce point avec l'aide d'ergonomes et de psychologues.

ANNEXE A

Matériel de mesure

A.1 Le microphone à directivité variable

Les microphones SOUNDFIELD sont construits en disposant quatre capteurs, dont la directivité est de type cardioïde, au centre des faces d'un tétraèdre (cf. fig. A.1). Les capteurs sont appelés *capsules* alors que le terme *tête* fait référence à l'assemblage des quatre capsules. Le modèle ST 250 utilise des capsules électrostatiques. La tête du microphone est disposée au sommet d'un boîtier cylindrique conte-

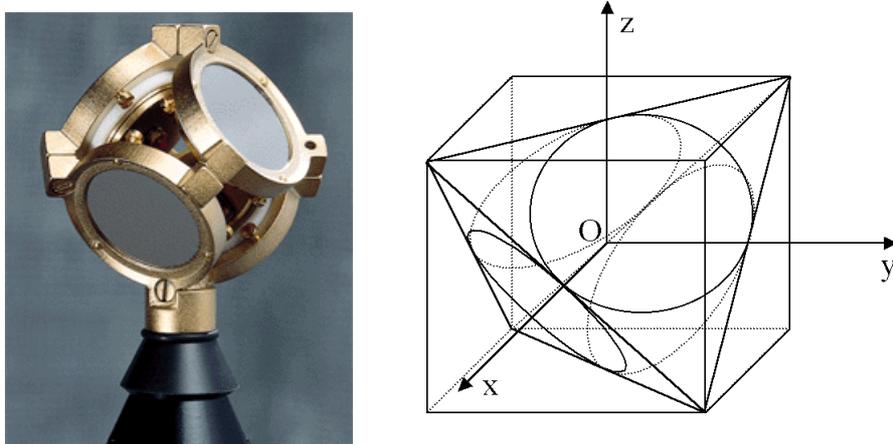


FIGURE A.1 – Disposition des capsules dans la tête du microphone ST 250

nant l'électronique de préamplification, et est protégée par une grille acoustiquement transparente.

Les signaux issus des quatre capsules constituent le « format A » ; celles-ci sont orientées régulièrement dans l'espace des directivités, et on peut les repérer par les couples

$$\begin{aligned}(\theta_1, \phi_1) &= \left(\frac{\pi}{4}, \frac{\pi}{4}\right) \\(\theta_2, \phi_2) &= \left(\frac{3\pi}{4}, \frac{3\pi}{4}\right) \\(\theta_3, \phi_3) &= \left(\frac{\pi}{4}, \frac{5\pi}{4}\right) \\(\theta_4, \phi_4) &= \left(\frac{3\pi}{4}, \frac{7\pi}{4}\right)\end{aligned} \quad \text{sur la sphère unité ;} \quad (\text{A.1})$$

une représentation schématique est donnée figure A.2. Les vecteurs ont pour coordonnées sphériques

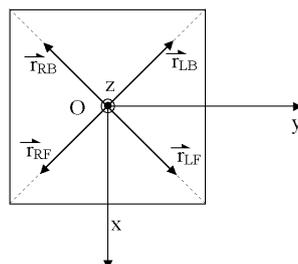


FIGURE A.2 – Orientation des capteurs du microphone SOUNDFIELD, symbolisées par quatre vecteurs unitaires

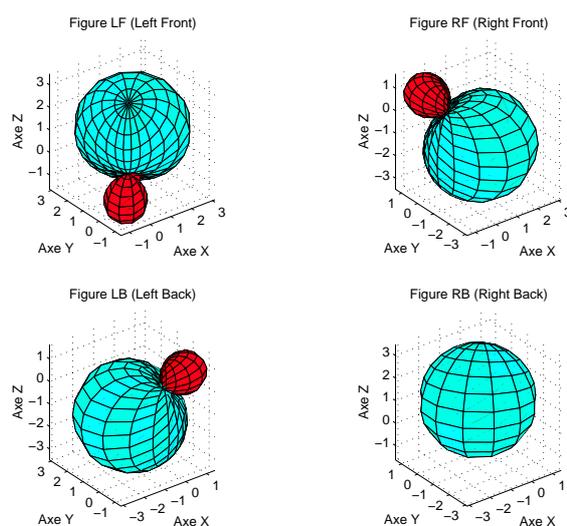


FIGURE A.3 – Figures de directivité idéales des capteurs du microphone SoundField (format A)

$$\begin{aligned}
 r_{LF} &= (1, \theta_1, \phi_1) \\
 r_{LB} &= (1, \theta_2, \phi_2) \\
 r_{RB} &= (1, \theta_3, \phi_3) \\
 r_{RF} &= (1, \theta_4, \phi_4)
 \end{aligned}
 \tag{A.2}$$

Les figures de directivités idéales des capsules sont présentées figure A.3.

Le microphone est fourni avec un transcodeur analogique réalisant la conversion des signaux issus des capsules en quatre signaux composés

- d'une directivité monopolaire (ou uniforme) ;
- de trois directivités dipolaires (ou en forme de huit) dans chacune des trois directions principales de l'espace.

Ces quatre signaux composent le « format B ». Le transcodage est réalisé en appliquant uniquement des gains et des inversions de phase selon les formules de transcodage idéales [133, 300]. Les directivités idéales sont présentées figure A.4.

Ce microphone a été mesuré par Laborie [173], qui constate que les figures de directivités des capsules s'approchent du modèle idéal dans une bande limitée de fréquences comprises entre 1500 et 5000 Hz. Une divergence en phase est également remarquée.

Les imperfections des figures de directivités ainsi que l'absence de coïncidence spatiale des capsules les unes par rapport aux autres ne sont pas pris en compte ; un modèle amélioré de transcodage est proposé par Laborie [173].

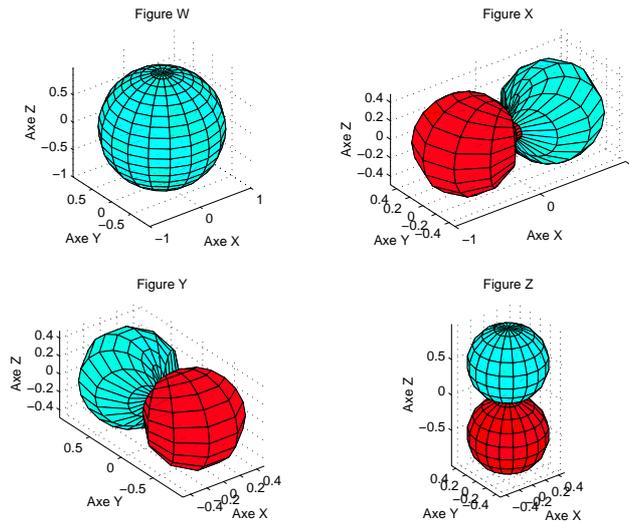


FIGURE A.4 – Figures de directivité idéales obtenues après transcodage (format B)

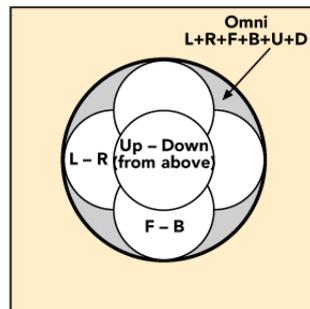


FIGURE A.5 – Les quatre figures de directivité utilisées lors d'une prise de son Ambisonic

Les quatre signaux du format B constituent les quatre canaux du système *Ambisonic*; celui-ci est apparu dans les années 80 comme une extension du principe de la stéréophonie pour atteindre une reproduction tridimensionnelle [95]. L'idée consistait à compléter une prise de son M-S coïncidente (c.-à-d. un microphone omnidirectionnel et un microphone bidirectionnel latéral gauche-droite) par l'ajout de microphones supplémentaires en s'assurant de l'absence de redondance d'information. Cette contrainte est satisfaite en ajoutant seulement deux microphones présentant des figures de directivité en huit et orientées respectivement selon les directions avant-arrière et haut-bas. La figure A.5 illustre le principe de la prise de son *Ambisonic* avec les quatre figures de directivité employées. Le signal obtenu au moyen d'une prise de son *Ambisonic* possède un réel sens physique, et réalise donc un encodage directif du champ sonore. Les signaux captés par les trois directivités en huit s'interprètent comme la projection de l'onde incidente sur chacune des trois directions principales de l'espace; le signal issu de la directivité uniforme représente la pression de l'onde incidente. En faisant référence à ce sens physique, on note $W(\theta, \phi)$ la figure de directivité uniforme, et $X(\theta, \phi)$, $Y(\theta, \phi)$ et $Z(\theta, \phi)$ les figures de directivité dipolaire. Les quatre signaux obtenus constituent le format B; les composantes *Ambisonic* associées à ces quatre figures de directivités sont notées w , z , y et x . La supériorité du système *Ambisonic* est la conséquence directe de la capture, de la manipulation et de la transmission des scènes sonores directement au format B, au lieu de travailler sur des signaux destinés à alimenter les haut-parleurs du système de restitution [173]. Un encodage spatial au format B est indépendant du système de restitution; il est susceptible d'être restitué sur un nombre quelconque de haut-parleurs disposés selon une configuration géométrique quelconque. L'adaptation au système de restitution est réalisé au moyen d'un décodeur dont le rôle est d'exploiter de manière optimale ce dispositif de restitution; la restitution sera d'autant meilleure que le nombre de

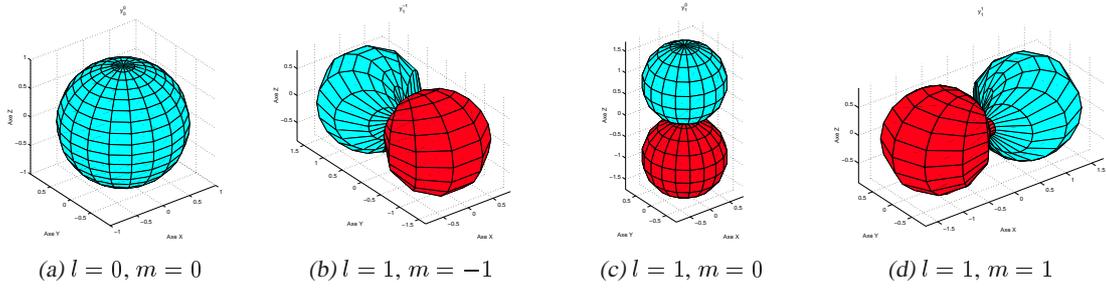


FIGURE A.6 – Représentation des quatre premières harmoniques sphériques réelles correspondant au développement jusque $l = 1$

haut-parleurs sera élevé.

Les figures de directivité associées aux composantes *Ambisonic* se déduisent de la base d'harmoniques sphériques réelles $\{y_l(\theta, \phi)\}_{lm}$ d'ordre 1 d'après les relations

$$\begin{cases} W(\theta, \phi) = y_0^0(\theta, \phi) \\ X(\theta, \phi) = \frac{1}{\sqrt{3}} y_1^1(\theta, \phi) \\ Y(\theta, \phi) = \frac{1}{\sqrt{3}} y_1^{-1}(\theta, \phi) \\ Z(\theta, \phi) = \frac{1}{\sqrt{3}} y_1^0(\theta, \phi) \end{cases} \quad \text{avec } y_l^m = \begin{cases} P_l^{|m|}(\cos \theta) \frac{\cos m\phi}{\sqrt{\pi}} & , m > 0 \\ P_l^0(\cos \theta) \frac{1}{\sqrt{2\pi}} & , m = 0 \\ P_l^{|m|}(\cos \theta) \frac{\sin m\phi}{\sqrt{\pi}} & , m < 0 \end{cases} \quad (\text{A.3})$$

$P_l^m(x)$ désigne le polynôme de Legendre associé, c.-à-d.

$$P_l^m(x) = \sqrt{\frac{2l+1}{2}} \sqrt{\frac{(l-m)!}{(l+m)!}} (1-x^2)^{\frac{m}{2}} \frac{d^m P_l(x)}{dx^m}, \quad (\text{A.4})$$

avec $P_l(x)$ le polynôme de Legendre dont l'expression est donnée par

$$P_l(x) = \frac{1}{2^l l!} \frac{d^l}{dx^l} (x^2 - 1)^l. \quad (\text{A.5})$$

Les fonctions de bases obtenues à l'ordre 0 et 1 sont représentées figure A.6. Pour une description complète sur le formalisme mathématique le lecteur se reportera à [1] [264, chap. 6]; on trouvera un bon résumé dans [173]. Plusieurs applications dans le domaine de l'acoustique sont données par [182, 220, 315].

La cohérence du champ sonore et l'absence de redondance d'information spatiale entre les différents canaux *Ambisonic* est garantie de manière théorique par les propriétés d'orthonormalités de la base des harmoniques sphériques. Cette étroite relation entre les composantes *Ambisonic* et la base des harmoniques sphériques permet d'envisager un encodage du champ sonore aux ordres supérieurs [54].

En effectuant une combinaison linéaire des différents canaux w , x , y et z , on peut créer n'importe quelle figure de directivité (omni, infra-cardioïde, cardioïde, hyper-cardioïde, figure huit). C'est en cela que nous faisons référence à un microphone à directivité *variable*.

ANNEXE B

Méthodes d'analyse statistique

B.1 Analyse statistique classique

B.1.1 La méthode des comparaisons par paires et le modèle de Bradley-Terry

La méthode des comparaisons par paires [58] permet de dégager un ordre de préférence concernant un attribut sémantique non symétrique, c.-à-d. mesurable sur une échelle continue et orientée. C'est *a priori* une méthode fiable lorsque les différences entre les stimuli sont faibles ; elle permet en outre de réduire les désaccords entre les sujets. Mais elle suppose que les comparaisons sont indépendantes entre elles ; dans le cas contraire, il est préférable d'appliquer une méthode directe d'extraction des rangs. D'autre part, des hypothèses plus restrictives (comme la transitivité) peuvent être posées concernant les probabilités de préférence des stimuli.

Le modèle nécessite la connaissance des probabilités de préférences de chaque stimulus, et génère pour chacun une mesure chiffrée appelée mérite. On suppose que ces valeurs sous-tendent le processus de jugement. Elles peuvent être positionnées sur une échelle linéaire, et leurs positions relatives indiquent la manière dont ont été jugés les stimuli : si deux stimuli ont des valeurs de mérite comparables, leur probabilités de préférences seront proche de 0.5, tandis que si les valeurs de mérite sont éloignées, cette probabilité sera proche de 1 (ou 0). De plus, les valeurs de mérite permettent d'établir des corrélations avec des indices objectifs mesurables.

Supposons que n sujets ($\gamma = 1, 2, \dots, n$) doivent comparer par paires t objets (A_1, A_2, \dots, A_t). Soit x_{ij} ($i, j = 1, 2, \dots, t, i \neq j$) une variable aléatoire prenant les valeurs 1 ou 0 selon que le sujet γ a préféré A_i ou A_j . On suppose que tous les x_{ij} sont indépendants, avec $x_{ij} + x_{ji} = 1$.

Soit $\pi_{ij} = \Pr\{x_{ij} = 1\}$ la probabilité que A_i soit préféré à A_j . Il est admis que $0 \leq \pi_{ij} \leq 1$ et que $\pi_{ij} = 1 - \pi_{ji}$. À ce stade, les objets peuvent être ordonnés en accord avec les probabilités de préférence moyenne

$$\pi_{i\cdot} = \frac{1}{t-1} \sum_{j \neq i} \pi_{ij}. \quad (\text{B.1})$$

Un autre ordonnancement est possible en introduisant l'hypothèse de transitivité pour chaque triplet de stimuli, c.-à-d.

$$\pi_{ij} \geq \frac{1}{2} \quad \text{et} \quad \pi_{jk} \geq \frac{1}{2} \quad \Rightarrow \quad \pi_{ik} \geq \frac{1}{2}; \quad (\text{B.2})$$

le classement se fait de manière triviale à l'intérieur de chaque triplet, et de proche en proche pour l'ensemble des stimuli.

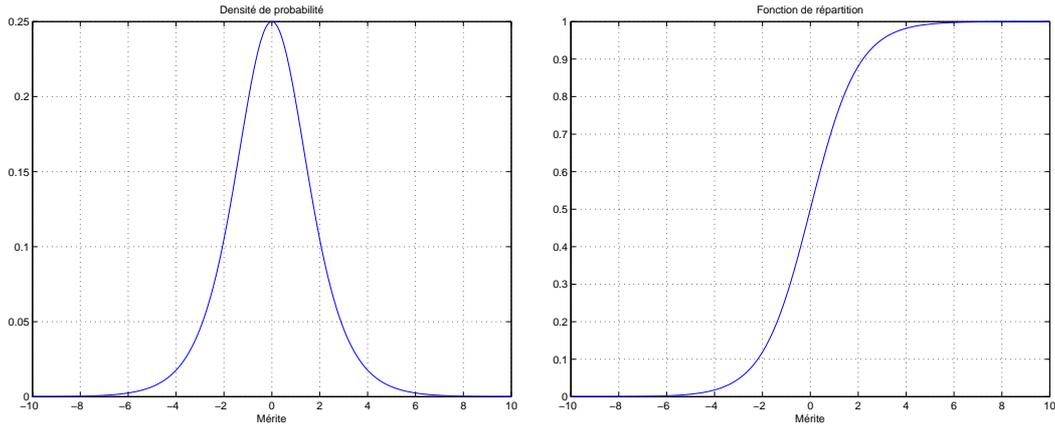


FIGURE B.1 – Densité de probabilité et fonction de répartition de la variable aléatoire z utilisées dans le modèle Bradley-Terry

Supposons que les objets A_i ($i = 1, 2, \dots, t$) aient chacun une véritable valeur ou « mérite » V_i quand on les note sur une caractéristique donnée. Ce mérite est représenté par une variable aléatoire continue y_i ($-\infty < y_i < +\infty$): A_i est préféré à A_j si $y_i > y_j$. Soit alors $z_i = y_i - V_i$, $i = 1, 2, \dots, n$. Si les z_i sont indépendants et identiquement distribués, $z_i - z_j$ aura la même distribution que $z_j - z_i$, ou en d'autres termes $z_i - z_j$ aura une distribution symétrique autour de 0. En écrivant que $H(x) = \Pr\{z_i - z_j < x\}$, il suit que

$$\begin{aligned}
 \pi_{ij} &= \Pr\{y_i - y_j > 0\} \\
 &= \Pr\{z_i - z_j > -(V_i - V_j)\} \\
 &= H(V_i - V_j) \\
 &= H(\delta_{ij}).
 \end{aligned} \tag{B.3}$$

Lorsque les probabilités de préférences sont exprimées à l'aide d'une fonction de répartition symétrique, on dit que les y_i satisfont un *modèle linéaire*. En effet, on peut exprimer les π_{ij} uniquement à l'aide de $t - 1$ différences indépendantes de V_i , avec $\pi_{ij} \geq \frac{1}{2}$ selon $V_i \geq V_j$. Les mérites des objets peuvent donc être représentés par t points sur une échelle *linéaire* muni d'une origine arbitraire.

Dans le modèle de Bradley-Terry, la probabilité π_{ij} est donnée par

$$H(V_i - V_j) = \frac{1}{4} \int_{-(V_i - V_j)}^{\infty} \frac{1}{\cosh^2(\frac{1}{2}x)} dx = \frac{1}{2} (1 + \tanh(\frac{V_i - V_j}{2})), \tag{B.4}$$

avec $V_i = \log \pi_i$, ce qui nécessite $\pi_i \geq 0$; en outre, on impose que $\sum \pi_i = 1$. Le lecteur déduira aisément la probabilité de préférence de l'objet A_i sur l'objet A_j

$$\pi_{ij} = \frac{\pi_i}{\pi_i + \pi_j}. \tag{B.5}$$

La densité de probabilité et la fonction de répartition H sont représentées figure B.1. La fonction de répartition est symétrique, tout comme la densité de probabilité, mais cette dernière n'est pas uniforme; au contraire, on favorise implicitement une estimation des valeurs de mérite à biais constant ($z_i = z_j$).

Les véritables valeurs de mérite (V) et les véritables probabilités (π), exposées ci-dessus, ne sont en fait obtenues que dans le cas d'une population infinie. Lorsque nous ne sommes pas dans ce cas limite, nous devons estimer ces quantités. Nous désignons par v les estimateurs des V , et par p les estimateurs des π .

Le but de cet algorithme est de calculer les estimateurs v_i des V_i . Soit $p_{ij} = \frac{a_{ij}}{n}$ la proportion de préférence de l'objet A_i sur l'objet A_j , avec a_{ij} le nombre de fois où A_i est préféré à A_j . On définit

alors d_{ij} par $H(d_{ij}) = p_{ij}$, avec $d_{ji} = -d_{ij}$ pour garantir la symétrie de la distribution. Comme il est en général impossible de satisfaire les relations $v_i - v_j = d_{ij}$, puisqu'il y a plus d'équations que d'inconnues, on choisit de les satisfaire en moyenne, en calculant

$$v_i = \frac{1}{t} \sum_{j \neq i} d_{ij}. \quad (\text{B.6})$$

Les v_i ainsi obtenus sont une estimation des V_i , au sens des moindres carrés. Avec la densité de probabilité calculée par le modèle de Bradley-Terry donnée en (B.4), on trouve que

$$p_{ij} \equiv H(d_{ij}) = \frac{1}{2} (1 + \tanh(\frac{1}{2} d_{ij})), \quad (\text{B.7})$$

et donc que

$$d_{ij} = \ln(p_{ij}/p_{ji}), \quad (\text{B.8})$$

d'où

$$v_i = \frac{1}{t} \ln \prod_{j \neq i} (a_{ij}/a_{ji}). \quad (\text{B.9})$$

Comme l'origine de l'échelle linéaire est arbitraire, il est commode de la choisir telle que

$$\sum_{i=1}^t V_i = 0, \quad (\text{B.10})$$

ce qui garantit l'unicité de l'estimateur v_i , mais implique de normaliser les π_i par leur somme pour que celle-ci soit égale à 1.

Les v_i peuvent alors être placés sur une échelle continue et orientée.

Un test de performance statistique permet de savoir si l'algorithme a réellement pu distinguer les stimuli : il s'agit de tester (et de rejeter) l'hypothèse selon laquelle les v_i sont identiques entre eux au moyen d'une loi statistique particulière possédant une distribution du χ^2 à $t-1$ degrés de libertés (cf. [58, p. 59]).

B.1.2 L'analyse de variance

Cette technique est destinée à apprécier l'effet de variables qualitatives sur une variable numérique, et revient dans le cas simple à comparer plusieurs moyennes d'échantillons gaussiens.

Supposons que N sujets ($n = 1, 2, \dots, N$) doivent comparer par paires t objets (A_1, A_2, \dots, A_t), soit $P = \frac{t(t-1)}{2}$ paires possibles. Soit δ_p^n ($p = 1, 2, \dots, P$, et $\delta \in [0, 10]$) une variable aléatoire représentant la mesure, par l'individu n , de la dissemblance entre les stimuli A_i et A_j constituant la paire p . Soit également $m_{suj} = \frac{1}{N} \sum_{n=1}^N \delta_p^n$ la moyenne des dissemblances entre les individus, $m_{sti} = \frac{1}{P} \sum_{p=1}^P \delta_p^n$ la moyenne des dissemblances entre les stimuli, et $m_{tot} = \frac{1}{NP} \sum_{p,n=1}^{P,N} \delta_p^n$ la moyenne totale.

Nous disposons maintenant d'un tableau à N lignes et P colonnes, et nous allons effectuer un test de confusion des P moyennes m_{suj} . En supposant que chaque N échantillons de valeurs de dissemblances pour une paire soit issu d'une variable aléatoire gaussienne de moyenne m_p et d'écart type σ , le problème est de trancher entre les deux hypothèses suivantes :

$$\begin{cases} H_0 & : m_1 = m_2 = \dots = m_p = m \\ H_1 & : \exists i, j m_i \neq m_j \end{cases} \quad (\text{B.11})$$

En remarquant que $\delta_p^n - m_{tot} = \delta_p^n - m_{suj} + m_{suj} - m_{tot}$, on peut facilement écrire

$$\frac{1}{NP} \sum_n \sum_p (\delta_p^n - m_{tot})^2 = \frac{1}{NP} \sum_n \sum_p (\delta_p^n - m_{suj})^2 + \frac{1}{P} \sum_p (m_{suj} - m_{tot})^2, \quad (\text{B.12})$$

formule qui n'est autre que celle de la variance totale décomposée en moyenne des variances et variance des moyennes. Si on pose $S^2 = \frac{1}{NP} \sum_n \sum_p (\delta_p^n - m_{tot})^2$, $S_A^2 = \frac{1}{P} \sum_p (m_{suj} - m_{tot})^2$ et $S_R^2 = \frac{1}{NP} \sum_n \sum_p (\delta_p^n - m_{suj})^2$, on peut récrire

$$S^2 = S_A^2 + S_R^2, \quad (\text{B.13})$$

formule dite « d'analyse de variance » ; S_A^2 représente la variance entre les stimuli, S_R^2 la variance résiduelle.

Si on écrit $S_R^2 = \frac{1}{P} \sum_{p=1}^P S_p^2$ avec $S_p^2 = \frac{1}{N} \sum_{n=1}^N (\delta_p^n - m_{suj})^2$ en introduisant les variances de chaque paire, on trouve que $\frac{NP S_R^2}{\sigma^2}$ est une variable χ^2 à $NP - P$ degrés de liberté car $\frac{NP S_R^2}{\sigma^2} = \sum_{p=1}^P \frac{N S_p^2}{\sigma^2}$ et $\frac{N S_p^2}{\sigma^2}$ est une variable χ_{N-1}^2 .

Sous l'hypothèse H_0 , les m_{suj} sont des variables de même loi, et dans ce cas $\frac{NP S^2}{\sigma^2}$ suit un χ_{NP-1}^2 car S^2 est la variance d'un échantillon de taille NP issu d'une gaussienne de moyenne m et d'écart type σ , et $\frac{NP S_A^2}{\sigma^2}$ suit un χ_{P-1}^2 car S_A^2 est analogue à la variance des P moyennes m_{suj} .

Or on peut écrire

$$\chi_{NP-1}^2 = \chi_{P-1}^2 + \chi_{NP-P}^2, \quad (\text{B.14})$$

ce qui équivaut au fait que S_R^2 et S_A^2 soient des variables aléatoires indépendantes (th. de Cochran) ; toujours si H_0 est vraie, on peut alors écrire

$$\frac{S_A^2/P - 1}{S_R^2/NP - P} = F(P - 1; NP - P), \quad (\text{B.15})$$

avec F la loi de Fisher. Le test statistique est alors le suivant : on forme le rapport $\frac{S_A^2/P - 1}{S_R^2/NP - P} = f$, et on calcule la probabilité de rejet de l'hypothèse nulle, à savoir

$$\text{Pr}\{0 \leq \phi \leq f\} = \int_0^f F(\phi) d\phi; \quad (\text{B.16})$$

si cette probabilité est supérieure à un seuil fixé à l'avance (typiquement 95%), alors on rejette l'hypothèse H_0 et on conclut à une influence significative des différences entre stimuli ; S_R^2 est alors un estimateur sans biais de σ^2 . Le rejet de H_0 ne signifie pas que les m_{suj} sont toutes différentes entre elles, et on peut chercher à tester l'égalité à 0 des différences $m_{suj_i} - m_{suj_j}$ appelés « contrastes ». Diverses méthodes existent ; le lecteur se reportera à [277].

B.1.3 La technique du bootstrap

Cette méthode a été mise au point par Efron [68]. Elle permet d'évaluer la précision d'une analyse statistique à partir des données d'un seul échantillon de population. Le principe consiste à créer, à partir de l'échantillon original, un grand nombre d'échantillons de même taille afin de calculer la probabilité pour que la quantité estimée soit comprise dans différentes fourchettes de valeurs. Pour déterminer ces nouveaux échantillons, on effectue un tirage au sort, avec remise, parmi la population de l'échantillon de départ ; ces échantillons sont dits « simulés ». Le nom de cette méthode de simulation « bootstrap », qui signifie tirant de bottes, indique que le seul échantillon dont on dispose permet d'en créer d'autres. Il n'est, d'une part, plus nécessaire de supposer que les données suivent une loi normale, et d'autre part, il n'est plus indispensable de ne se cantonner qu'aux mesures statistiques dont les propriétés sont calculables analytiquement.

On peut considérer que la partie redonne le tout. Mais les avantages de la méthode bootstrap ne sont valables qu'en moyenne et, comme toute autre procédure statistique, celle-ci peut donner des résultats

erronés pour un faible pourcentage d'échantillons.

Nous proposons ici un rapide survol de la méthode. Supposons que nous voulions estimer un paramètre θ défini par

$$\theta = t(F), \quad (\text{B.17})$$

avec F la fonction de distribution de probabilité et t la procédure d'évaluation. L'estimateur étendu du paramètre θ est défini par

$$\hat{\theta} = t(\hat{F}), \quad (\text{B.18})$$

avec $\hat{\theta}$ la distribution empirique. En d'autres termes, on estime la fonction $\theta = t(F)$ de la distribution de probabilité F par la même fonction de la distribution empirique \hat{F} , $\hat{\theta} = t(\hat{F})$. On utilise le bootstrap pour étudier l'erreur standard¹ et le biais de l'estimateur $\hat{\theta}$. Ces deux quantités sont obtenues automatiquement de manière numérique, indépendamment de la complexité de l'estimateur $\hat{\theta}$.

Supposons qu'un échantillon $\mathbf{x} = (x_1, \dots, x_n)$ soit extrait d'une distribution de probabilité inconnue F , et que nous voulions estimer un paramètre $\theta = t(F)$ à partir de \mathbf{x} . Pour cela, nous calculons un estimateur $\hat{\theta} = s(\mathbf{x})$ à partir de \mathbf{x} ; $s(\mathbf{x})$ peut être l'estimateur étendu $t(\hat{F})$, mais pas nécessairement. Notons \mathbf{x}^* un échantillon de bootstrap, et

$$\hat{\theta}^* = s(\mathbf{x}^*) \quad (\text{B.19})$$

une réplication bootstrap de $\hat{\theta}$. La quantité $s(\mathbf{x}^*)$ est obtenue en appliquant la même fonction s à \mathbf{x}^* au lieu de l'appliquer à \mathbf{x} . L'estimateur bootstrap de $se_F(\hat{\theta})$, l'*erreur standard* de la statistique $\hat{\theta}$, est un estimateur étendu qui utilise la fonction de distribution empirique \hat{F} à la place de la distribution théorique inconnue F . Cet estimateur bootstrap de $se_F(\hat{\theta})$ est défini par :

$$se_{\hat{F}}(\hat{\theta}^*). \quad (\text{B.20})$$

En d'autres termes, l'estimateur bootstrap de $se_F(\hat{\theta})$ est l'erreur standard de $\hat{\theta}$ pour l'ensemble des échantillons de taille n construits par échantillonnage aléatoire à partir de la distribution empirique \hat{F} . Malheureusement, dans la quasi totalité des cas, il n'existe pas de formule analytique permettant de calculer cet estimateur. Cependant, on peut obtenir une bonne approximation numérique de l'erreur standard de $\hat{\theta}$ en évaluant la déviation standard (ou écart type) empirique des réplifications : le résultat est appelé estimateur bootstrap de l'erreur standard, noté \widehat{se}_B , où B est le nombre de réplifications. En appelant $\mathbf{x}^{*1}, \mathbf{x}^{*2}, \dots, \mathbf{x}^{*B}$ les échantillons de bootstrap, et

$$\hat{\theta}^*(\cdot) = \sum_{b=1}^B \hat{\theta}^*(b)/B = \sum_{b=1}^B s(\mathbf{x}^{*B})/B \quad (\text{B.21})$$

la moyenne des estimateurs obtenus à partir des B échantillons de bootstrap, l'estimateur bootstrap de l'erreur standard est calculé par

$$\widehat{se}_B = \left(\sum_{b=1}^B \frac{(\hat{\theta}^*(b) - \hat{\theta}^*(\cdot))^2}{B-1} \right)^{1/2}, \quad (\text{B.22})$$

avec $\hat{\theta}^*(b) = s(\mathbf{x}^{*b})$, $b = 1, 2, \dots, B$, la réplication bootstrap correspondant à chaque échantillon de bootstrap. La limite de \widehat{se}_B lorsque B tend vers l'infini est en fait l'estimateur bootstrap idéal de $se_F(\hat{\theta})$, c.-à-d.

$$\lim_{B \rightarrow \infty} \widehat{se}_B = se_{\hat{F}} = se_{\hat{F}}(\hat{\theta}^*). \quad (\text{B.23})$$

1. L'erreur standard est un terme général pour la déviation standard (ou écart type) d'une statistique décrite par une somme ; ce n'est pas un estimateur de la déviation standard basé sur les données. Dans le cas de l'estimateur de la moyenne \bar{x} , l'erreur standard vaut $se_F(\bar{x}) = \sigma_F/\sqrt{n}$

Le fait que \widehat{se}_B approche $se_{\hat{F}}$ quand B tende vers l'infini conduit à dire que la déviation standard empirique approche la déviation standard de la population lorsque le nombre de réplifications devient très grand. La population, dans ce cas, est la population des valeurs $\hat{\theta}^* = s(\mathbf{x}^*)$. D'après [68], 200 réplifications sont nécessaires pour obtenir une bonne approximation de l'erreur standard du bootstrap.

Un autre paramètre servant à contrôler la précision d'un estimateur $\hat{\theta}$ est le *biais*. Par définition, c'est la différence entre l'espérance de l'estimateur $\hat{\theta}$ et la vraie valeur du paramètre θ , c.-à-d.

$$\text{biais}_F = E_F[s(\mathbf{x})] - t(F). \quad (\text{B.24})$$

Les estimateurs étendus $\hat{\theta} = t(\hat{F})$ ne sont pas nécessairement non biaisés, mais on constate généralement qu'ils tendent à avoir un biais faible comparé à leur erreur standard. On peut donc utiliser le bootstrap pour calculer le biais d'un estimateur $\hat{\theta} = s(\mathbf{x})$. L'estimateur bootstrap du biais est défini comme étant l'estimateur de biais_F obtenu en substituant \hat{F} dans l'équation (B.24),

$$\text{biais}_{\hat{F}} = E_{\hat{F}}[s(\mathbf{X}^*)] - t(\hat{F}), \quad (\text{B.25})$$

ou encore basé sur les B réplifications :

$$\widehat{\text{biais}}_B = \hat{\theta}^*(\cdot) - t(\hat{F}), \quad (\text{B.26})$$

avec $t(\hat{F})$ l'estimateur étendu de θ , qui peut différer de $\hat{\theta} = s(\mathbf{X})$.

Un estimateur amélioré du biais peut être calculé lorsque $\hat{\theta}$ est l'estimateur étendu $t(\hat{F})$ de $\theta = t(F)$. Soit P_j^* la proportion du j -ème point contenue dans l'échantillon bootstrap $\mathbf{x}^* = (x_1^*, \dots, x_n^*)$, c.-à-d.

$$P_j^* = \frac{1}{n} \text{Card}\{x_i^* = x_j\} \quad \text{pour } j = 1, 2, \dots, n, \quad (\text{B.27})$$

et on note

$$\mathbf{P}^* = (P_1^*, \dots, P_n^*) \quad (\text{B.28})$$

le vecteur de rééchantillonnage. Dans le cas où $\hat{\theta} = t(\hat{F})$ est l'estimateur étendu de θ , on a alors

$$\hat{\theta}^* = T(\mathbf{P}^*) \quad (\text{B.29})$$

pour indiquer que $\hat{\theta}^*$ est fonction du vecteur de rééchantillonnage. Les B échantillons de bootstrap $\mathbf{x}^{*1}, \mathbf{x}^{*2}, \dots, \mathbf{x}^{*B}$ fournissent donc B vecteurs de rééchantillonnage $\mathbf{P}^{*1}, \mathbf{P}^{*2}, \dots, \mathbf{P}^{*B}$, chaque vecteur \mathbf{P}^{*b} étant de la forme (B.28). On définit enfin $\bar{\mathbf{P}}^*$ la moyenne de ces vecteurs, c.-à-d.

$$\bar{\mathbf{P}}^* = \sum_{b=1}^B \frac{1}{B} \mathbf{P}^{*b} \quad (\text{B.30})$$

pour obtenir un meilleur estimateur bootstrap du biais

$$\overline{\text{biais}}_B = \hat{\theta}^*(\cdot) - T(\bar{\mathbf{P}}^*); \quad (\text{B.31})$$

$\widehat{\text{biais}}_B$ et $\overline{\text{biais}}_B$ convergent tous les deux vers $\widehat{\text{biais}}_\infty = \text{biais}_{\hat{F}}$, l'estimateur idéal bootstrap du biais, quand B tend vers l'infini, mais la convergence est plus rapide pour $\overline{\text{biais}}_B$.

Lors d'une analyse mettant en oeuvre le bootstrap, il peut être d'une grande utilité d'établir un intervalle de confiance pour θ , ce qui est plus significatif qu'un estimateur isolé $\hat{\theta}$. Pris ensemble, l'estimateur ponctuel et l'intervalle de confiance fournissent les renseignements les plus complets sur θ . Plusieurs méthodes existent pour calculer cet intervalle de confiance ; nous donnons ici le résultat de la méthode basée sur les percentiles de la distribution cumulée, car elle comporte le meilleur rapport qualité-coût.

Soit \hat{G} la fonction de répartition de $\hat{\theta}^*$. L'intervalle de confiance, au risque α , du paramètre θ est fourni par

$$[\hat{\theta}_{inf}; \hat{\theta}_{sup}] = [\hat{G}^{-1}(\alpha); \hat{G}^{-1}(1 - \alpha)]. \quad (B.32)$$

Dans le cas idéal, c.-à-d. lorsque B tend vers l'infini, on a $\hat{G}^{-1}(\alpha) = \hat{\theta}^{*(\alpha)}$ le percentile $100.\alpha$ de la distribution du bootstrap, et

$$[\hat{\theta}_{inf}; \hat{\theta}_{sup}] = [\hat{\theta}^{*(\alpha)}; \hat{\theta}^{*(1-\alpha)}]. \quad (B.33)$$

Lorsque B est fini, $\hat{\theta}_B^{*(\alpha)}$ est le $100.\alpha$ percentile empirique des $\hat{\theta}^*(b)$, c.-à-d. la $B.\alpha$ valeur dans la liste ordonnée des B réplifications de $\hat{\theta}^*$. Donc, si $B = 200$ et si $\alpha = 0.05$, $\hat{\theta}_B^{*(\alpha)}$ est la 10ème valeur ordonnée des réplifications. L'intervalle de confiance s'exprime alors par

$$[\hat{\theta}_{inf}; \hat{\theta}_{sup}] \approx [\hat{\theta}_B^{*(\alpha)}; \hat{\theta}_B^{*(1-\alpha)}]. \quad (B.34)$$

Remarquons que nous n'obtenons qu'une *approximation* de l'intervalle de confiance, dont la précision augmente avec le nombre de réplifications.

B.2 Méthodes multidimensionnelles

B.2.1 L'analyse en composantes principales

L'analyse en composantes principales ACP est une méthode d'analyse factorielle qui a pour objectif de synthétiser l'information donnée par un grand nombre de variables en un nombre restreint de variables nouvelle. L'ACP est ainsi une méthode descriptive permettant de visualiser et réduire l'information contenue dans un tableau individus \times variables ; dans la suite du paragraphe, nous supposons que n individus ont attribué une note à p variables, $p < n$, et nous appellerons x_{ij} la valeur prise par la variable X_j sur l'individu i . Matriciellement,

$$\mathbf{X} = \begin{pmatrix} x_{11} & x_{12} & \cdots & \cdots & \cdots & x_{1p} \\ x_{21} & x_{22} & \cdots & \cdots & \cdots & x_{2p} \\ \vdots & \vdots & \ddots & & & \vdots \\ \vdots & \vdots & & \mathcal{X}_{ij} & & \vdots \\ \vdots & \vdots & & & \ddots & \vdots \\ x_{n1} & x_{n2} & \cdots & \cdots & \cdots & x_{np} \end{pmatrix}.$$

Pour représenter graphiquement ce tableau, on considère qu'une ligne i du tableau décrit les coordonnées d'un point individu ; on doit donc effectuer donc une représentation d'un nuage de n points dans \mathbb{R}^p . Un problème se pose dès que $p > 3$...

Un des buts de l'ACP est de réduire la taille des matrices originelles en projetant le nuage de points sur un sous-espace où converge au mieux la topologie ; en d'autres termes, il s'agit d'unifier les variables fortement corrélées. Le principe consiste à décomposer la matrice \mathbf{X} de taille $n \times p$ en valeurs singulières, ce qui s'écrit

$$\mathbf{X} = \mathbf{TKU}^t = \sum_{i=1}^p k_i \mathbf{v}_i \mathbf{u}_i^t, \quad (B.35)$$

avec \mathbf{K} la matrice diagonale des valeurs singulières réelles, et \mathbf{T} et \mathbf{U} deux matrices orthogonales. La réduction consiste à trouver $q < p$ tel que $\mathbf{X}^* = \sum_{i=1}^q k_i \mathbf{v}_i \mathbf{u}_i^t$ soit suffisamment proche de \mathbf{X} . Les vecteurs \mathbf{u}_i sont appelés facteurs.

Soit maintenant un nuage de n points dans \mathbb{R}^p muni de la distance euclidienne, et soit \mathbf{u}_1 le vecteur directeur de F_1 , supposé meilleur axe factoriel au sens des moindres carrés. Il s'agit de définir F_1 comme la droite le long de laquelle la somme des distances quadratiques des points à F_1 est minimale. Le lecteur

démontrera sans difficulté que le vecteur \mathbf{u}_1 cherché n'est autre que le vecteur propre de $\mathbf{X}^t \mathbf{X}$, matrice des produits scalaires des points individus.

Pour généraliser, on peut écrire que le i -ème axe factoriel ajustant au mieux le nuage de points individus est défini par le vecteur directeur \mathbf{u}_i vecteur propre associé à la i -ème plus grande valeur propre λ_i de la matrice $\mathbf{X}^t \mathbf{X}$. On peut démontrer que les axes factoriels passent par le centre de gravité du nuage de points. La qualité de la représentation obtenue en se limitant aux q premiers axes factoriels est estimée en calculant

$$T_q = \frac{\sum_{i=1}^q |\lambda_i|}{\sum_{i=1}^r |\lambda_i|}, \quad (\text{B.36})$$

avec r le nombre de valeurs propres non nulles, tel que $r = \text{rang}(\mathbf{X}^t \mathbf{X}) < \min(n, p)$. On assimile donc $1 - T_q$ comme le taux d'information perdu en réduisant \mathbf{X} à \mathbf{X}^* .

Pratiquement, il faut commencer par centrer le nuage de points par rapport à son centre de gravité, ce qui revient à centrer la matrice \mathbf{X} par rapport aux moyennes des colonnes. De plus, si les données ne sont pas homogènes (toutes les variables n'ont pas la même unité), il faut essayer d'atténuer l'effet d'échelle en faisant des homothéties de rapport variable suivant la direction de \mathbb{R}^p . On normalise donc la matrice \mathbf{X} par l'estimation de l'écart type σ_j associé à la variable j , et on note \mathbf{Y} cette nouvelle matrice. Puis il ne reste plus qu'à diagonaliser la matrice dans la base des axes factoriels. La diagonalisation peut s'opérer soit par une décomposition en vecteurs et valeurs propres, soit par une décomposition en valeurs singulières ; le résultat est identique avec l'une ou l'autre méthode, à ceci près que les valeurs propres sont les variances des nouvelles coordonnées sur les axes principaux, alors que les valeurs singulières sont les écarts types.

La k -ème composante principale est le facteur \mathbf{f}_k dont les n composantes sont les coordonnées des n points individus sur le k -ème axe factoriel \mathbf{u}_k . Matriciellement,

$$\mathbf{Y} = \begin{pmatrix} y_{11} & y_{12} & \cdots & \cdots & \cdots & y_{1p} \\ y_{21} & y_{22} & \cdots & \cdots & \cdots & y_{2p} \\ \vdots & \vdots & \ddots & & & \vdots \\ \vdots & \vdots & & y_{ij} & & \vdots \\ \vdots & \vdots & & & \ddots & \vdots \\ y_{n1} & y_{n2} & \cdots & \cdots & \cdots & y_{np} \end{pmatrix} \xrightarrow{\text{ACP}} \begin{pmatrix} f_{11} & f_{12} & \cdots & f_{1q} \\ f_{21} & f_{22} & \cdots & f_{2q} \\ \vdots & \vdots & & \vdots \\ \vdots & \vdots & y_{ij} & \vdots \\ \vdots & \vdots & & \vdots \\ f_{n1} & f_{n2} & \cdots & f_{nq} \end{pmatrix}, \quad (\text{B.37})$$

ou encore $\mathbf{f}_k = \mathbf{Y} \mathbf{u}_k$ avec \mathbf{f}_k de taille $n \times 1$, \mathbf{Y} de taille $n \times p$ et \mathbf{u}_k de taille $p \times 1$.

On peut montrer que :

1. La moyenne de chaque composante principale est nulle.
2. La variance de la k -ème composante principale est égale à λ_k , k -ème valeur propre de $\mathbf{Y}^t \mathbf{Y}$.
3. Les composantes principales sont non corrélées entre elles.

La représentation graphique consiste à projeter le nuage des individus sur un plan principal ($\mathbf{u}_k, \mathbf{u}_{k'}$) qui est généralement défini par deux des trois premiers axes principaux ; plus on va loin, plus on déforme la représentation. Les coordonnées de l'individu i dans ce système d'axe sont les composantes \mathbf{f}_{ik} et $\mathbf{f}_{ik'}$.

La qualité globale de la représentation est estimée en calculant

$$T_q(\mathbf{u}_k, \mathbf{u}_{k'}) = \frac{\lambda_k + \lambda_{k'}}{\sum_{i=1}^r |\lambda_i|}; \quad (\text{B.38})$$

la qualité de la représentation est d'autant meilleure que T_q est proche de 1.

La qualité ponctuelle de la représentation est estimée pour l'individu i sur le plan $(\mathbf{u}_k, \mathbf{u}_{k'})$ en considérant le \cos^2 de l'angle α qu'il fait avec le plan, c.-à-d.

$$p = \cos^2 \alpha = \frac{\mathbf{f}_{ik} + \mathbf{f}_{ik'}}{\sum_{j=1}^p y_{ij}^2} = \frac{\|\hat{\mathbf{y}}_i\|^2}{\|\mathbf{y}_i\|^2}, \quad (\text{B.39})$$

avec $\hat{\mathbf{y}}_i$ la projection du vecteur \mathbf{y}_i dans le plan ; plus p est proche de 1, meilleure est la représentation.

Avec cette méthode, on interprète les proximités entre individus en terme de similitude de comportement (règle « qui se ressemble s'assemble ») ; deux individus sont proches dans \mathbb{R}^p si ils sont dans le plan factoriel $(\mathbf{u}_k, \mathbf{u}_{k'})$ et si ils sont bien représentés dans ce plan.

B.2.2 L'analyse MDSCAL

Cette méthode de positionnement est non métrique, ou ordinale [169] ; elle consiste à rechercher une configuration de t points dans un espace euclidien de dimension fixée p telle que les distances d entre ces points respectent au mieux l'ordre défini par les dissemblances δ : si $\delta(A_i, A_j) < \delta(A_k, A_l)$, on cherche à avoir $d(A_i, A_j) < d(A_k, A_l)$ pour le maximum de points. La relation existant entre dissemblance initiale et distance finale est monotone. De plus, le choix des axes est arbitraire : le modèle est invariant par rotation. Pour une description détaillée, le lecteur se reportera à [279]. Globalement, l'algorithme transforme Δ , une matrice $t \times t$ symétrique (ou une demi-matrice) contenant l'information de proximité entre les t stimuli (sous forme de ressemblance ou de dissemblance), en \mathbf{X} , une matrice $t \times r$ contenant les coordonnées des t stimuli sur r dimensions. Matriciellement,

$$\begin{pmatrix} \delta_{11} & \delta_{12} & \cdots & \cdots & \cdots & \delta_{1t} \\ \delta_{21} & \delta_{22} & \cdots & \cdots & \cdots & \delta_{2t} \\ \vdots & \vdots & \ddots & \delta_{jk} & \vdots & \vdots \\ \delta_{t1} & \delta_{t2} & \cdots & \cdots & \ddots & \delta_{tt} \end{pmatrix} \implies \begin{pmatrix} x_{11} & x_{12} & \cdots & x_{1r} \\ x_{21} & x_{22} & \cdots & x_{2r} \\ \vdots & \vdots & x_{jp} & \vdots \\ x_{t1} & x_{t2} & \cdots & x_{tr} \end{pmatrix} \quad (\text{B.40})$$

avec δ_{jk} la proximité entre les stimuli j et k , et x_{jp} la coordonnée du stimulus j sur la dimension p .

Une fois que les dissemblances ont été transformées en distances euclidiennes par l'ajout d'une constante (appelée *constante additive*), le programme génère une matrice de produits scalaires, suivant la procédure décrite par [306], et reprise dans [277, 279], soit

$$b_{jk} = -\frac{1}{2}(d_{jk}^2 - d_{j\cdot}^2 - d_{\cdot k}^2 + d_{\cdot\cdot}^2), \quad (\text{B.41})$$

avec

$$d_{j\cdot}^2 = \frac{1}{t} \sum_{k=1}^t d_{jk}^2, \quad d_{\cdot k}^2 = \frac{1}{t} \sum_{j=1}^t d_{jk}^2 \quad \text{et} \quad d_{\cdot\cdot}^2 = \frac{1}{t^2} \sum_{j,k} d_{jk}^2. \quad (\text{B.42})$$

Une matrice de configuration initiale \mathbf{X} est calculée à partir des p premiers vecteurs propres de la matrice des produits scalaires ; en effet, on sait que les vecteurs propres d'une matrice de produits scalaires représentent un ensemble de composantes principales (cf. § B.2.1), c.-à-d. un ensemble de coordonnées à variance maximale ; l'initialisation est donc métrique (elle respecte au mieux les distances) et non ordinale. Puis l'optimisation non métrique débute.

À chaque itération, la configuration est normalisée pour avoir une moyenne nulle et une variance unité. Les coordonnées calculées permettent d'établir une matrice de distances. Les *disparités* D_{ij} sont alors calculées par transformation des distances d_{ij} , en regard des données : la transformation peut être monotone ou mieux linéaire, ce qui évite les solutions dégénérées (valeurs de disparités identiques). Les disparités sont alors dans le même ordre que les données, et collent aux distances le mieux possible (au

sens des moindres carrés). Puis le *stress*, ou coefficient d'ajustement du modèle, compris entre 0 et 1, est calculé selon la formule

$$\text{stress} = \frac{\sum_{i>j} (d_{ij} - D_{ij})^2}{\sum_{i>j} (d_{ij})^2}. \quad (\text{B.43})$$

Si il devient inférieur à un seuil, ou si il reste constant au cours des dernières itérations, l'itération courante clos le processus. Dans le cas contraire, la position des points est modifiée à l'aide d'une procédure de descente de gradient. Et le processus recommence : calcul des disparités et calcul du stress.

Après la dernière itération, la configuration est normalisée pour que les coordonnées sur chaque axe aient une moyenne nulle et une somme des carrés unité. De plus, si la distance choisie est la distance euclidienne, c.-à-d.

$$d(A_j, A_k) = \left(\sum_{p=1}^r |x_{jp} - x_{kp}|^w \right)^{1/w} \quad (\text{B.44})$$

avec w appelée constante de Minkowski telle que $w = 2$, la configuration subit une rotation vers ses axes principaux d'inertie.

Les paramètres fournis au programme sont :

- la constante de Minkowski : $w = 2$;
- le coefficient d'ajustement du modèle : le stress ;
- la transformation des distances en disparités : linéaire ;
- le nombre de dimension.

Les deux témoins de la bonne optimisation de la solution sont :

- le stress : il doit être le plus faible possible ;
- le diagramme de Shepard² : il doit être le plus régulier possible.

La variation du stress avec le nombre de dimensions présente une décroissance assez forte suivie d'une décroissance plutôt lente ; on remarque facilement un coude, qui indique que les paramètres supplémentaires associés aux dimensions supplémentaires ne contribuent plus à améliorer la précision du modèle. Les dimensions de l'espace correspondent au nombre d'attributs perceptifs sur lesquels le sujet a construit sa notation.

B.2.3 L'analyse INDSCAL

Si, comme nous le supposons, nous utilisons tous les mêmes facteurs pour appréhender un phénomène, par contre nous avons chacun notre propre sensibilité. On peut donc imaginer qu'un programme puisse superposer les espaces des individus en les plaçant de façon à trouver alternativement les axes de dilatation possibles, et à faire coïncider les positions des configurations dans l'espace. Ces axes de dilatation sont les facteurs perceptifs communs à tous les sujets ; ils correspondent à la notion de sensibilité. L'espace moyen, appelé *espace perceptif de groupe*, permet de retrouver tous les espaces individuels, non par rotation, mais par dilatation dans un sens ou dans un autre suivant le sujet. La prise en compte des aspirations propres de chaque sujet implique nécessairement que le modèle INDSCAL (Individual Differences SCALing) n'est pas invariant par rotation. Les coefficients de dilatation sont les poids que les sujets accordent aux différents facteurs.

Pour une description précise de cet algorithme, le lecteur se reportera à [34, 279] ; un rappel des points importants est présenté dans [181]. Globalement, l'algorithme transforme n matrices Δ $t \times t$ symétriques (ou demi-matrices) contenant l'information de proximité entre les t stimuli (sous forme de

2. tracé des distances de la configuration finale en fonction des dissemblances initiales

ressemblance ou de dissemblance), en \mathbf{W} , une matrice $n \times r$ de poids des n sujets sur les r dimensions définissant l'espace sujet, et \mathbf{X} , une matrice $t \times r$ contenant les coordonnées des t stimuli sur r dimensions définissant l'espace perceptif de groupe. Un espace individuel peut alors être construit pour chaque sujet en appliquant la racine carrée des poids du sujet aux dimensions des stimuli. Matriciellement,

$$\begin{pmatrix} \delta_{11} & \delta_{12} & \cdots & \cdots & \cdots & \delta_{1t} \\ \delta_{21} & \delta_{22} & \cdots & \cdots & \cdots & \delta_{2t} \\ \vdots & \vdots & \ddots & \delta_{jk} & \vdots & \vdots \\ \delta_{t1} & \delta_{t2} & \cdots & \cdots & \ddots & \delta_{tt} \end{pmatrix} \quad \text{n fois} \quad (B.45)$$

$$\Downarrow$$

$$\begin{pmatrix} w_{11} & w_{12} & \cdots & w_{1r} \\ w_{21} & w_{22} & \cdots & w_{2r} \\ \vdots & \vdots & w_{\gamma p} & \vdots \\ w_{n1} & w_{n2} & \cdots & w_{nr} \end{pmatrix} \quad \text{et} \quad \begin{pmatrix} x_{11} & x_{12} & \cdots & x_{1r} \\ x_{21} & x_{22} & \cdots & x_{2r} \\ \vdots & \vdots & x_{jp} & \vdots \\ x_{t1} & x_{t2} & \cdots & x_{tr} \end{pmatrix}$$

avec δ_{jk} la proximité entre les stimuli j et k , $w_{\gamma p}$ le poids accordé par le sujet γ à la dimension p , et x_{jp} la coordonnée du stimulus j sur la dimension p . À la différence de MDSCAL, INDSCAL est une analyse métrique, c.-à-d. que la relation existant entre les dissemblances initiales et les distances finales est linéaire. La distance, pour l'individu γ , entre les stimuli j et k est une généralisation de la distance euclidienne ; elle est donnée par

$$d^\gamma(A_j, A_k) = \left(\sum_{p=1}^r w_p^\gamma |x_{jp} - x_{kp}|^2 \right)^{1/2}. \quad (B.46)$$

Une interprétation géométrique est fournie figure B.2. L'espace perceptif de groupe, en haut à gauche, montre neuf stimuli. L'espace sujet, en haut à droite, montre les poids (forces perceptives) accordés par les neuf sujets aux deux dimensions ; ces poids peuvent être vus comme des facteurs de dilatation applicables aux dimensions de l'espace perceptif de groupe. Ainsi, le sujet 3, qui possède des poids identiques sur les deux dimensions, aura un espace individuel identique à l'espace de groupe (à un facteur de zoom près). Par contre, il n'en est pas de même pour les individus 2 et 4, pour qui les dimensions sont dilatées ou comprimées. Le sujet 9, à l'origine du repère, doit être exclu de l'analyse : ou il a répondu au hasard, ou il a jugé par rapport à d'autres dimensions.

Rappelons que les dimensions sont déterminées de manière unique, et qu'il est impossible d'envisager une quelconque rotation des axes ; ceux-ci correspondent à un processus perceptif.

L'algorithme débute comme pour MDSCAL. Après ajout d'une constante (appelée *constante additive*) sur les dissemblances pour leur faire respecter l'inégalité triangulaire (d_{jk}^γ), celles-ci sont converties en produits scalaires b_{jk}^γ (éq. (B.41)), et une configuration initiale est calculée par analyse en composantes principales. Les poids initiaux sont tous égaux. L'optimisation consiste à minimiser, au sens des moindres carrés, la fonction d'erreur

$$S(\mathbf{W}, \mathbf{X}) = \sum_{\gamma, j, k} (b_{jk}^\gamma - \hat{b}_{jk}^\gamma)^2, \quad (B.47)$$

avec

$$\hat{b}_{jk}^\gamma = \sum_{p=1}^r y_{jp}^\gamma y_{kp}^\gamma \quad (B.48)$$

le produit scalaire calculé à partir de la configuration courante avec $y_{jp}^\gamma = \sqrt{w_p^\gamma} x_{jp}$ et y centré. L'itération s'effectue alternativement sur les poids w_p^γ et sur les coordonnées x_{jp} . La convergence vers un minimum

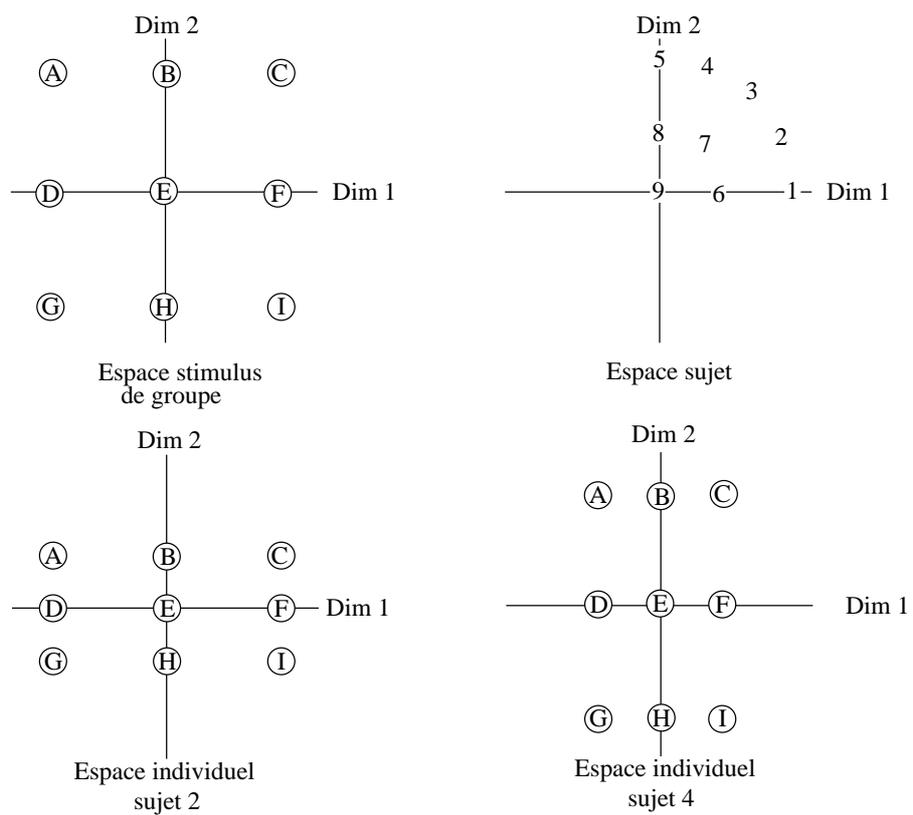


FIGURE B.2 – Exemple illustrant le modèle INDSCAL. Les poids de l'espace sujet sont appliqués aux dimensions de l'espace stimulus de groupe pour produire les espaces individuels pour les sujets 2 et 4

local est assurée, même si un minimum global est atteint dans la majorité des cas. Une dernière normalisation permet d'assurer une meilleure visualisation des plans représentant l'espace perceptif, à savoir $\sum_{j=1}^t x_{jp}^2 = 1$.

Si tous les sujets répondent de la même manière au test, les espaces individuels seront aussi les mêmes et le programme ne trouvera aucun axe de dilatation. Il donnera la position des configurations les unes par rapport aux autres, mais le repère de l'espace proposé sera arbitraire.

La corrélation entre les poids des différents individus sur les axes renseigne sur leur dispersion dans l'espace ; elle est donnée par

$$S_{pp'} = \frac{\sum_{\gamma=1}^n w_p^\gamma w_{p'}^\gamma}{\sqrt{\sum_{\gamma=1}^n (w_p^\gamma)^2 \sum_{\gamma=1}^n (w_{p'}^\gamma)^2}}. \quad (\text{B.49})$$

Si cette corrélation tend vers 1, tous les sujets sont regroupés et une rotation des axes est alors permise. Si cette corrélation est nulle, les sujets sont alors séparés en deux groupes, les uns étant insensibles à l'axe p , les autres étant insensibles à l'axe p' . Dans le cas où les sujets sont uniformément répartis dans le plan, $S_{pp'}$ prend la valeur caractéristique $2/\pi \approx 0.64$.

La corrélation entre les coordonnées des configurations sur les axes p et p' est donnée par

$$S_{pp'}^* = \frac{\sum_{j=1}^t x_{jp} x_{jp'}}{\sqrt{\sum_{j=1}^t (x_{jp})^2 \cdot (x_{jp'})^2}}. \quad (\text{B.50})$$

La valeur de $S_{pp'}^*$ mesure la quantité d'information conservée par les axes p et p' par rapport à la quantité d'information injectée *a priori* par les valeurs des critères des configurations. Si les valeurs des critères varient initialement de façon indépendante, la corrélation entre les facteurs perceptifs correspondants informe sur la dépendance ($S_{pp'}^* = 1$) ou sur l'indépendance conservée des facteurs ($S_{pp'}^* = 0$). Par contre, si les valeurs des critères sont initialement corrélés, le calcul de $S_{pp'}^*$ ne permet pas de mesurer l'indépendance des facteurs ; même si $S_{pp'}^* = 1$, cela ne remet pas en cause l'existence de ces facteurs perceptifs p et p' .

Le modèle INDSCAL est toujours capable de trouver une solution dès lors qu'on lui impose le nombre de dimensions r . Le problème est de choisir le nombre de dimensions optimal, en terme d'adéquation du modèle aux données initiales, mais aussi en terme d'efficacité du modèle, c.-à-d. que ce nombre de dimensions doit être le plus faible possible pour décrire de manière satisfaisante le modèle. Or, il est impossible de choisir *ex nihilo* le bon nombre de dimensions. Par contre, il existe *a posteriori* plusieurs indicateurs de l'optimisation du modèle.

L'indicateur principal est la *variance expliquée* par le modèle ; l'idée est de connaître la quantité d'information « sauvegardée » lors du passage dans le modèle. La variance expliquée R^2 est le carré de la moyenne des corrélations sur l'ensemble des sujets entre les produits scalaires initiaux b_{jk}^γ et ceux estimés par le modèle \hat{b}_{jk}^γ ; c'est un nombre compris entre 0 et 1 qui, plus il est proche de 1, témoigne d'un ajustement de plus en plus précis du modèle aux données initiales. Ainsi, on écrit

$$R = \frac{\sum_{\gamma,j,k} (b_{jk}^\gamma \cdot \sum_{p=1}^r w_p^\gamma x_{jp} x_{kp})}{\sqrt{(\sum_{\gamma,j,k} (b_{jk}^\gamma)^2) \cdot (\sum_{\gamma,j,k} (\sum_{p=1}^r w_p^\gamma x_{jp} x_{kp})^2)}}. \quad (\text{B.51})$$

Le pourcentage de variance sur chaque axe est estimé en pondérant R^2 par le poids normalisé moyen que les sujets ont accordé à cet axe, ce qui s'écrit

$$\nu_p = \frac{\sum_{\gamma=1}^n (w_p^\gamma)^2}{\sum_{p=1}^r \sum_{\gamma=1}^n (w_p^\gamma)^2} \cdot R^2. \quad (\text{B.52})$$

En théorie, la courbe d'évolution de R^2 en fonction du nombre de dimensions imposé à l'analyse présente une croissance rapide puis un coude au niveau du nombre de dimensions qui semble le mieux approprié ; ensuite, malgré une légère amélioration de la variance par addition d'une dimension supplémentaire, l'interprétation ne s'en trouve pas améliorée. En pratique, on se limitera donc au nombre de dimensions qui correspond à la rupture de la pente de la courbe.

D'autre part, l'algorithme débutant comme pour MDSCAL, il est également possible de calculer le *stress* (cf. § B.2.2) ; ce coefficient, compris entre 0 et 1, doit être le plus petit possible.

Enfin, le *jackknife* (ou couteau suisse) est une méthode qui permet d'étudier la stabilité d'une solution proposée par une analyse multidimensionnelle. Plus la solution est stable, moins l'introduction d'une perturbation aura de l'effet sur celle-ci. L'idée est d'étudier la perturbation due à la suppression d'un sujet ou d'un stimulus. Pour une explication détaillée, le lecteur se reportera à [60].

B.2.4 L'analyse EXSCAL

Le modèle INDSCAL peut ne pas être approprié si, en plus des dimensions partagées par les stimuli, ceux-ci diffèrent par des dimensions spécifiques à chaque stimulus, appelées *spécificités*. Winsberg & Caroll [325–327] proposent un modèle dans lequel la distance entre le stimulus j et k pour le sujet γ est définie par

$$d^\gamma(A_j, A_k) = \left(\sum_{p=1}^r w_p^\gamma |x_{jp} - x_{kp}|^2 + u^\gamma (s_j + s_k) \right)^{1/2}, \quad (\text{B.53})$$

avec s_j et s_k les spécificités des stimuli j et k , et u_γ le poids pour le sujet γ pour les spécificités. La spécificité s_j , pour le stimulus j , peut être vue comme la somme pondérée des carrés des coordonnées sur cette dimension spécifique au stimulus j , toutes étant nulle sauf celle pour le stimulus en question ; en d'autres termes, il s'agit du carré de la force perceptive d'une caractéristique possédée par le stimulus. Ce modèle très général est appelé EXSCAL (EXTENDED INDSCAL model). Bien entendu, certaines quantités peuvent être nulles, comme u^γ par exemple.

Dans ce modèle, la relation existant entre les dissemblances initiales et les distances finales est définie par

$$\delta_{jk}^\gamma = F^\gamma(d_{jk}^\gamma) + \text{erreur}. \quad (\text{B.54})$$

avec δ les dissemblances et d les distances. La fonction F n'est pas linéaire, donc le modèle est non métrique. Cependant, la relation n'est pas monotone quelconque, comme dans le cas d'un traditionnel modèle non métrique, puisqu'il s'agit de combinaisons linéaires de fonctions splines ; on parlera donc ici de modèle quasi non métrique. Ces fonctions ont l'avantage d'être aisément manipulables, puisqu'elles sont décrites par un nombre réduit de paramètres, et qu'elles sont différentiables ; d'après Caroll [33], elles permettent de minimiser les dégénérescences des solutions.

L'estimation des paramètres du modèle est du type maximum de vraisemblance. Si on suppose que les erreurs de l'équation (B.54) sont indépendantes et distribuées normalement, on peut écrire le logarithme de la fonction de vraisemblance, soit

$$\log L \propto \sum_{\gamma=1}^n \sum_{k>j} \left(\delta_{jk}^\gamma - F^\gamma(d_{jk}^\gamma) \right)^2. \quad (\text{B.55})$$

L'estimation par maximum de vraisemblance se transforme en une estimation au sens des moindres carrés. Les paramètres à estimer sont classés en trois ensembles : les paramètres spatiaux, x_{jp} et s_j ($j = 1 \dots t, p = 1 \dots r$), les paramètres de la transformation c_q (ordre des fonctions splines $q = 1 \dots Q$), et les poids w_p^γ ($\gamma = 1 \dots n$) et u^γ . Nous ne détaillerons pas le mécanisme de cette optimisation ; le lecteur se reportera aux références citées en début de paragraphe. Globalement, l'algorithme commence par choisir un ensemble de valeurs pour les paramètres de transformation et les paramètres de poids, et

cherche la configuration spatiale qui maximise le logarithme de la vraisemblance ; puis, conditionnellement à la configuration et aux poids, l'algorithme cherche la transformation qui maximise la vraisemblance ; et enfin, conditionnellement à la configuration et à la transformation, il cherche les poids qui maximisent la vraisemblance. L'optimisation fait intervenir un hessien sur les coordonnées ; elle est du second ordre.

Comme l'estimation est du type maximum de vraisemblance, on peut comparer divers modèles, en terme de qualité d'ajustement des paramètres. En toute rigueur, les deux critères exposés ci-dessous ne sont pas suffisants pour décider du choix d'un modèle ; ils ne fournissent qu'une présomption. Pour lever le doute, il faudrait procéder à une analyse de type Monte Carlo entre deux modèles concurrents [127].

Les critères statistiques suivants s'appliquent uniquement lorsque le modèle est asymptotique par rapport au nombre d'individus [327] ; ce n'est pas le cas pour un modèle avec pondérations individuelles (et *a fortiori* avec spécificités) puisqu'on introduit plus d'un paramètre par nouvel individu. Ces critères ne constituent alors pas un véritable test statistique, mais plutôt un indicateur de l'adéquation du modèle. Il faut alors procéder à une interprétation éclairée de la situation...

Le premier, noté AIC (Aikake Information Criteria) et introduit par Aikake [3], est défini par

$$\text{AIC} = -2 \log L + 2\nu, \quad (\text{B.56})$$

avec ν le nombre de degrés de libertés du modèle. Cette statistique ne prend pas en compte la taille de l'échantillon (nombre de sujets et nombre de stimuli), et peut conduire à sélectionner un modèle trop coûteux vis-à-vis du nombre de paramètres. Schwartz [291] introduit un second critère, noté BIC (Bayesian Information Criteria), qui répond à cette critique. Dans le cas d'une comparaison par paires, il est défini par

$$\text{BIC} = -2 \log L + \nu \log \left(n \frac{t(t-1)}{2} \right). \quad (\text{B.57})$$

Le modèle avec la plus petite valeur de ces statistiques est élu pour donner la meilleure représentation des données. Empiriquement, Winsberg & De Soete [328] conseillent de se fier au BIC. En général, pour une analyse qui ne comporte pas de minima locaux, ces critères diminuent avec l'augmentation du nombre de paramètres du modèle jusqu'à un certain point, avant d'augmenter. En pratique, le critère le plus important sur lequel le choix final doit être fondé est l'interprétabilité des dimensions et spécificités fournies par le modèle.

Les spécificités ne doivent être prises en compte que si le modèle se trouve considérablement amélioré.

Page blanche

ANNEXE C

Filtrage numérique

C.1 Filtrage linéaire des signaux à temps discret

Nous donnons ici un rappel des propriétés importantes ; pour une étude plus approfondie, le lecteur se reportera à [243].

C.1.1 Définitions

Un système discret est un système dont l'entrée $x(t)$ et la sortie $y(t)$ sont des signaux échantillonnés, vérifiant le théorème de Shannon.

Comme dans le cas des signaux à temps continu, on peut définir les notions de linéarité et d'invariance dans le temps.

Un système F est dit *linéaire* si $F\{\lambda x_1(n) + \mu x_2(n)\} = \lambda F\{x_1(n)\} + \mu F\{x_2(n)\}$, avec λ et μ constantes réelles.

Un système F est dit *invariant dans le temps* si un décalage en temps de l'entrée produit un même décalage en temps de la sortie, c.-à-d. $F\{x(n - k)\} = F\{x(n)\}_{n=n-k}$.

On appelle *filtre linéaire* un système linéaire et invariant dans le temps.

Un filtre linéaire répond à une *équation de convolution* entre le signal d'entrée et une fonction caractéristique appelée *réponse impulsionnelle*. Cette propriété se traduit en temporel par la relation

$$y(n) = \sum_{k=-\infty}^{+\infty} x(k)h(n - k). \quad (\text{C.1})$$

La fonction $h(n)$ est définie comme la réponse au signal impulsion-unité, c.-à-d.

$$h(n) = F\{\delta(n)\}. \quad (\text{C.2})$$

La fonction $H(z)$, définie comme la transformée en z de $h(n)$, à savoir

$$H(z) = \sum_{n=-\infty}^{n=+\infty} h(n)z^{-n}, \quad z \in D_h, \quad (\text{C.3})$$

avec D_h le domaine de convergence de h , s'appelle la *fonction de transfert*.

La fonction $H(e^{2j\pi f})$, définie comme la transformée de Fourier à temps discret (TFTD) de $h(n)$, à savoir

$$H(e^{2j\pi f}) = \sum_{n=-\infty}^{+\infty} h(n) \exp(-2j\pi n f), \quad (\text{C.4})$$

s'appelle le *gain complexe*.

L'équation (C.1) peut se mettre, dans le domaine fréquentiel, sous la forme

$$Y(e^{2j\pi f}) = H(e^{2j\pi f})X(e^{2j\pi f}). \quad (\text{C.5})$$

On montre facilement que les exponentielles complexes $\exp(2j\pi n f)$ sont les *fonctions propres* des filtres linéaires, associées aux *valeurs propres* $H(e^{2j\pi f})$; cela s'écrit

$$F\{\exp(2j\pi n f)\} = H(e^{2j\pi f}) \exp(2j\pi n f). \quad (\text{C.6})$$

C.1.2 Équation linéaire aux différences à coefficients constants

Une importante sous-classe de filtres linéaires est constituée par les systèmes qui répondent à une équation linéaire aux différences à coefficients constants, c.-à-d.

$$y(n) + a_1 y(n-1) + \dots + a_p y(n-p) = b_0 x(n) + b_1 x(n-1) + \dots + b_q x(n-q). \quad (\text{C.7})$$

En utilisant les propriétés de la transformée en z , on obtient pour la fonction de transfert une fraction rationnelle

$$H(z) = \frac{B(z)}{A(z)} = \frac{b_0 + \dots + b_q z^{-q}}{1 + a_1 z^{-1} + \dots + a_p z^{-p}}. \quad (\text{C.8})$$

Les zéros et les pôles de la fonction de transfert correspondent respectivement aux racines du numérateur et du dénominateur. Ils sont soit réels, soit complexes conjugués.

Si tous les a_i sont nuls, le filtre est à *réponse impulsionnelle finie*, ou RIF; dans le cas contraire, il est dit à *réponse impulsionnelle infinie*, ou RII.

Si tous les b_i sont nuls, le filtre est dit *auto régressif*, ou AR; dans le cas contraire, il est dit *auto régressif à moyenne ajustée*, ou ARMA.

C.1.3 Propriétés

Un système est dit *causal* si sa réponse impulsionnelle est nulle pour les instants négatifs, ce qui s'écrit

$$h(n) = 0 \quad \forall n < 0; \quad (\text{C.9})$$

cette condition implique que les échantillons en sortie du filtre ne dépendent que des échantillons aux instants précédents.

Un système est dit *stable* si à toute entrée bornée correspond une sortie bornée (BIBO, ou Bounded Input Bounded Output). On peut montrer qu'une condition nécessaire et suffisante pour qu'un filtre linéaire soit stable est que sa réponse impulsionnelle $h(n)$ soit de module sommable, ce qui s'écrit

$$\sum_{n=-\infty}^{+\infty} |h(n)| < +\infty. \quad (\text{C.10})$$

On en déduit que le filtre est stable si et seulement si le domaine de convergence de sa fonction de transfert $H(z)$ contient le cercle unité. Comme le domaine de convergence ne doit contenir aucun pôle, on en déduit que la fonction de transfert ne doit avoir aucun pôle sur le cercle unité. Mais on peut montrer par ailleurs que si $H(z)$ correspond à la solution *causale*, alors le domaine de convergence s'étend au-delà du pôle de plus grand module. Cela implique qu'un filtre linéaire est à la fois *causal* et *stable* si et seulement si tous ses pôles sont à l'intérieur du cercle unité.

La linéarité de la phase d'un filtre est une propriété très importante en audio, car elle garantit le respect de la forme temporelle du signal, dans la mesure où son contenu fréquentiel n'est pas modifié. Pour les signaux transitoires, c'est une caractéristique à laquelle l'oreille est très sensible. Les filtres à phase linéaire sont très utilisés pour la réalisation de filtres passe-bas, passe-bande et passe-haut, pour lesquels il est capital que le signal dans la bande ne soit pas déformé.

La linéarité de la phase est assurée pour un filtre à réponse impulsionnelle finie par l'importante propriété

$$h(N - n) = h(N + n), \quad (\text{C.11})$$

où N représente le milieu de la réponse impulsionnelle ; un tel filtre introduit un retard de N échantillons.

Les filtres à réponse impulsionnelle infinie causaux et stables ne peuvent en aucun cas avoir une phase linéaire.

On distingue deux quantités qui mesurent le retard introduit par un filtrage. Il s'agit du *retard de phase*

$$\tau_\varphi = -\frac{1}{2\pi} \frac{\text{Arg}(H(e^{2j\pi f}))}{f}, \quad (\text{C.12})$$

et du *retard de groupe*

$$\tau_g = -\frac{1}{2\pi} \frac{d \text{Arg}(H(e^{2j\pi f}))}{df}. \quad (\text{C.13})$$

Ces deux retards ne sont égaux que si le filtre est à phase linéaire. Notons que les formules ci-dessus ne sont valables *que* pour un signal à *bande étroite* ; en particulier, on ne peut pas dire, en général, que le retard introduit par un filtre soit caractérisé par son temps de propagation de groupe ; cela n'est vrai que pour un signal à bande étroite dont le support coïncide avec la bande passante du filtre H .

Tout filtre peut être vu comme la mise en série d'un filtre à *phase minimale* et d'un filtre *passe-tout*, de telle sorte que sa fonction de transfert peut s'écrire (la dépendance fréquentielle étant omise par souci de clarté)

$$\begin{aligned} H &= |H| e^{j\varphi} \\ &= |H| e^{j\varphi_{\text{mph}}} e^{j\varphi_{\text{eph}}} \\ &= H_{\text{min}} H_{\text{exc}}, \end{aligned} \quad (\text{C.14})$$

avec $H_{\text{min}} = |H| e^{j\varphi_{\text{mph}}}$ le filtre à phase minimale, et $H_{\text{exc}} = e^{j\varphi_{\text{eph}}}$ le filtre passe-tout. De plus, on peut montrer que le module et la phase d'un filtre à phase minimale forment une paire de transformée de Hilbert [243], à savoir

$$\text{mph} = \Im(\text{Hilbert}(-\log |H|)). \quad (\text{C.15})$$

Tout filtre H est donc entièrement déterminé par son module $|H|$ et par l'excès de phase eph de sa composante passe-tout.

On dit qu'un filtre linéaire est à *minimum de phase* si les pôles et les zéros de sa fonction de transfert sont à l'intérieur du cercle unité ; le filtre inverse se retrouve donc à la fois causal et stable.

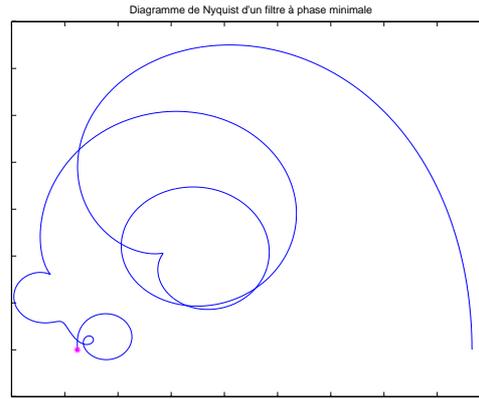


FIGURE C.1 – Diagramme de Nyquist d'un filtre à phase minimale, sur lequel on peut constater que l'origine (*) n'est jamais entourée

Dans le cas où la fonction de transfert est une fraction rationnelle à coefficients constants, un filtre à minimum de phase est uniquement défini par la donnée du module de son gain complexe (ce n'est pas vrai dans le cas général, puisque les pôles et les zéros sont soit réels soit complexes conjugués). Cette approche est souvent adoptée dans la synthèse d'un filtre à partir de la seule donnée de son gain.

Le filtre à minimum de phase est, parmi tous les filtres ayant le même gain, celui qui répond le plus rapidement. En effet, si on note $y_{\min}(n)$ la réponse du filtre à minimum de phase, et $y(n)$ celle de tout autre filtre ayant le même gain, on peut montrer que pour tout signal d'entrée $x(n)$, on a

$$\sum_{k=-\infty}^n |y(k)|^2 \leq \sum_{k=-\infty}^n |y_{\min}(k)|^2, \quad (\text{C.16})$$

ce qui veut bien dire que le filtre à minimum de phase est celui qui a l'énergie de sa réponse impulsionnelle la plus concentrée autour de 0.

D'autre part, si un filtre est à minimum de phase, son retard de groupe est nul ; au pire, il est constant si la partie passe-tout se réduit à un retard pur.

Enfin, on peut établir une condition nécessaire et suffisante pour qu'un filtre soit à minimum de phase. Pour cela, on utilise le critère de Nyquist afin de détecter la présence de zéros à l'extérieur du cercle unité. Le critère de Nyquist est basé sur un théorème de Cauchy dans le plan complexe : si une variable complexe z dans le plan z décrit un contour C_1 dans le sens trigonométrique, alors $H(z)$, fonction de la variable z , décrira un contour C_2 dans le plan $H(z)$ qui entourera N fois l'origine dans le sens trigonométrique, où N est la différence entre le nombre de pôles et de zéros de $H(z)$ inclus dans C_1 [229, 278]. Pour les systèmes discrets, le contour C_1 n'est autre que le cercle unité. Donc si $H(z)$ est la transformée en z d'un système causal et stable, alors $H(z)$ n'a aucun pôle à l'extérieur du cercle unité ; donc le nombre de tours de $H(z)$ autour de l'origine, lorsque z décrit le cercle unité, est précisément le nombre de zéros de $H(z)$ extérieurs au cercle unité dans le plan z .

Pratiquement, le test se résume à tracer un diagramme pôlaire (ou courbe de Lissajous), c.-à-d. la partie réelle de $H(e^{2j\pi f})$ en abscisse, et sa partie imaginaire en ordonnée ; si le tracé n'entoure pas le cercle unité, alors le filtre est à phase minimale (cf. fig. C.1). Pour cela, il est nécessaire que $H(e^{2j\pi f})$ soit échantillonné sur le cercle unité avec un pas suffisamment fin pour éviter les brusques variations de phase. Si l'on dispose de la réponse impulsionnelle du filtre, on aura donc soin de calculer sa TFD sur un grand nombre de points, et au besoin d'effectuer un zero-padding ; et si il existe un retard avant l'arrivée du son direct, celui-ci doit être éliminé puisqu'il est équivalent à un filtre passe-tout.

On dit qu'un filtre linéaire est *passé-tout* si son gain est indépendant de la fréquence. Les zéros de $H(z)$ sont les inverses (conjugués) de pôles de $H(z)$, et la phase est toujours décroissante (avec une

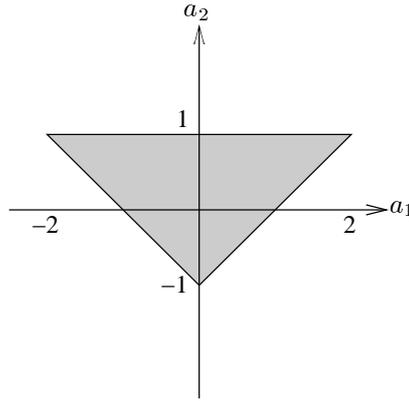


FIGURE C.2 – Domaine de stabilité de la cellule du second ordre (en gris)

brusque variation au niveau des pôles). On peut montrer que pour tout signal d'entrée $x(n)$ et tout signal de sortie $y(n)$ on a la propriété

$$\sum_{k=-\infty}^n |x(k)|^2 \geq \sum_{k=-\infty}^n |y(k)|^2, \quad (\text{C.17})$$

ce qui veut dire qu'à l'instant t , il y a moins d'énergie en sortie.

C.1.4 La cellule d'ordre deux

Le filtre récursif d'ordre deux constitue la brique élémentaire du filtrage dans les logiciels Max/MSP® et jMax®. Son équation de récurrence s'écrit

$$y(n) = b_0x(n) + b_1x(n-1) + b_2x(n-2) - a_1y(n-1) - a_2y(n-2), \quad (\text{C.18})$$

Ce filtre est bien causal car sa sortie dépend des entrées et des sorties précédentes ; sa fonction de transfert s'écrit

$$H(z) = \frac{B(z)}{A(z)} = \frac{b_0 + b_1z^{-1} + b_2z^{-2}}{1 + a_1z^{-1} + a_2z^{-2}}. \quad (\text{C.19})$$

Il est aisé de calculer le critère de stabilité, c.-à-d. l'ensemble des valeurs de a_1 et a_2 qui garantissent la stabilité du filtre : il suffit de calculer les racines du polyôme $A(z)$ et de voir sous quelles conditions elles ont un module inférieur à 1 ; on obtient la domaine connexe représenté sur la figure C.2. De plus, on peut montrer que si

$$\left| \frac{a_1(1+a_2)}{4a_2} \right| < 1, \quad (\text{C.20})$$

alors le gain du filtre passe par un maximum dont la fréquence (dite fréquence de résonance) est donnée par

$$f_r = \frac{1}{2\pi} \arccos \left(- \frac{a_1(1+a_2)}{4a_2} \right); \quad (\text{C.21})$$

la valeur du maximum (appelé surtension) est donnée par

$$H(f_r) = \frac{1}{(1-a_2) \sin \varphi}, \quad (\text{C.22})$$

avec $a_1 = -2r \cos \varphi$ et $a_2 = r^2$, où l'on vérifie que r représente le module des pôles et $\pm\varphi$ leur argument. On remarquera que la surtension à la résonance est d'autant plus grande que a_2 est voisin de 1.

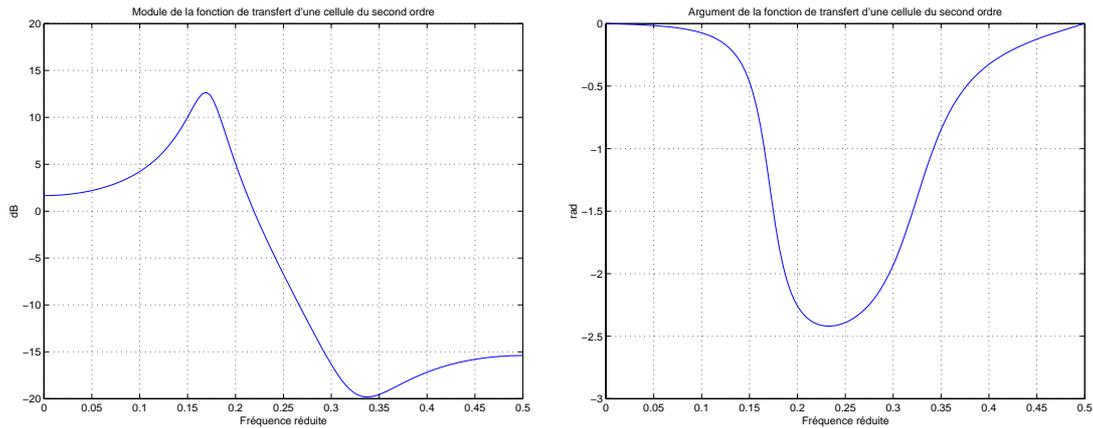


FIGURE C.3 – Réponse en fréquence d'une cellule d'ordre deux. À gauche : module ; à droite : argument

La partie non récursive, représentée par le polynôme $B(z)$ sert à créer une antirésonance dans le gain du filtre ; afin de retomber sur les mêmes formules que ci-dessus pour la fréquence de l'extremum (ici un minimum) et la valeur du gain correspondant (mais inversé), il est commode de normaliser le polynôme $B(z)$ par b_0 qui n'est autre qu'un gain global.

Le filtre complet $H(z)$ permet donc de créer une résonance et une antirésonance, comme on peut le voir sur la figure C.3. On remarquera la brusque variation de la phase autour de la résonance. Le lecteur se référera à [56, 194, 324] pour plus d'informations sur la cellule du second ordre et son utilisation en audio.

C.1.5 Structures des filtres

Nous ne détaillerons pas ici la multitude de façons d'implanter un filtre numérique RIF ou RII dans un calculateur numérique ; pour une revue complète, le lecteur se reportera encore une fois à [243].

Pour un filtre RIF, il y a plusieurs techniques associées à la notion de filtrage, c.-à-d. relatives à l'implantation de la convolution.

La technique temporelle (encore appelée filtrage transverse ou structure directe), met en œuvre l'équation (C.1) ; cette implantation est à proscrire, vue l'énorme quantité de calculs à effectuer : N opérations par échantillon pour un filtre de longueur N (près de 50 000 opérations par échantillon pour une réponse impulsionnelle de 1 seconde échantillonnée à 44 kHz).

L'implantation fréquentielle, ou l'utilisation de l'équation (C.5), à l'aide des techniques de convolution rapide basées sur l'algorithme de FFT, permet de réduire le coût de calcul dans des proportions considérables, mais nécessite toujours un retard de traitement correspondant au chargement d'un bloc de signal d'une durée supérieure ou égale à celle de la réponse (technique addition-recouvrement ou sauvegarde-recouvrement). Cette contrainte est uniquement acceptable pour une application de convolution dans laquelle le temps réel n'est pas primordial. Pour un filtre de longueur N et pour des blocs de taille $M \geq N$, en notant $L = M + N - 1$ le coût de calcul est environ de $4 \log_2(L) + 4$ opérations par échantillon (toujours pour une réponse impulsionnelle de 1 seconde échantillonnée à 44 kHz, le coût de calcul est ainsi ramené à moins de 100 opérations par échantillon).

On peut mentionner également l'approche polynômiale (algorithme de Toom-Cook) basé sur l'interpolation de Lagrange, mais qui nécessite de séparer les multiplications générales et les multiplications par des constantes ou des additions.

La convolution hybride est une méthode combinant un filtrage temporel RIF pour le début de la réponse impulsionnelle, et une convolution en fréquence pour la fin de la réponse (elle-même divisée en blocs). Cette méthode permet d'effectuer la convolution avec un coût de calcul raisonnable donc en

temps réel, mais surtout *sans retard* [31, 94]. Notons que le coût de calcul est variable, car il dépend de la taille des blocs de FFT, et aussi d'un certain retard pouvant être introduit.

Pour un filtre RII, la structure la plus efficace en terme de coût de calcul est la structure en série ; tout filtre $H(z)$ est factorisé en cellules du second ordre, soit pour un filtre d'ordre N pair

$$H(z) = \prod_{k=1}^{N/2} \frac{b_{0k} + b_{1k}z^{-1} + b_{2k}z^{-2}}{1 + a_{1k}z^{-1} + a_{2k}z^{-2}}. \quad (\text{C.23})$$

Grâce à l'appariement des pôles et des zéros, et grâce à l'ordre des sections en cascade, il est possible de réduire l'importance du bruit de quantification des coefficients, et d'éviter les erreurs d'arrondis [324] ; numérateur et dénominateur sont donc regroupés de sorte que la fréquence de résonance et la fréquence d'antirésonance de la cellule soient proches, et ainsi que le gain maximal de la cellule soit proche de 1 ; les cellules sont donc ordonnées par gain maximal décroissant. On peut même stocker le gain global du filtre dans le coefficient b_{01} de la première cellule, tandis que les $N/2 - 1$ suivantes n'ont besoin que de quatre coefficients ; on obtient alors un coût de calcul de $4 \times (N/2) + 1 = 2N + 1$ opérations par échantillons.

C.1.6 Synthèse d'un filtre non récursif

La fonction de transfert d'un filtre non récursif ou à réponse impulsionnelle finie (RIF) est un polynôme en z , c.-à-d.

$$H(z) = \sum_{n=0}^{N-1} h_n z^{-n}, \quad (\text{C.24})$$

et sa réalisation ne pose pas de problème *a priori* puisque la fonction de transfert n'a pas de pôle.

Parmi les avantages des filtres RIF, on peut citer :

- la possibilité de réaliser des filtres à phase linéaire, particulièrement intéressants pour les applications dans le domaine de l'audio (l'oreille n'étant pas sensible à la phase proprement dite, mais plutôt aux variations de phase) ;
- la stabilité inconditionnelle, puisqu'il n'y a pas de pôle ;
- la linéarité des méthodes de spécifications ;
- la possibilité d'obtenir un bruit de calcul assez faible.

Parmi les inconvénients :

- la nécessité d'un ordre élevé pour obtenir des filtres à coupure raide (en d'autres termes, les filtres RIF ont un niveau de performances plus faible que les filtres RII pour un ordre donné) ;
- un temps de propagation de groupe qui n'est pas forcément un nombre entier d'échantillons.

L'existence d'une symétrie, paire ou impaire, des coefficients du filtre par rapport au milieu de la fenêtre permet de réaliser un filtre à *phase linéaire*.

Dans la suite, on suppose donné le gain en fréquence $H(e^{2j\pi f})$ du filtre à synthétiser.

Méthode de la fenêtre La réponse impulsionnelle $h(n)$ peut être obtenue par la formule de transformation inverse suivant l'expression (dans le domaine échantillonné)

$$h(n) = \int_{-\frac{1}{2}}^{+\frac{1}{2}} H(e^{2j\pi f}) \exp(2j\pi n f) df. \quad (\text{C.25})$$

En général, la suite $h(n)$ est infinie. Aussi pour pouvoir réaliser l'opération de filtrage directement à partir de la relation de convolution, il ne faut conserver qu'un nombre fini P de valeurs. De façon à

effectuer une approximation aussi précise que possible, on prendra les P valeurs les plus significatives, c.-à-d. celles qui maximisent $\sum_P |h(n)|^2$. Ce choix est facilité en pratique par le fait que la suite $|h(n)|$ est souvent paire et décroissante en fonction de n . Dans ce cas, il suffit de prendre les $P = 2K + 1$ valeurs successives allant de $-K$ à $+K$.

Mais l'opération de troncature de la suite des coefficients sur le gain en fréquence revient à multiplier la suite $h(n)$ par la fonction rectangle ; elle a donc pour conséquence d'introduire, par convolution, des ondulations sur le gain en fréquence, ondulations qui peuvent même ne pas être évanescentes lorsque P tend vers l'infini (phénomène de Gibbs). Toutefois, il est possible de limiter ces ondulations en utilisant la *méthode des fenêtres*. La suite $h(n)$, calculée précédemment par la formule d'inversion, est alors multipliée, terme à terme, par une fenêtre de pondération ; ces fenêtres réduisent les ondulations constatées sur le gain en fréquence. Mais gardons à l'esprit que la réduction des ondulations (c.-à-d. la réduction de la hauteur des lobes secondaires) s'accompagne toujours de l'élargissement de la bande de transition (c.-à-d. l'élargissement du lobe principal).

Méthodes itératives La synthèse du filtre $h(n)$ est ici considérée comme un problème d'approximation pondérée au sens de Chebyshev. Pour cela, il faut déterminer un ensemble de conditions pour lesquelles il est possible de montrer que la solution est optimale (au sens où l'erreur d'approximation est minimisée sur l'intervalle entier), et unique. L'idée est d'exprimer $|H(e^{2j\pi f})|$ sous la forme d'un produit $|H(e^{2j\pi f})| = P(e^{2j\pi f})Q(e^{2j\pi f})$, produit dans lequel $Q(e^{2j\pi f})$ est connu *a priori*, et $P(e^{2j\pi f})$ est le polynôme trigonométrique dont on doit déterminer les coefficients. L'erreur d'approximation est exprimée par

$$E(e^{2j\pi f}) = W(e^{2j\pi f})[H(e^{2j\pi f}) - Q(e^{2j\pi f})P(e^{2j\pi f})], \quad (\text{C.26})$$

avec $W(e^{2j\pi f})$ une fonction de pondération qui permet de donner une importance différente à l'erreur suivant la fréquence f . La minimisation au sens de Chebyshev est celle de la norme L_∞ , soit

$$\|E(e^{2j\pi f})\| = \min_{\text{coef de } P} \left(\max_{f \in [0, \frac{1}{2}]} |E(e^{2j\pi f})| \right). \quad (\text{C.27})$$

Un algorithme d'échange ou de Remez repose sur un théorème d'alternance, donne un ensemble de conditions nécessaires et suffisantes sur l'erreur commise ; cet algorithme permet de résoudre le problème d'approximation en cherchant les fréquences extrêmes de la meilleure approximation [152].

C.1.7 Synthèse d'un filtre récursif

Contrairement aux filtres non récursifs, il est physiquement impossible d'obtenir avec les filtres récursifs causaux des fonctions de transfert à déphasage linéaire sur toute la bande de fréquence. La fonction de transfert d'un filtre récursif ou à réponse impulsionnelle infinie (RII) est de la forme

$$H(z) = \frac{B(z)}{A(z)} = \frac{\sum_{i=0}^M b_i z^{-i}}{\sum_{j=0}^N a_j z^{-j}} = \sum_{n=0}^{\infty} h_n z^{-n}. \quad (\text{C.28})$$

Les filtres récursifs présentent les avantages suivants :

- une bande de transition étroite pour un ordre équivalent à celui d'un filtre non récursif ;
- la possibilité de réaliser des déphaseurs purs, dits aussi filtres passe-tout.

Les techniques utilisées pour effectuer la synthèse d'un filtre récursif peuvent être classées en trois catégories :

1. La transposition d'un filtre continu.
2. La synthèse directe dans le plan complexe, par recherche de la position des pôles et des zéros.
3. La synthèse à l'aide de méthode d'optimisation qui en général sont des méthode itératives.

Transposition à partir de filtres continus La première méthode est celle de l'invariant impulsionnel. Comme son nom l'indique, cette méthode consiste à dire que la réponse impulsionnelle du filtre numérique est la version échantillonnée de la réponse impulsionnelle du filtre continu. Cette méthode donne de bons résultats pour des filtres de type Butterworth, Bessel ou Chebyshev, en ce sens qu'elle conserve les propriétés de phase aussi bien que les propriétés d'atténuation ; cette méthode est simple à mettre en œuvre, moyennant le fait que la fonction de transfert du filtre continu vérifie $H_c(e^{2j\pi f}) \approx 0$ pour $|f| \geq \frac{f_c}{2}$, ce qui permet de définir une fréquence d'échantillonnage f_e et de limiter les recouvrements de spectres liés à l'échantillonnage ; une deuxième condition est que $h_c(0) = 0$, ce qui traduit l'absence de discontinuité à l'origine. Cependant, cette méthode nécessite de connaître $H(e^{2j\pi f})$ par ses pôles et ses zéros, ce qui nécessite souvent un calcul de racines de polynômes qu'il n'est pas évident de toujours mener correctement pour des ordres trop élevés.

Transformation bilinéaire La méthode précédente permet d'obtenir un filtre numérique dont seule la réponse impulsionnelle est invariante. Mais il est possible de déterminer une méthode pour laquelle le filtre numérique présentera le même comportement que le filtre continu quelque soit l'entrée. La transformation peut être abordée de deux façons : soit comme une transformation géométrique du plan complexe qui transforme l'axe imaginaire du plan de Laplace en cercle unité du plan z , soit comme une identification de l'opérateur d'intégration. C'est souvent le deuxième principe qui est utilisé. On obtient un filtre numérique équivalent en effectuant la transformation

$$H(z) = H(p) \Big|_{p=\frac{2}{T} \frac{z-1}{z+1}}, \quad (\text{C.29})$$

avec T la période d'échantillonnage. Le filtre ainsi obtenu aura approximativement la même réponse temporelle que le filtre de départ continu ; cette transformation est appelée transformation bilinéaire. Le demi-plan gauche devient l'intérieur du cercle unité, l'axe imaginaire le cercle unité, et le demi-plan droit l'extérieur du cercle unité ; il y a donc conservation de la stabilité.

La relation entre les fréquences continues et numériques est un peu plus compliquée ; il y a bel et bien distorsion de fréquence. Soit $z = e^{2j\pi f_N T}$ où f_N représente la fréquence du domaine numérique, il vient

$$f_c = \frac{1}{\pi T} \tan(\pi f_N T), \quad (\text{C.30})$$

avec f_c la fréquence correspondante du domaine continu. Il s'ensuit que lorsque f_c varie de façon continue de $-\infty$ à $+\infty$, alors $f_N T$ varie de $-\frac{1}{2}$ à $+\frac{1}{2}$: le fait de décrire l'axe imaginaire tout entier fait décrire un seul tour sur le cercle unité.

La procédure de synthèse est donc la suivante :

1. Définir le gabarit du filtre numérique.
2. Redéfinir le gabarit dans le domaine continu à l'aide de la distorsion de fréquences.
3. Effectuer la synthèse à l'aide de méthodes du continu (Butterworth, Bessel, Chebyshev, Legendre, Cauer, elliptique, etc.)
4. Effectuer la transformation de l'équation (C.29) qui donne directement $H(z)$.

Conception directe Il est possible de spécifier un filtre directement dans le domaine numérique, par approximation de sa fonction de transfert. À la différence des méthodes de spécifications à partir du domaine continu, les méthodes de conception directes ne sont pas limitées à des des filtres passe-bas, passe-haut, passe-bande ou stop-bande.

MÉTHODE DE YULE-WALKER

Cette méthode spécifie un filtre récursif par ajustement (au sens des moindres carrés dans le domaine temporel) du gain d'une fonction de transfert donné. L'origine du nom provient de la méthode utilisée

pour calculer les coefficients du dénominateur : l'algorithme calcule la TFD inverse du spectre de puissance idéalement désiré, et résout les équations de Yule-Walker modifiées en utilisant les coefficients de la fonction d'autocorrélation. Le calcul du numérateur est plus complexe, et fait intervenir l'ajustement temporel sus-cité ; pour un descriptif complet de la méthode, le lecteur pourra se référer à [89]. Cette méthode n'accepte aucune information de phase.

MÉTHODE DE LA TRANSFORMÉE EN z INVERSE

À la différence de la méthode précédente, cette méthode accepte une fonction de transfert complexe en entrée. Les coefficients du numérateur et du dénominateur sont calculés à l'aide d'un ajustement au sens des moindres carrés dans le domaine fréquentiel discret, par le biais de la minimisation d'une fonction d'erreur, c.-à-d.

$$\min_{b,a} \sum_{k=0}^{N-1} w(k) \left| H(k) - \frac{B(k)}{A(k)} \right|^2, \quad (\text{C.31})$$

avec k fréquence discrète, N taille de la TFD, w fonction de pondération de l'erreur, H la fonction de transfert désirée, $\frac{B}{A}$ le filtre récursif calculé ; celui-ci est toujours stable.

MÉTHODE PAR PRÉDICTION LINÉAIRE

Cette méthode génère exclusivement un filtre autorégressif (AR), ou tout pôles, dont la réponse impulsionnelle s'ajuste au mieux à une séquence temporelle définie. Elle suppose que chaque échantillon de sortie du filtre est une combinaison linéaire de n échantillons passés, et que les coefficients sont constants au cours du temps. Ceux-ci sont calculés par inversion de la matrice d'autocorrélation (en général grâce à l'algorithme récursif de Durbin-Levinson appliqué à la dite matrice, de forme Toeplitz symétrique). Le filtre généré est toujours stable, mais il peut ne pas modéliser exactement la séquence définie, car le calcul de l'autocorrélation suppose implicitement que les échantillons sont nuls au-delà de la longueur de la séquence.

MÉTHODE DE PRONY

À la différence de la méthode précédente, cette méthode génère un filtre autorégressif à moyenne ajustée (ARMA) dont la réponse impulsionnelle est la meilleure approximation d'une séquence temporelle ; pour une description complète, le lecteur se reportera à [251, p. 226-228]. Cette méthode utilise une variante de la méthode précédente pour calculer les coefficients du dénominateur, puis calcule les coefficients du numérateur pour lesquels la réponse impulsionnelle du filtre résultant correspond exactement aux $n_b + 1$ échantillons de la séquence définie (n_b étant l'ordre du polynôme du numérateur). Le filtre généré n'est pas forcément stable, mais à la différence de la méthode précédente, il peut retrouver exactement les coefficients de la séquence d'entrée si celle-ci est véritablement un processus ARMA du même ordre. Les capacités de modélisation sont, pour un ordre donné, supérieures à la méthode par prédiction linéaire

Remarquons cependant qu'à la différence de la modélisation ARMA classique, qui suppose que l'entrée du système à modéliser est du bruit, la méthode de Prony suppose que l'entrée du système est une impulsion, et la méthode de Steiglitz-MacBride que l'entrée du système est arbitraire.

MÉTHODE DE STEIGLITZ-MACBRIDE

Comme la méthode de Prony, cette méthode détermine les coefficients d'un modèle ARMA dont la réponse impulsionnelle est la meilleure approximation d'une séquence temporelle donnée ; pour une description complète de la méthode, le lecteur se reportera à [185, 301]. Mais à la différence de la méthode de Prony, la méthode de Steiglitz-MacBride réalise l'identification complète d'un système, basée

sur la connaissance d'une séquence d'entrée et d'une séquence de sortie correspondante. L'algorithme est itératif ; il débute comme pour la méthode de Prony, c.-à-d. que les coefficients du dénominateur sont d'abord calculés par la méthode de la covariance modifiée ; puis l'entrée et la sortie sont filtrées par le filtre AR précédemment calculé, et les coefficients du numérateur sont calculés par la minimisation d'une fonction d'erreur entre la sortie du filtre calculé et la séquence de sortie désirée. L'algorithme converge très rapidement ; mais il peut aussi ne pas converger du tout si l'ordre du modèle est trop grand. Comme pour la méthode de Prony, le filtre généré n'est pas nécessairement stable. Les capacités de modélisation sont, pour un ordre donné, supérieures à la méthode de Prony.

C.1.8 Lissage fréquentiel

Le système auditif a la propriété bien connue d'analyser et d'interpréter les signaux captés avec une résolution fréquentielle de plus en plus faible à mesure que la fréquence augmente [120, 332]. Ce phénomène peut être modélisé par un banc de filtres dont la largeur augmente avec la fréquence [253, 295]. Classiquement, le banc de filtres en question est à Q constant, le plus utilisé étant celui en 1/3 d'octaves car il approxime au mieux la largeur des bandes critiques de l'oreille.

Les bancs de filtres auditifs ont été caractérisés à l'aide de mesures psychoacoustiques, conduisant à une approximation de la réponse en fréquence des filtres en termes de fonction de Gauss [252], ou de « rounded exponential » [254] ; puis ont été introduites l'échelle des Barks et celle des ERB [97] ; plus récemment Patterson a introduit les bancs de filtres gammatones [253, 294], puis les gammachirps, ajoutant au gammatones une correction asymétrique en fonction du niveau [130].

Le processus d'analyse d'un signal sonore par l'oreille peut être modélisé par une analyse temps-fréquence sur une échelle fréquentielle variable : citons la TFD warpée [90, 128, 153, 154, 241], la transformée en ondelettes avec warping variable en fonction du temps [2], les bancs de filtres non uniformes [129, 201], les filtres de Kautz permettant une allocation arbitraire de la résolution fréquentielle [248], et la transformée en fraction d'octaves [192]. Récemment, a été proposé un algorithme basé sur la transformée de Fourier à court terme, utilisant une échelle fréquentielle variable calquée sur l'analyse temps fréquence effectuée par le système auditif [134].

De plus, pour étudier les systèmes de transmissions multiples tels que le phénomène de propagation en acoustique des salles, il faut se souvenir que l'oreille tend à détecter l'attaque des signaux, c.-à-d. qu'elle est sensible à une large étendue fréquentielle des premières réflexions ; parallèlement, l'oreille tend à ignorer les composantes hautes-fréquences des réflexions tardives. Ces caractéristiques ont conduit à la mise en œuvre de traitements fréquentiels avec une fenêtre de longueur variable, tels que ceux implantés dans le système MLSSA [269].

Outre l'intérêt concernant l'adéquation de la représentation au fonctionnement de notre système auditif, il en existe un autre concernant la modélisation d'un filtre. Une fois l'opération de lissage effectuée, la modélisation requiert un plus petit nombre de coefficients.

Lissage du spectre de puissance Les bancs de filtres peuvent être utilisés dans le domaine fréquentiel, sur le module d'un gain complexe obtenu par transformée de Fourier (assimilable à un banc de filtres de largeurs constantes). Le lissage est effectué par l'intermédiaire d'une moyenne pondérée des échantillons du spectre ; mais le nombre de points résultants est moindre (il est égal au nombre de filtres dans le banc). Si on veut ensuite procéder à une modélisation, il faut effectuer une interpolation entre ces valeurs pour revenir à une échelle fréquentielle constante.

Le lissage peut aussi être réalisé par l'intermédiaire d'une moyenne pondérée des échantillons du spectre, avec une fenêtre de pondération symétrique sur une échelle logarithmique (en général une fenêtre de Hann) définie par son facteur de qualité Q, et glissante. On obtient donc un spectre lissé avec le même nombre de points que le spectre original, mais chaque nouveau point représente une moyenne

effectuée sur un intervalle d'autant plus large que la fréquence augmente. Dans le cas des signaux discrets, considérons le spectre de puissance $|H(k)|^2$, où k est l'indice fréquentiel tel que $0 \leq k \leq N - 1$; l'opération de lissage peut être exprimée à l'aide de l'équation de convolution

$$\begin{aligned} H_1(k) &= |H(k)|^2 \circledast W_{\text{sm}}(m(k), k) \\ &= \sum_{i=0}^{N-1} |H(k-i) \bmod N|^2 W_{\text{sm}}(m(k), i), \end{aligned} \quad (\text{C.32})$$

où \circledast représente l'opération de convolution circulaire, et $W_{\text{sm}}(m(k), k)$ est une fenêtre de lissage ayant la forme générale d'un filtre passe-bas (en général une combinaison linéaire de fonctions cosinus), dont la demi-largeur $m(k)$ varie en fonction de l'indice fréquentiel k . Ordinairement, les fenêtres sont définies par c.-à-d.

$$W(k) = \sum_{i=0}^{L-1} b_i \cos\left(\frac{2ik\pi}{N}\right), \quad (\text{C.33})$$

avec L le nombre de coefficients, et $0 \leq n \leq N - 1$ avec N impair (cf. [119]). La fenêtre rectangulaire correspond à $L = 1$ avec $b_0 = 1$, et la fenêtre de Hann correspond à $L = 2$ avec $b_0 = 0.5$ et $b_1 = -0.5$. Plus précisément, dans le cas d'un lissage du spectre complexe, la forme de la fenêtre est donnée par

$$W_{\text{sm}}(m(k), k) = \begin{cases} \frac{b-(b-1) \cos[(\pi/m(k))k]}{2b(m(k)+1)-1} & k = 0, 1, \dots, m(k) \\ \frac{b-(b-1) \cos[(\pi/m(k))(k-N)]}{2b(m(k)+1)-1} & k = N - m(k), N - (m(k) - 1), \dots, N - 1 \\ 0 & k = m(k) + 1, \dots, N - (m(k) + 1) \end{cases} \quad (\text{C.34})$$

On peut, par exemple, choisir $m(k)$ pour effectuer un lissage sur une échelle de $1/3$ d'octave, ou mieux directement sur l'échelle des ERB. Un tel lissage, effectué avant l'étape de modélisation précédemment décrite, permet de répartir uniformément l'erreur de modélisation dans les différentes bandes critiques.

Le lissage peut aussi être effectué dans le domaine cepstral [243]; le principe est d'effectuer une troncature du cepstre, le nombre de points conservés correspondant à l'idée de la fréquence de coupure d'un filtre passe-bas. Aucune référence ne peut être donnée quant à une quelconque résolution fréquentielle.

Dans le même ordre d'idée, on peut décomposer un signal temporel (ou un spectre complexe) sur une base d'exponentielles complexes, et ne garder qu'un nombre fini de coefficients.

Lissage du spectre complexe Si le filtrage est effectué sur le module du gain complexe, aucun traitement n'est effectué sur la phase. Le processus de lissage du module est donc *non réversible*, en ce sens qu'il est impossible de retrouver la réponse impulsionnelle dont le module de son gain complexe serait justement le module lissé (sauf à supposer que la réponse impulsionnelle est à phase nulle). Or, la réponse temporelle modifiée (calculée à partir d'un lissage du module et de la phase) peut être aussi pertinente, d'un point de vue perceptif, que le spectre de puissance lissé. L'idée du lissage complexe [121] est de modifier une réponse impulsionnelle pour que, d'une part elle acquière les propriétés d'un spectre de puissance lissé, et d'autre part elle puisse mettre en évidence les caractéristiques temporelles perceptivement pertinentes du système étudié. L'équation de convolution devient dans ce cas

$$\begin{aligned} H_{1c}(k) &= H(k) \circledast W_{\text{sm}}(m(k), k) \\ &= \sum_{i=0}^{N-1} [H(k-i) \bmod N] W_{\text{sm}}(m(k), i). \end{aligned} \quad (\text{C.35})$$

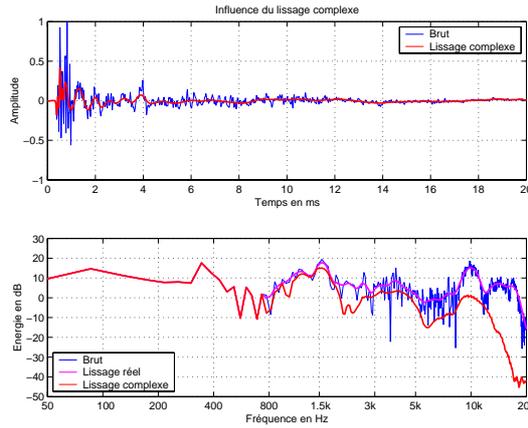


FIGURE C.4 – Influence d'un lissage complexe en 1/3 d'octave sur une réponse impulsionnelle mesurée à l'intérieur d'un véhicule

Le résultat est présenté figure C.4 ; il est clair que les composantes hautes fréquences sont conservées au début de la réponse impulsionnelle, et qu'un effet passe-bas est progressivement introduit. Cet effet correspond bien à notre processus auditif, spécialement dans le cas de l'analyse d'une réponse impulsionnelle de salle. De plus, l'effet du lissage complexe peut être mis à profit dans le calcul de filtres inverses pour l'égalisation acoustique, puisqu'il permet de réaliser des filtres avec une sensibilité plus faible –d'un point de vue perceptif– aux modifications de la position l'auditeur dans la salle.

C.1.9 La technique du warping

La méthode d'approximation d'une fonction de transfert $H(z)$ consiste à chercher les coefficients des polynômes $B(z)$ et $A(z)$ tels que $H(z) \approx \frac{B(z)}{A(z)}$; comme nous l'avons vu ci-dessus, il existe de nombreuses approches envisageables. Smith propose la notion de *warping* [297–299], ou distorsion de l'échelle des fréquences, avant l'étape de modélisation. L'échelle fréquentielle est dilatée vers les hautes fréquences, afin de réduire l'importance relative des hautes fréquences par rapport aux basses fréquences ; en effet, il n'y a perceptivement aucun intérêt à ce que la modélisation accorde autant d'importance à l'intervalle [10-12] kHz qu'à l'intervalle [0-2] kHz (par exemple), alors que justement l'intervalle [0-2] kHz présente une pertinence auditive accrue. Une fois la modélisation effectuée, il faut repasser le filtre dans le domaine fréquentiel originel.

Le warping est réalisé par la transformation bilinéaire

$$\zeta = \frac{z - r}{1 - rz}, \quad (\text{C.36})$$

où r est le paramètre de warping compris entre -1 et +1 ; on peut voir sur la figure C.5 l'effet de cette opération sur l'axe fréquentiel. D'après Smith, la valeur optimale (au sens de la minisation à l'aide de la norme L_∞ de l'erreur relative de la largeur de bande) du coefficient de warping permettant d'approcher au mieux l'échelle des Barks est donnée par

$$r^*(f_e) \approx 1.0674 \sqrt{\frac{2}{\pi} \arctan(0.06583 f_e)} - 0.1916, \quad (\text{C.37})$$

avec f_e la fréquence d'échantillonnage en kHz ; la formule est extrêmement précise pour des valeurs de f_e voisines de 40 kHz. Pour une fréquence d'échantillonnage de 44 kHz, on obtient une valeur de r^* proche de 0,756.

La valeur optimale (au sens du même critère) du coefficient de warping permettant d'approcher au mieux l'échelle des ERB de Moore et Glasberg est donnée par

$$r^*(f_e) \approx 0.7446 \sqrt{\frac{2}{\pi} \arctan(0.1418 f_e)} + 0.03237, \quad (\text{C.38})$$

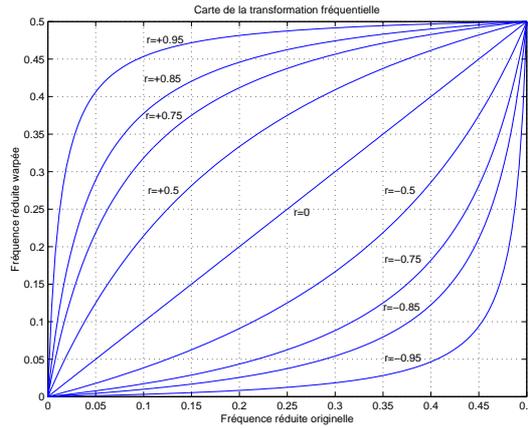


FIGURE C.5 – Distorsion de l'axe fréquentiel sous l'influence de l'opération de transformation bilinéaire suggérée par Smith ; r est le paramètre de warping

avec f_e la fréquence d'échantillonnage en kHz ; on obtient la valeur 0,738 pour une fréquence d'échantillonnage de 44kHz. Due à la résolution fréquentielle plus élevée de l'échelle des ERB, principalement en basses fréquences, le filtre du premier ordre utilisé est moins efficace, et l'erreur d'approximation est deux à trois fois plus importante qu'avec l'échelle des Barks.

Le warping est une opération de transformation complète du plan complexe, de telle sorte que le plan complexe originel z est transformé en un autre plan complexe ζ . Cette opération peut être réalisée directement dans le plan complexe (c.-à-d. sur les pôles et les zéros de $F(z)$), ou sur le module de $F(z)$; mais elle peut aussi être réalisée directement sur le signal temporel.

Dans le plan complexe, on peut directement remplacer les pôles et les zéros de $F(z)$ par leur nouvelles valeurs, calculées en écrivant que $\zeta = \frac{z-r}{1-rz}$.

Sur la fonction de transfert $F(z)$, on peut remplacer z par $\frac{\zeta+r}{1+r\zeta}$; en écrivant $\zeta = \cos \theta + j \sin \theta$, on trouve que

$$\arg z = \arctan \left(\frac{\sin \theta (1 - r^2)}{(1 + r^2) \cos \theta + 2r} \right) ; \quad (\text{C.39})$$

il est donc possible de ré-échantillonner l'axe fréquentiel sur lequel est évalué le module de $F(z)$ pour obtenir le module sur un axe fréquentiel warpé.

Sur le signal temporel, l'opération est plus complexe. Il faut se reporter à la structure directe d'un filtre RIF (cf. § C.1.5 et fig. C.6), dans laquelle le retard z^{-1} sera remplacé par sa version warpée, c.-à-d.

$$\frac{z^{-1} - r}{1 - rz^{-1}} , \quad (\text{C.40})$$

qui n'est autre que la fonction de transfert d'un filtre passe-tout [154] ; une structure équivalente à l'implantation transverse, mais avec un coût de calcul relativement faible, a été développée à cette occasion (cf. fig. C.7). Les avantages de cette technique sont nombreux. D'abord, elle permet de réaliser le warping « à la volée », c.-à-d. d'obtenir directement le signal temporel résultant de la convolution d'un signal source avec une réponse impulsionnelle warpée. Ensuite, cette structure est valide quelle que soit l'entrée, donc en particulier pour une impulsion ; il est donc possible d'obtenir une réponse impulsionnelle (temporelle) dans le domaine warpé. Celle-ci présente une information basses fréquences comprimée dans ses premiers échantillons, tandis que l'information hautes fréquences est repoussée vers ses derniers échantillons ; on peut expliquer ceci par la diminution (resp. l'augmentation) de la constante de temps des pôles proches du point $z = 1$ (resp. $z = -1$) [75]. En d'autres termes, pour un nombre donné de coefficients, la réponse impulsionnelle warpée est, par rapport à la version non warpée, plus précise en basses fréquences et plus grossière en hautes fréquences. La réponse impulsionnelle warpée peut

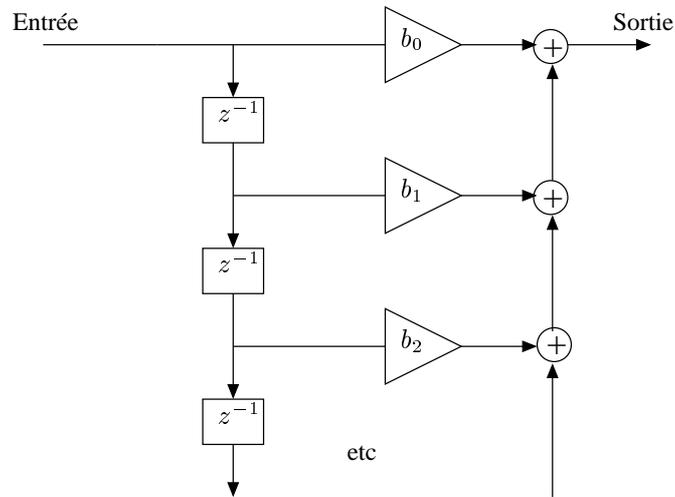


FIGURE C.6 – Schéma de câblage d'un filtrage RIF transverse

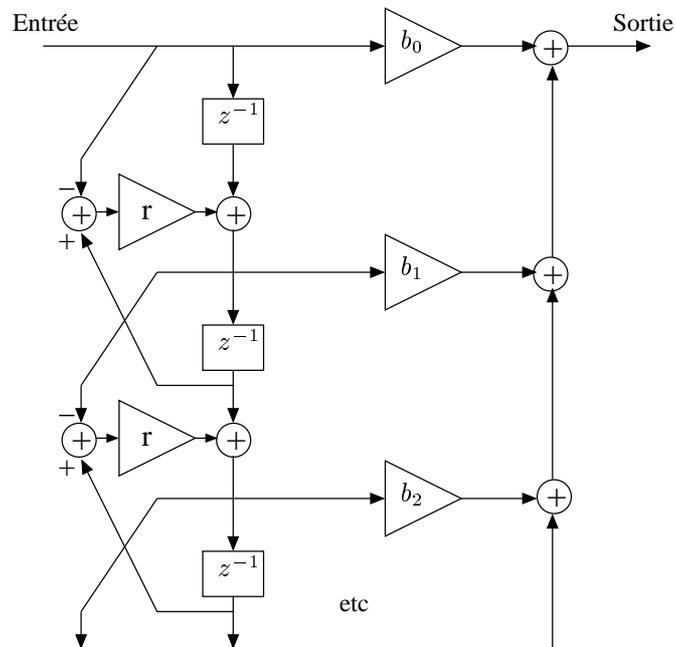


FIGURE C.7 – Schéma de câblage d'un filtrage RIF avec warping intégré

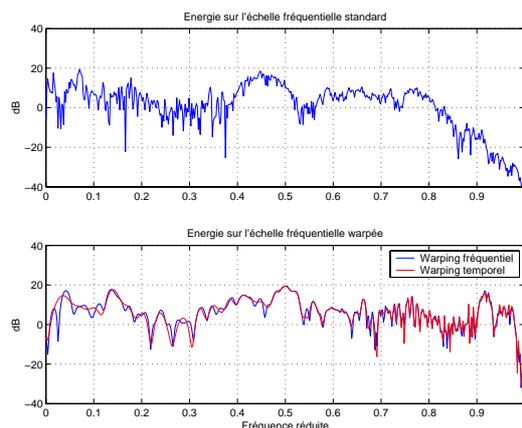


FIGURE C.8 – Comparaison du warping effectué dans le domaine temporel et dans le domaine fréquentiel

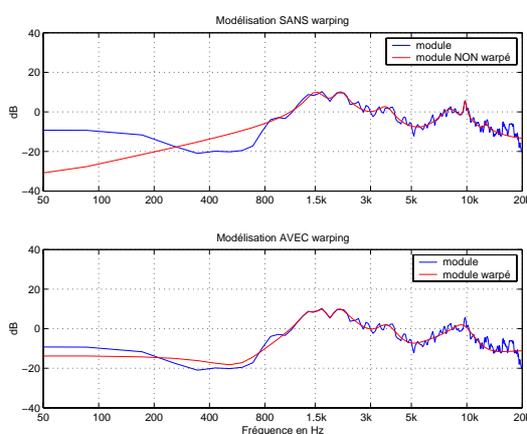


FIGURE C.9 – Mesure et modélisation du module d'une fonction de transfert sous la forme d'un filtre récursif d'ordre 12 à l'aide de la méthode de Steiglitz-MacBride ; le warping est effectué sur l'échelle des Barks

donc être tronquée après ses premiers échantillons sans perdre trop d'information en basses fréquences, spécialement pour les valeurs élevées du coefficient de warping r ; le coût de calcul de la convolution s'en trouve réduit de 40%. Farina exploite cette technique pour réaliser une égalisation en fréquence : la réponse impulsionnelle du canal acoustique à égaliser est d'abord warpée (filtrage avec warping et une impulsion en entrée), puis inversée dans le domaine warpé, et stockée, avant d'être convoluée avec un signal source quelconque (filtrage avec warping inverse et le signal temporel en entrée) [17, 307].

La figure C.8 compare les résultats d'un warping effectué directement dans le domaine temporel sur une réponse impulsionnelle, et dans le domaine fréquentiel sur le module de la fonction de transfert ; la technique temporelle, bien que nettement plus coûteuse, est légèrement plus précise en basses fréquences.

Terminons avec un exemple de modélisation d'un filtre (fig. C.9) ; on veut modéliser le module de la fonction de transfert représenté en bleu sous forme d'un filtre récursif d'ordre 12, grâce de la méthode de Steiglitz-MacBride. D'une part, nous avons modélisé directement le module ; d'autre part, nous avons warpé le module sur l'échelle des Barks, puis effectué l'étape de modélisation, et warpé en sens inverse les pôles et les zéros. Il est clair que la solution warpée fournit une meilleure évaluation globale que la solution directe, qui sacrifie la précision en dessous de 4 kHz pour obtenir une meilleure approximation au-dessus de 10 kHz.

C.2 Transformée de Fourier à court terme

De nombreuses références bibliographiques traitent de la transformée de Fourier à court terme (TFCT). La présentation la plus complète est le livre de Crochiere et Rabiner [51] ; une approche moins détaillée est envisagée par Allen et Rabiner [4]. Dans l'écriture passe-bande de la TFCT, les voies du banc de filtre équivalent sont des signaux complexes à bande étroite. La référence temporelle est glissante : les trames à court terme sont référencées par rapport au début de la fenêtre courante. Dans cette convention, la TFCT s'écrit

$$X(p, k) = \sum_{n=0}^{N-1} h(-n)x(pR + n) e^{-2j\pi kn/N}, \quad (\text{C.41})$$

avec $h(n)$ la fenêtre d'analyse de support temporel $[-(N-1); 0]$, N la longueur de la fenêtre d'analyse, n l'échantillon temporel, R le facteur de sous-échantillonnage temporel, p l'échantillon temporel sous-échantillonné, et k l'échantillon fréquentiel. Pour une fréquence discrète k fixée, le signal de sous-bande $X(p, k)$ est indicé par p et bien échantillonné à une fréquence R fois plus faible que le signal analysé $x(n)$ indicé par n . Chaque coupe de la TFCT à k constant représente l'évolution temporelle du signal dans le canal centré autour de la fréquence réduite k/N : la largeur de ce canal est donnée par la résolution fréquentielle du fenêtrage employé.

Une analyse avec fenêtres de Hann ne fournit pas un résultat satisfaisant car la décroissance en haute fréquence est perturbée par la décroissance en basse fréquence. Le choix de la fenêtre d'analyse s'est donc porté sur celle de Blackman-Harris d'ordre 4. Cette fenêtre possède d'excellentes caractéristiques globales [119, 239], les plus intéressantes étant la hauteur extrêmement faible de ses lobes secondaires (-92 dB) et le faible écart entre la largeur à -3 dB du lobe principal de sa transformée de Fourier discrète (TFD) et la bande de bruit équivalente. Même si le taux de décroissance des lobes secondaires de sa TFD est faible (6 dB par octave), il est suffisant, car l'objectif de l'analyse est d'obtenir une dynamique maximale à toutes les fréquences. La résolution fréquentielle relativement faible de cette fenêtre, fixée par la largeur à -6 dB du lobe principal de sa transformée de Fourier (2.72 bins) [119, 239], est un avantage, puisque plusieurs points fréquentiels voisins se trouveront naturellement moyennés grâce à la convolution fréquentielle. En effet, il ne sert à rien d'avoir une résolution fréquentielle extrême, puisqu'il y a fort à parier que les caractéristiques que l'on recherche (temps de réverbération et densité de puissance initiale) ne varient pas brutalement en fréquence.

La longueur de la fenêtre a été fixée à 11,6 ms, ce qui correspond à 512 échantillons (on rappelle que la fréquence d'échantillonnage des réponses impulsionnelles mesurées est de 44000 Hz). La taille de la transformée de Fourier discrète, implémentée sous forme de FFT, a été fixée aussi à 512 points. On n'effectue donc pas de « zero padding ». Celui-ci serait inutile, puisqu'il augmenterait l'échantillonnage fréquentiel de la transformée de Fourier à temps discret, mais sans apporter aucune information, puisque les points rajoutés ne sont pas indépendants. La résolution fréquentielle de notre fenêtre est donc de 86 Hz.

Le pas de décalage des fenêtres, ou facteur de sous-échantillonnage des signaux de sous-bande, est choisi en accord avec les conditions de reconstruction parfaite.

Lorsque l'on cherche à reconstruire un signal analysé par TFCT, la méthode la plus employée, lorsque R est différent de 1, est l'*overlap-add*. Toujours dans le cas de notre approche passe-bande, elle s'écrit

$$y(n) = \sum_{p=-\infty}^{+\infty} f(n - pR) \frac{1}{N} \sum_{k=0}^{N-1} Y(p, k) e^{2j\pi k(n-pR)/N}, \quad (\text{C.42})$$

avec $f(n)$ la fenêtre de synthèse. La condition de transparence de l'analyse-synthèse par TFCT implique

que la somme du produit des fenêtres d'analyse et de synthèse décalées soit égal à 1, pour tout n , c.-à-d.

$$\sum_{p=-\infty}^{+\infty} f(n - pR)h(pR - n) = 1, \forall n. \quad (\text{C.43})$$

Dans le cas où la fenêtre de synthèse est une fenêtre carrée, la condition de transparence devient

$$\sum_{p=-\infty}^{+\infty} h(pR - n) = 1, \forall n. \quad (\text{C.44})$$

Il existe une règle simple, portant sur le pas d'avancement des fenêtres, permettant de vérifier théoriquement cette condition [32]. La fenêtre de Blackman-Harris d'ordre 4, comme beaucoup de fenêtres, est obtenue par échantillonnage d'une fenêtre continue qui est une combinaison linéaire de P termes en cosinus (quatre termes ici), à savoir

$$h(t) = \sum_{p=0}^P a_p \cos(2\pi pt) \quad \text{avec } t \in \left[-\frac{1}{2}; \frac{1}{2}\right]. \quad (\text{C.45})$$

Si le recouvrement entre fenêtres s'écrit sous la forme $(1 - 1/K)$, il faut choisir $K > P$ pour que la fenêtre continue vérifie la condition (C.44), au facteur d'échelle $K a_0$ près. Tous les recouvrements de la forme $(1 - 1/K)$ conviennent. Par contre, il n'est pas sûr que la fenêtre échantillonnée puisse satisfaire cette condition, spécialement si sa longueur est une puissance de 2 ($K = 3$, par exemple). Dans notre cas, le plus petit recouvrement envisageable est de $3/4$ ou 384 échantillons. Le pas d'avancement maximal est donc de 128 points.

La largeur à -6 dB de la transformée de Fourier de la fenêtre de Blackman étant de 2.72 bins, si la longueur de la fenêtre est de 512 points et la longueur de la FFT de 512 points, la résolution maximale de la TFCT est de 2.72 échantillons fréquentiels. On aurait donc pu se contenter d'un échantillonnage à 188 points fréquentiels (soit 1 point tous les 170 Hz).

Enfin, comme ce qui nous intéresse est une représentation temps-fréquence *énergétique*, on va considérer le carré du module de la TFCT. Cette représentation prend encore le nom de *spectrogramme* (cf. § I.4.2.1). Dans ce cas, il convient de prendre en compte l'énergie de la fenêtre dans sa représentation continue, donnée par $\frac{1}{Fe} \sum_{n=0}^{N-1} h^2(n)$. Le spectrogramme est donc à normaliser par cette quantité, comme on le fait avec un périodogramme moyenné et fenêtré pour garantir qu'il soit sans biais lorsque la longueur des trames tend vers l'infini [316].

ANNEXE D

Inversion de réponses impulsionnelles

D.1 Inversion de réponses impulsionnelles

D.1.1 Cadre théorique

Lorsque le dispositif de restitution n'est pas symétrique, le concept de phase minimale jointe, introduit par Cooper et Bauck dans l'implantation shuffler, ne peut plus s'appliquer. Il faut procéder à une inversion complète (module et phase) du déterminant Δ . Or la création d'un filtre inverse causal et stable est une tâche insolvable [43] ; seule existe une solution approchée. Ce problème a été largement étudié par Mourjopoulos [224], dans le but d'effectuer l'inversion d'une réponse impulsionnelle de salle.

Rappelons les phénomènes en présence lors de l'inversion d'un filtre défini dans le domaine temporel discret sur un support fini.

D.1.1.1 Présentation du problème

Pour une description approfondie des problèmes d'inversion et de déconvolution, le lecteur se reportera à [45, 63, 117, 167, 203].

Dans un système linéaire et invariant dans le temps, l'entrée x et la sortie y sont reliées par l'équation intégrale

$$y(t) = \int_D h(t-u)x(u) du ; \quad (\text{D.1})$$

le noyau de cette intégrale est la réponse impulsionnelle du système. La déconvolution est le problème qui consiste à résoudre l'équation intégrale de convolution pour restaurer l'entrée du système à partir de sa sortie. Ce problème n'est pas limité à l'acoustique des salles, puisque l'équation (D.1) se rencontre dans de nombreux domaines de la physique où les phénomènes étudiés vérifient assez bien l'hypothèse de linéarité. La résolution d'une équation de convolution n'est absolument pas triviale, puisqu'il s'agit d'un *problème inverse* et, comme la plupart des problèmes de ce type, d'un *problème mal posé*. Les difficultés proviennent du caractère lissant de l'opérateur de convolution : l'équation (D.1) montre qu'au lieu d'accéder directement à la grandeur à mesurer $x(t)$, on obtient une grandeur expérimentale $y(t)$ qui est une moyenne pondérée de la vraie grandeur. La moyenne effectuée par l'intégrale de convolution a pour effet d'adoucir toutes les irrégularités de $x(t)$. La résolution effective d'un problème de déconvolution est numérique : elle suppose que le problème soit discrétisé. La résolution du problème nécessite aussi de donner des structures d'espace métriques à l'ensemble Ω_x des paramètres théoriques et à l'ensemble Ω_y des paramètres expérimentaux. Nous utiliserons, dans tout ce paragraphe, l'espace de Hilbert

qui est un espace complet pour sa norme associée, laquelle norme provient d'un produit scalaire, c.-à-d. $\|\vec{u}\| = \sqrt{\vec{u} \cdot \vec{u}}$. Soient Ω_x et Ω_y deux espaces de Hilbert définis sur le corps des réels ; l'équation de convolution est de la forme

$$y = Hx, \quad x \in \Omega_x \text{ et } y \in \Omega_y, \quad (\text{D.2})$$

où $H: \Omega_x \rightarrow \Omega_y$ est un opérateur linéaire, borné et compact. Pour mémoire, on rappelle que :

- H est borné si il existe un réel $c \geq 0$ tel que $\|Hx\| \leq c\|x\|$, quelque soit $x \in \Omega_x$;
- H est dit compact si il applique tout ensemble borné $S \subset \Omega_x$ dans un ensemble $A(S)$ dont la fermeture $\overline{A(S)}$ est compacte. En d'autres termes, A est compact si et seulement si, pour toute suite bornée $\{x_n, n \in \mathbb{N}\} \subset \Omega_x$, il existe une sous-suite $\{x_{n_k}, k \in \mathbb{N}\}$ et un élément $y \in \Omega_y$ tels que $A(x_{n_k}) \xrightarrow{\|\cdot\|_{\Omega_y}} y$ lorsque $k \rightarrow \infty$; un opérateur linéaire et compact est aussi borné (la réciproque est vraie si Ω_y est de dimension finie).

Toutes les difficultés évoquées plus haut proviennent du caractère compact de l'opérateur de convolution.

Un problème est dit stable ou bien posé au sens de Hadamard lorsque à une faible perturbation des données correspond une faible variation de la solution. Ici, si l'on observe $y + \delta y$ au lieu de y , l'équation doit rester soluble et $\delta y \rightarrow 0$ doit entraîner $\delta x \rightarrow 0$.

Pour que l'équation (D.2) soit un problème bien posé au sens de Hadamard, il faut satisfaire trois conditions :

1. Existence : $\forall y \in \Omega_y, \exists x / Hx = y \iff y \in \mathcal{I}m(H)$.
2. Unicité : liée à la trivialité de l'équation homogène $\mathcal{K}er(H) = 0$.
3. Stabilité : la solution x doit dépendre continûment de y .

Il est évident que Ω_y doit être l'image de Ω_x par H , mais il est moins évident que les normes dans Ω_x et Ω_y doivent être choisies soigneusement pour que H et son inverse soient continues. En effet, les conditions $\Omega_y = \mathcal{I}m(H)$ et $\mathcal{K}er(H) = 0$ sont les conditions habituelles d'existence et d'unicité de la solution, auxquelles il faut ajouter, comme pour tout problème physique, une condition de stabilité, qui est ici une condition de fermeture topologique $\mathcal{I}m(H) = \overline{\mathcal{I}m(H)}$. Malheureusement, il est prouvé que l'image d'un opérateur borné peut ne pas être fermée, et que si de plus l'opérateur est compact, l'image n'est presque jamais fermée.

Mentionnons que la symétrisation de H par son adjoint H^* n'apporte rien, car $H^*y = H^*Hx$ reste un problème mal posé en regard de la stabilité.

Si l'on se restreint au cas où x, y et h sont des fonctions possédant des transformées de Fourier (notées \mathcal{F}), une méthode de résolution utilisant les propriétés de la transformée de Fourier semble s'imposer. En effet, l'équation (D.1) devient dans le domaine fréquentiel

$$Y(f) = H(f)X(f), \quad (\text{D.3})$$

d'où la solution par filtrage inverse

$$X(f) = Y(f)H^{-1}(f) \quad \text{et} \quad x(t) = \mathcal{F}^{-1}[X(f)]. \quad (\text{D.4})$$

Comme la transformée $X(f)$ détermine complètement $x(t)$, la connaissance de $Y(f)$ et de $H(f)$ suffit en principe. En fait, les choses ne sont pas si simples.

Si l'on se place dans une algèbre de convolution \mathcal{A} (c.-à-d. un espace vectoriel de distributions sur lequel on peut définir le produit de convolution), la condition nécessaire et suffisante pour que $y = h * x$ possède une solution dans \mathcal{A} , et ce $\forall y \in \mathcal{A}$, est que h possède une inverse de convolution h^{*-1} telle que $h * h^{*-1} = \delta$. La résolution est alors immédiate, c.-à-d.

$$y * h^{*-1} = x * h * h^{*-1} = x * \delta = x. \quad (\text{D.5})$$

Mais h^{*-1} n'existe pas toujours si h est, par exemple, une distribution. Plus grave encore, h^{*-1} peut n'exister que pour certaines fonctions y . De plus, si l'algèbre \mathcal{A} admet des diviseurs de zéro, l'équation (D.1) admet une infinité de solutions.

Il ne faut pas perdre de vue l'aspect purement formel de l'équation (D.3). Pour qu'elle ait un sens, il faut que $H^{-1}(f)$ existe, que $H(f)$ soit une fonction qui ne s'annule pour aucune valeur de f et ne tende pas vers zéro plus vite que toute puissance de $1/f$. Ces conditions ne sont jamais réalisées en pratique. De plus, pour des systèmes physiques, $H(f) \rightarrow 0$ quand $f \rightarrow \infty$, et $H(f)$ n'est donc en principe connue que sur un intervalle fini borné $[-F_H; F_H]$. Quand $H(f)$ est à support borné, la relation (D.3) montre que $Y(f)$ a le même support. Ceci entraîne que, dans le cas général, $X(f)$ est complètement inconnue, et la solution impossible. L'élaboration d'une méthode universelle de déconvolution (c.-à-d. s'appliquant à tous les signaux ayant une transformée de Fourier) n'a donc aucun sens, et il faut, d'une certaine manière, la spécialiser. Cette spécialisation repose sur une information *a priori* permettant de restreindre la classe des entrées de façon à obtenir une solution unique.

Si la classe des entrées est limitée aux fonctions à spectre limité, c.-à-d. aux fonctions dont la transformée de Fourier est à support borné ($X(f) = 0$ pour $|f| \geq F_X$), il est clair que la largeur $2F_X$ du spectre de l'entrée doit être inférieure à la largeur de bande $2F_H$ de la fonction de transfert.

Si la classe des entrées est limitée aux fonctions à support borné ($x(t) = 0$ en dehors de $[a, b]$), il est clair d'après la relation (D.3) que l'information disponible sur $x(t)$ est une faible partie du spectre $X(f)$. Il est cependant possible de déterminer uniquement la solution. $H(f)$ est connue est identiquement nulle sur l'intervalle fini $|f| \geq F_H$; $x(t)$ étant à support borné, $X(f)$ est une fonction analytique régulière qui ne peut s'annuler identiquement sur un intervalle, si petit soit-il. Une conséquence particulière est que les valeurs de $X(f)$ pour $|f| \geq F_H$ déterminent de façon unique son comportement pour tout f . Il est donc théoriquement possible de reconstituer une entrée à support borné à partir de la mesure exacte de la sortie, quelles que soient les largeurs des spectres de $h(t)$ et de $x(t)$.

Ces méthodes sont d'une mise en œuvre difficile et sont très sensibles au bruit. En effet, quand les inévitables erreurs de mesure sont prises en compte, le problème change totalement, et il est possible que deux entrées très différentes donnent deux sorties différant d'une quantité partout inférieure aux erreurs de mesure. Sans perdre de généralité, on peut supposer que sur l'intervalle $|f| \geq F_H$ la fonction $Y(f)$ est connue avec une erreur $B(f)$, c.-à-d.

$$Y(f) = H(f)X(f) + B(f), \quad (\text{D.6})$$

d'où

$$X(f) = [Y(f) - B(f)]/H(f). \quad (\text{D.7})$$

Les erreurs $b(t)$ ont en général un contenu spectral étendu, et il n'y a aucune raison que $B(f)/H(f)$ se comporte comme on le souhaite quand $f \rightarrow \infty$. Le problème est à nouveau mal posé et la solution instable.

Les difficultés évoquées plus haut pour déconvoluer dans le cas analogique vont évidemment se retrouver dans le problème discrétisé qu'il faut bien résoudre pour une déconvolution effective. On peut montrer qu'il est possible de se ramener à un système d'équations linéaires dans lesquelles on choisit, par souci de commodité, de normaliser les pas d'échantillonnage à l'unité. L'entrée à restaurer étant discrétisée sur M points, le système comportant N points, et l'observateur disposant de $L = N + M - 1$ valeurs de la sortie, l'équation de convolution discrète s'écrit

$$y(i) = \sum_{j=0}^{M-1} h(i-j)x(j) \text{ avec } i = 0, 1, \dots, L-1, \quad (\text{D.8})$$

ou matriciellement

$$\mathbf{y} = \mathbf{H}\mathbf{x}; \quad (\text{D.9})$$

Pour que les erreurs dues à l'échantillonnage soient inférieures aux erreurs de mesure, il faut que $x(t)$ soit à support borné, et que les M points de l'échantillonnage soient répartis régulièrement sur ce support. Faisons l'hypothèse que le signal d'entrée et le système sont causaux, c.-à-d. $x(i) = 0$ et $h(i) = 0$ pour $i < 0$, on obtient alors

$$\begin{pmatrix} y(0) \\ y(1) \\ y(2) \\ \vdots \\ \vdots \\ \vdots \\ y(L-1) \end{pmatrix} = \begin{pmatrix} h(0) & 0 & \cdots & \cdots & 0 \\ \vdots & h(0) & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \ddots & \vdots \\ \vdots & \vdots & & \ddots & 0 \\ \vdots & \vdots & & & h(0) \\ h(N-1) & \vdots & & & \vdots \\ 0 & h(N-1) & & & \vdots \\ \vdots & 0 & \ddots & & \vdots \\ \vdots & \vdots & \ddots & \ddots & \vdots \\ 0 & 0 & \cdots & 0 & h(N-1) \end{pmatrix} \begin{pmatrix} x(0) \\ x(1) \\ \vdots \\ \vdots \\ \vdots \\ x(M-1) \end{pmatrix}. \quad (\text{D.10})$$

La matrice \mathbf{H} possède la propriété

$$h(i+p, j+p) = h(i, j), \quad \forall p, \quad i = 0, 1, \dots, L-1 \quad j = 0, 1, \dots, M-1; \quad (\text{D.11})$$

une telle matrice est dite de Toeplitz, entièrement définie par sa première ligne et sa première colonne.

Dans sa forme discrète, le problème de la déconvolution n'est qu'un cas particulier (correspondant à des espaces X et Y de dimensions finies) du problème général évoqué par l'équation (D.2), et la première démarche consiste à symétriser le problème en appliquant à l'équation (D.9) l'opérateur adjoint \mathbf{H}^t , c.-à-d.

$$\mathbf{H}^t \mathbf{y} = \mathbf{H}^t \mathbf{H} \mathbf{x}. \quad (\text{D.12})$$

Le problème semble donc se limiter à la résolution d'un système d'équations linéaires élémentaire

$$\hat{\mathbf{x}} = (\mathbf{H}^t \mathbf{H})^{-1} \mathbf{H}^t \mathbf{y}. \quad (\text{D.13})$$

Ceci suppose évidemment que la matrice symétrique $\mathbf{H}^t \mathbf{H}$ soit régulière, c.-à-d. qu'il y ait au moins autant d'observations $y(i)$ que d'inconnues $x(i)$. Si cette condition est vérifiée, le problème semble résolu ; en fait il n'en est rien, car même si $\mathbf{H}^t \mathbf{H}$ est de rang plein, l'inversion directe suivant l'équation (D.13) peut conduire à des résultats inacceptables, cette dernière difficulté provenant du caractère mal conditionné de la matrice $\mathbf{H}^t \mathbf{H}$.

En effet, le conditionnement de ces matrices dépend du pas d'échantillonnage. Pour bien comprendre cela, il faut remarquer que les lignes successives de \mathbf{H} sont très proches (due au caractère Toeplitz), et cela d'autant plus que le pas d'échantillonnage est petit. Mais on peut aussi expliquer ceci à l'aide des propriétés des valeurs propres et des vecteurs propres des matrices de Toeplitz.

La matrice $\mathbf{H}^t \mathbf{H}$ étant supposée réelle et symétrique de taille $M \times M$, elle possède des valeurs propres réelles $\lambda_j^2 \geq 0$, supposées ordonnées

$$\lambda_0^2 \geq \lambda_1^2 \geq \dots \lambda_{M-1}^2; \quad (\text{D.14})$$

ces valeurs propres correspondent aux carrés des valeurs singulières de la matrice \mathbf{H} . La matrice \mathbf{H} peut se factoriser

$$\mathbf{H} = \mathbf{U} \mathbf{\Lambda} \mathbf{V}^t, \quad (\text{D.15})$$

où \mathbf{U} et \mathbf{V} sont deux matrices orthonormales, et $\mathbf{\Lambda}$ est la matrice diagonale formée des valeurs singulières λ_j . Le rang K de la matrice symétrique $\mathbf{H}^t \mathbf{H}$ est tel que $K \leq \inf(L, M)$. On peut montrer que

$$\frac{\|\delta \mathbf{x}\|^2}{\|\mathbf{x}\|^2} \geq \left(\frac{\lambda_{\max}}{\lambda_{\min}} \right)^2 \frac{\|\delta \mathbf{y}\|^2}{\|\mathbf{y}\|^2}, \quad (\text{D.16})$$

avec $\delta \mathbf{y}$ l'erreur d'observation sur \mathbf{y} et $\delta \mathbf{x}$ l'erreur d'observation sur \mathbf{x} . On voit que, même dans le cas *a priori* favorable où la matrice est régulière ($K = M$), l'erreur relative sur \mathbf{x} est bornée par l'erreur relative sur \mathbf{y} affectée du rapport des valeurs singulières extrêmes. Ce rapport mesure le *conditionnement* de \mathbf{H} communément utilisé en analyse numérique pour mesurer les difficultés d'inversion d'une matrice.

Comme dans la pratique l'évaluation donnée par la borne de l'équation (D.16) s'avère tout à fait correcte, il est important de connaître *a priori* les valeurs de λ_{\max} et λ_{\min} qui renseignent sur les difficultés que l'on pourrait rencontrer dans une déconvolution particulière. Lorsque les dimensions de \mathbf{H} sont faibles, il est possible de calculer ces valeurs avec un programme informatique ; mais lorsque les dimensions augmentent, il devient de plus en plus difficile de calculer avec précisions les plus faibles valeurs singulières. Il est alors préférable d'utiliser les relations qui existent entre l'ensemble des valeurs singulières de \mathbf{H} et la réponse impulsionnelle $h(t)$.

On a vu que \mathbf{H} est une matrice de Toeplitz, et cette propriété est conservée dans le produit $\mathbf{H}^t \mathbf{H}$ notée \mathbf{R} dont le terme général est

$$r(i - j) = \sum_{k=0}^{N-1} h(k)h(k + i - j); \quad (\text{D.17})$$

il est clair que la première ligne et la première colonne de \mathbf{R} correspondent à la séquence d'autocorrélation de la réponse impulsionnelle discrétisée $h(n)$. Une telle matrice est d'ailleurs appelée matrice de corrélation. Comme les valeurs singulières de \mathbf{H} sont les valeurs propres de \mathbf{R} , on peut utiliser les résultats généraux concernant les matrices de Toeplitz. On sait en particulier que l'ensemble des valeurs propres d'une matrice de Toeplitz est l'image, par la transformation de Fourier discrète, de la séquence discrète construite à partir de sa première ligne et de sa première colonne (le principe est en fait de transformer la matrice de Toeplitz en une matrice circulante, nous y reviendrons plus tard).

Si $H(f)$ est la transformée de Fourier à temps discret de $h(n)$, alors la transformée de Fourier discrète s'exprime par

$$H(k) = \sum_{n=0}^{N-1} h(n) \exp(-2j\pi kn/N) = H(e^{2j\pi f}) \Big|_{f=k/N}; \quad (\text{D.18})$$

en conséquence, la transformée de la séquence d'autocorrélation donnée à l'équation (D.17) est $|H(k)|^2$, et l'on sait alors que

$$\lambda_i^2 \in [\min |H(k)|^2, \max |H(k)|^2], \quad i = 0, 1, \dots, M - 1. \quad (\text{D.19})$$

On ne possède pas de résultat sur les valeurs exactes prises par λ_i^2 pour des matrices de dimensions finies, mais une extension du théorème de Szegö [103, 104, 200] indique que, lorsque ces dimensions augmentent, l'ensemble des valeurs propres de \mathbf{R} se répartit approximativement comme la fonction $|H(k)|^2$ évaluée en M points équidistants, soit

$$\lambda_i = H(e^{2j\pi f}) \Big|_{f=i/M}; \quad (\text{D.20})$$

le mauvais conditionnement d'une matrice $\mathbf{H}^t \mathbf{H}$ peut donc être évalué à l'aide d'un algorithme du type FFT.

L'examen de la figure D.1 permet de comprendre comment la diminution de la fréquence d'échantillonnage accroît le mauvais conditionnement de $\mathbf{H}^t \mathbf{H}$. Un échantillonnage fin est nécessaire pour amé-

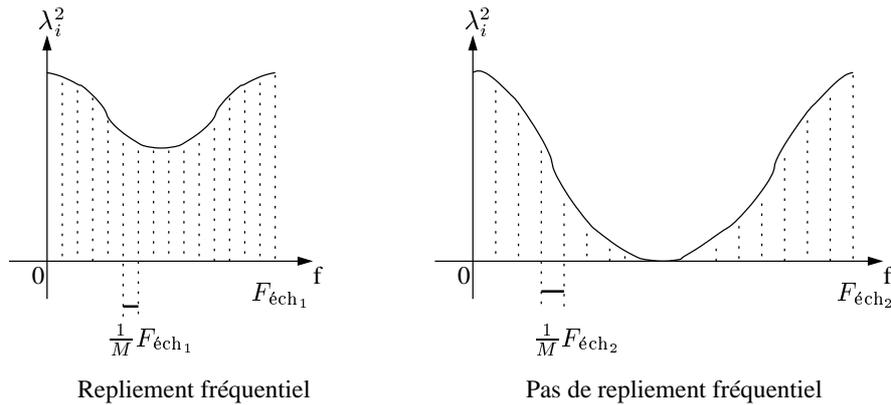


FIGURE D.1 – Variation de l'ensemble des valeurs singulières avec la fréquence d'échantillonnage

liorer la résolution, et aussi pour réduire les erreurs d'approximation d'une intégrale par une somme finie. La réponse $h(t)$ est, pour des raisons pratiques, à spectre limité et $|H(f)|$ s'annule pour $|f| \geq F_H$. Quand la fréquence d'échantillonnage augmente, $\min \lambda \rightarrow 0$ alors que $\max \lambda$ reste inchangé.

Ceci explique le paradoxe de la déconvolution : plus l'erreur d'approximation est faible, plus la solution directe est sujette à caution. Il est clair que le conditionnement du problème dépend aussi de la douceur de $|H(f)|$, c.-à-d. de sa vitesse de décroissance vers 0. Le cas des systèmes passe-bande sera donc plus délicat que celui des systèmes passe-bas.

Nous avons montré dans ce paragraphe que la déconvolution est un problème mal posé dont les difficultés sont celles des équations générales de première espèce. L'idée majeure de la plupart des nombreuses méthodes proposées dans la littérature est de construire une « solution » qui soit une approximation physiquement acceptable et qui soit suffisamment stable du point de vue du calcul.

À la différence des problèmes usuels, dans les problèmes mal posés, les propriétés des solutions acceptables ou significatives ne sont pas entièrement contenues dans l'équation de départ, et l'astuce est d'apporter une information supplémentaire sur la solution désirée afin de résoudre correctement le problème mal posé.

On désigne par *régularisation* d'un problème mal posé les différentes approches qui permettent de circonvenir le manque de dépendance continue de la solution vis-à-vis des données (et aussi d'apporter l'existence et l'unicité si nécessaire). En gros, cela revient à substituer au problème initial mal posé un problème bien posé associé dont l'analyse fournit une solution acceptable. On distingue habituellement trois aspects de la régularisation : la philosophie, les méthodes, et les algorithmes qui permettent de les mettre en œuvre.

La philosophie de la régularisation entremêle les idées suivantes :

- changement de la notion même de solution en introduisant les solutions approchées (\hat{x} : $\|H\hat{x} - y\| < \epsilon$, avec $\epsilon > 0$ fixé), c'est le cas de la solution des moindres carrés ;
- introduction d'opérateurs de régularisation ;
- modification de l'équation de départ ou du problème lui-même ;
- changement d'espaces et de topologies ;
- introduction de concepts probabilistes.

Les méthodes de régularisation sont extrêmement variées : régularisation au sens de Phillips-Twomey-Tikhonov, décomposition tronquée en valeurs singulières, méthodes de projection, etc.

Pour que la stratégie de résolution retenue soit effectivement mise en œuvre, il faut disposer d’algorithmes, ce qui nécessite en général la détermination de la valeur adéquate d’un paramètre particulier correspondant à la méthode adoptée (coefficient de régularisation, dimension du sous-espace de projection, coefficient de relaxation, etc.). Ce n’est pas un choix trivial, car il implique un compromis entre la précision et la stabilité numérique, une situation qui ne se présente pas dans les problèmes bien posés.

D.1.1.2 Inversion généralisée

On a indiqué précédemment d’un moyen de résoudre un problème mal posé consistait à changer la notion même de solution. Une démarche possible consiste à introduire, par le concept d’*inverse généralisée*, une quasi-solution, c.-à-d. une solution qui minimise $\|Hx - y\|$. Ce concept remédie au problème de la non unicité de la solution. La résolution de l’équation (D.2) présente trois possibilités :

1. Il y a une solution unique.
2. Il y a une infinité de solutions.
3. Il n’y a pas de solutions.

Dans le premier cas, il reste à calculer l’inverse habituelle H^{-1} . Dans le second cas (quand $y \in \mathcal{I}m(\mathcal{H})$), le problème est de choisir la meilleure solution, par exemple la solution de *norme minimale* ; l’opérateur H^+ qui permet de l’obtenir est appelé *inverse généralisé* de l’opérateur H . Dans le troisième cas (quand $y \notin \mathcal{I}m(\mathcal{H})$ mais le noyau de H n’est pas trivial¹), on aimerait encore définir la meilleure solution et on pourrait par exemple choisir un x qui minimise $\|Hx - y\|$, c.-à-d. une solution des *moindres carrés* ; mais comme le noyau de H n’est pas trivial, il faut encore effectuer un choix, et on peut par exemple choisir, comme dans le cas 2., le vecteur de norme minimale qui minimise $\|Hx - y\|$.

Moindres carrés On appelle $\hat{\mathbf{x}}$ une solution au sens des moindres carrés de l’équation matricielle $\mathbf{y} = \mathbf{H}\mathbf{x}$ une solution qui minimise la norme euclidienne $\|\mathbf{y} - \mathbf{H}\mathbf{x}\|^2$, c.-à-d.

$$\hat{\mathbf{x}} = \arg \min \{ \|\mathbf{y} - \mathbf{H}\mathbf{x}\|^2 \}. \quad (\text{D.21})$$

Cet ensemble coïncide avec l’ensemble des solutions de l’équation normale

$$\mathbf{H}^t \mathbf{y} = \mathbf{H}^t \mathbf{H} \hat{\mathbf{x}}. \quad (\text{D.22})$$

Rappelons que \mathbf{H} est de dimension $L \times M$ et par conséquent $\mathbf{H}^t \mathbf{H}$ de taille $M \times M$; donc

- si $L = M$ et $\text{rang}(\mathbf{H}) = M$, alors $\hat{\mathbf{x}} = \mathbf{x} = \mathbf{H}^{-1} \mathbf{y}$;
- si $L > M$ et $\text{rang}(\mathbf{H}) = M$, alors $(\mathbf{H}^t \mathbf{H})$ est de rang plein et $\hat{\mathbf{x}} = (\mathbf{H}^t \mathbf{H})^{-1} \mathbf{H}^t \mathbf{y}$;
- si $L < M$, alors $(\mathbf{H}^t \mathbf{H})$ est forcément singulière ; elle a un rang au maximum égal à L et l’équation $\mathbf{y} = \mathbf{H}\mathbf{x}$ n’a pas de solution (ou plutôt elle en a une infinité).

Inverse généralisée On appelle \mathbf{x}^+ la solution inverse généralisée de l’équation matricielle $\mathbf{y} = \mathbf{H}\mathbf{x}$ la solution de norme minimale au sens des moindres carrés de l’équation normale $\mathbf{H}^t \mathbf{y} = \mathbf{H}^t \mathbf{H} \hat{\mathbf{x}}$. Si $\mathbf{y} \in \mathcal{I}m(H)$ alors cette équation existe et elle est unique ; elle est donnée par

$$\mathbf{x}^+ = \mathbf{H}^+ \mathbf{y}, \quad (\text{D.23})$$

où \mathbf{H}^+ est l’inverse généralisée de \mathbf{H} . Deux cas se présentent :

- si $\text{rang}(\mathbf{H}) = L$, alors $\mathbf{H}^+ = \mathbf{H}^t (\mathbf{H}^t \mathbf{H})^{-1}$;
- si $\text{rang}(\mathbf{H}) = M$, alors $\mathbf{H}^+ = (\mathbf{H}^t \mathbf{H})^{-1} \mathbf{H}^t$.

1. c.-à-d. ni égal à l’ensemble vide, ni égal à tout l’espace de départ

Une matrice \mathbf{G} de taille $M \times L$ est dite inverse généralisée de la matrice \mathbf{H} au sens de Moore & Penrose si

$$\mathbf{HGH} = \mathbf{H}, \quad (\mathbf{HG})^t = \mathbf{HG}, \quad \mathbf{GHG} = \mathbf{G} \text{ et } (\mathbf{GH})^t = \mathbf{H}. \quad (\text{D.24})$$

Ces définitions sont identiques si les normes associées aux espaces sont du type euclidien. Il est facile de vérifier que l'erreur quadratique moyenne entre une estimation $\hat{\mathbf{x}}$ linéaire quelconque de la solution et \mathbf{x} est minimale si $\hat{\mathbf{x}} = \mathbf{H}^+ \mathbf{y}$ (elle est même nulle si $\mathbf{H}^+ \mathbf{H} = \mathbf{I}$).

En résumé, lorsque nous voulons calculer une solution au sens inverse généralisé de l'équation $\mathbf{y} = \mathbf{H}\mathbf{x}$ avec \mathbf{H} une matrice de taille $L \times M$, quatre situations se présentent :

- si $L = M$ et $\text{rang}(\mathbf{H}) = M$, alors $\mathbf{H}^+ = \mathbf{H}^{-1}$;
- si $L > M$ et $\text{rang}(\mathbf{H}) = M$, alors $\mathbf{H}^+ = (\mathbf{H}^t \mathbf{H})^{-1} \mathbf{H}^t$;
- si $L < M$ et $\text{rang}(\mathbf{H}) = L$, alors $\mathbf{H}^+ = \mathbf{H}^t (\mathbf{H} \mathbf{H}^t)^{-1}$;
- si $\text{rang}(\mathbf{H}) = K < \inf(L, M)$, alors il faut décomposer \mathbf{H} en K valeurs singulières, ou utiliser des méthodes itératives.

Décomposition en valeurs singulières Nous avons déjà exposé que toute matrice \mathbf{H} de taille $L \times M$ peut être séparée en plusieurs composantes orthogonales. La matrice $\mathbf{H}^t \mathbf{H}$ étant supposée réelle et symétrique de taille $M \times M$, elle possède des valeurs propres réelles $\lambda_j^2 \geq 0$, supposées ordonnées

$$\lambda_0^2 \geq \lambda_1^2 \geq \dots \lambda_{M-1}^2; \quad (\text{D.25})$$

ces valeurs propres correspondent aux carrés des valeurs singulières de la matrice \mathbf{H} , à laquelle on peut associer deux systèmes de vecteurs propres $\mathbf{u}_i, i = 0, 1, \dots, L-1$ et $\mathbf{v}_j, j = 0, 1, \dots, M-1$, tels que

$$\begin{cases} \mathbf{H}\mathbf{v}_j = \lambda_j \mathbf{u}_j \\ \mathbf{H}^t \mathbf{u}_i = \lambda_i \mathbf{v}_i \end{cases}, \quad (\text{D.26})$$

ce qui devient

$$\begin{cases} \mathbf{H}^t \mathbf{H} \mathbf{v}_j = \lambda_j^2 \mathbf{v}_j \\ \mathbf{H} \mathbf{H}^t \mathbf{u}_i = \lambda_i^2 \mathbf{u}_i \end{cases}. \quad (\text{D.27})$$

La matrice \mathbf{H} peut donc se décomposer sous la forme

$$\mathbf{H} = \mathbf{U} \mathbf{\Lambda} \mathbf{V}^t, \quad (\text{D.28})$$

avec \mathbf{U} la matrice dont les colonnes \mathbf{u}_i sont les vecteurs propres de $\mathbf{H}^t \mathbf{H}$, \mathbf{V} la matrice dont les colonnes \mathbf{v}_j sont les vecteurs propres de $\mathbf{H} \mathbf{H}^t$, et $\mathbf{\Lambda}$ la matrice diagonale formée des valeurs singulières λ_j . Les matrices \mathbf{U} et \mathbf{V} sont orthonormales, c.-à-d. que l'on a $\mathbf{U} \mathbf{U}^t = \mathbf{U}^t \mathbf{U} = \mathbf{I}$ et $\mathbf{V} \mathbf{V}^t = \mathbf{V}^t \mathbf{V} = \mathbf{I}$, avec \mathbf{I} la matrice identité.

Si $\text{rang}(\mathbf{H}) = K < \inf(L, M)$, la solution inverse généralisée au sens de Moore & Penrose est définie par

$$\mathbf{x}^+ = \mathbf{H}^+ \mathbf{y}, \text{ avec } \mathbf{H}^+ = \mathbf{V} \mathbf{\Lambda}^+ \mathbf{U}^t, \quad (\text{D.29})$$

où

$$\mathbf{\Lambda}^+ = \text{diag}(\alpha_i), \text{ avec } \begin{cases} \alpha_i = \frac{1}{\lambda_i} & \text{si } \lambda_i \neq 0 \\ \alpha_i = 0 & \text{si } \lambda_i \approx 0 \end{cases} \text{ et } i = 0, 1, \dots, K-1. \quad (\text{D.30})$$

D.1.1.3 Régularisation d'un problème mal posé

La régularisation d'un problème mal posé consiste à le transformer en un problème bien posé, c.-à-d. à définir une solution unique, et à assurer la stabilité de cette solution.

Définition d'un opérateur de régularisation Un « régulariseur » de l'équation $\mathbf{y} = \mathbf{H}\mathbf{x}$ est une famille d'opérateurs à un paramètre $\{\mathcal{R}_\alpha: \alpha \in \mathbb{A} \subset \mathbb{R}^+\}$ tels que :

- $\forall \mathbf{x} \in X, \lim_{\alpha \rightarrow 0} \|\mathcal{R}_\alpha \mathbf{H}\mathbf{x} - \mathbf{x}\| = 0$;
- $\forall \alpha \in \mathbb{A}, \mathcal{R}_\alpha$ est un opérateur continu de X dans Y .

Les problèmes mal posés de la physique sont habituellement régularisables dans ce sens. Ainsi, on peut calculer une solution approchée $\mathbf{x}_\epsilon = \mathcal{R}_\alpha \mathbf{y}_\epsilon$ à l'équation $\mathbf{y}_\epsilon = \mathbf{H}\mathbf{x}$ qui possède les propriétés

$$\lim_{\epsilon \rightarrow 0} \|\mathbf{x}_\epsilon - \mathbf{H}^{-1}\mathbf{y}\| = 0 \quad \text{et} \quad \lim_{\epsilon \rightarrow 0} \|\mathbf{H}\mathbf{x}_\epsilon - \mathbf{y}\| = 0; \quad (\text{D.31})$$

on a dans ce cas

$$\|\mathbf{x}_\epsilon - \mathbf{H}^{-1}\mathbf{y}\| \leq \|\mathcal{R}_\alpha \mathbf{y} - \mathbf{H}^{-1}\mathbf{y}\| + \|\mathcal{R}_\alpha(\mathbf{y}_\alpha - \mathbf{y})\|; \quad (\text{D.32})$$

le premier terme (erreur de régularisation) tend vers zéro quand $\alpha \rightarrow 0$, le second terme (amplification du bruit due au caractère mal posé) tend vers zéro, à α fixé, quand $\epsilon \rightarrow 0$. Régulariser consiste donc, pour un $\epsilon > 0$ donné, à choisir $\alpha = \alpha(\epsilon)$ qui minimise $\|\mathbf{x}_\epsilon - \mathbf{H}^{-1}\mathbf{y}\|$. Pour des raisons de simplicité, on impose souvent à \mathcal{R}_α d'être linéaire. Il faut remarquer que les opérateurs \mathcal{R}_α ne sont pas uniformément bornés puisque \mathbf{H}^{-1} n'est pas borné. L'erreur due au bruit tend donc vers l'infini quand $\alpha \rightarrow 0$.

L'idée de ces solutions approchées ou régularisées est donc la suivante : on accepte le fait que la résolution (au sens habituel) de l'équation $\mathbf{y} = \mathbf{H}\mathbf{x}$ soit impossible. Cette équation définit d'ailleurs, plutôt qu'une solution unique, une classe d'équivalence, l'ensemble des entrées telles que $\|\mathbf{y}_\alpha - \mathbf{H}\mathbf{x}\| < \epsilon$. Il faut donc une information supplémentaire pour choisir un élément particulier dans cette classe ; cette information supplémentaire est contenue dans le régulariseur.

Régularisation par contrôle d'une dimension Une question sous-jacente dans tout régularisation d'un problème mal posé est le choix d'une base appropriée pour représenter la solution. L'approche probablement la plus classique pour la résolution de tels problèmes consiste à rechercher une approximation de la solution par combinaison linéaire d'un petit nombre de fonctions judicieusement choisies. Le problème se réduit alors à déterminer au mieux un petit nombre de paramètres. Les méthodes de ce paragraphe montrent comment l'on peut régulariser en changeant à la fois la notion de solution et les espaces de travail, et ceci en contrôlant une dimension : le nombre de fonctions choisies. Nous verrons que ces fonctions peuvent être choisies à l'avance (TSVD) ou bien générées récursivement par filtrage répété d'une approximation initiale x_0 .

DÉCOMPOSITION TRONQUÉE EN VALEURS SINGULIÈRES

La méthode de décomposition tronquée en valeurs singulières (Truncated Singular Values Decomposition, ou TSVD), utilise les bases formées des vecteurs singuliers de \mathbf{H} . Pour un opérateur compact, le système singulier est formé des fonctions propres et des valeurs propres des opérateurs $\mathbf{H}\mathbf{H}^*$ et $\mathbf{H}^*\mathbf{H}$. L'exposant \star représente l'opérateur adjoint, c.-à-d. la matrice transposée et conjuguée. Comme déjà énoncé précédemment, on a la décomposition

$$\mathbf{H} = \mathbf{U}\mathbf{\Lambda}\mathbf{V}^*, \quad (\text{D.33})$$

avec \mathbf{U} la matrice dont les colonnes sont les vecteurs propres de $\mathbf{H}^*\mathbf{H}$, \mathbf{V} la matrice dont les colonnes sont les vecteurs propres de $\mathbf{H}\mathbf{H}^*$, et $\mathbf{\Lambda}$ la matrice diagonale formée des valeurs propres correspondantes.

Si on suppose la matrice \mathbf{H} de taille $L \times M$, avec $L \geq M$, $\mathbf{H}\mathbf{H}^*$ est de taille $M \times M$ et ses valeurs propres réelles λ^2 peuvent être ordonnées

$$\lambda_0^2 \geq \lambda_1^2 \geq \dots \lambda_{M-1}^2. \quad (\text{D.34})$$

On décide qu'à partir du rang K les valeurs singulières ordonnées ne sont plus significatives. Pour l'inversion, on utilise l'inverse généralisée au sens de Moore & Penrose de \mathbf{H} , c.-à-d.

$$\mathbf{H}^+ = \mathbf{V}\mathbf{\Lambda}^+\mathbf{U}^*, \quad (\text{D.35})$$

avec

$$\Lambda^+ = \text{diag}(\alpha_i), \text{ avec } \begin{cases} \alpha_i = \frac{1}{\lambda_i} & \text{si } \lambda_i \neq 0 & \text{significatif} \\ \alpha_i = 0 & \text{si } \lambda_i \approx 0 & \text{non significatif} \end{cases} \quad (\text{D.36})$$

L'équation initiale $\mathbf{H}\mathbf{x} = \mathbf{y}$ est donc remplacée par

$$\mathbf{H}\mathbf{x} = \bar{\mathbf{y}} = \mathbf{U}\bar{\mathbf{U}}^t \mathbf{y}, \quad (\text{D.37})$$

avec $\bar{\mathbf{U}} = (\mathbf{u}_0, \mathbf{u}_1, \dots, \mathbf{u}_{K-1}, 0, \dots, 0)$. Ceci équivaut à utiliser la matrice pseudo-inverse de $\bar{\mathbf{H}} = \mathbf{U}\bar{\mathbf{\Lambda}}\mathbf{V}^t$, où $\bar{\mathbf{\Lambda}} = \text{diag}(\lambda_0, \lambda_1, \dots, \lambda_{K-1}, 0, \dots, 0)$. Comme l'équation (D.37) est une projection sur la partie *significative* de l'espace, cette méthode fait implicitement l'hypothèse que les composantes des données correspondant aux valeurs singulières forcées à 0 sont négligeables (plus précisément $\|\mathbf{y} \cdot \mathbf{u}_j\| \ll \lambda_j$, $\forall j \geq K$). Ceci n'a aucune raison d'être vrai pour une matrice \mathbf{H} quelconque. D'autre part, la convolution étant un opérateur compact, 0 est un point d'accumulation de ses valeurs singulières ; celles-ci sont donc contenues dans une boule de rayon fini. En d'autres termes, les valeurs singulières varient de manière continue, et il n'est pas évident de définir un sous-espace significatif.

Lorsque les données sont contaminées par du bruit, l'erreur de la solution relative due au bruit \mathbf{b} s'annule si \mathbf{b} est orthogonal aux vecteurs \mathbf{u}_j , $j = 1, 2, \dots, K$; sinon, les projections non nulles de \mathbf{b} sur ces vecteurs doivent être suffisamment petites devant celles des données non bruitées \mathbf{y} .

On voit donc que cette méthode conduit à des problèmes numériquement très bien posées, sans pour autant assurer l'élimination des effets du bruit. Elle semble limitée en pratique aux problèmes de dimensions raisonnables car elle nécessite la détermination du système singulier de \mathbf{H} ; or il n'est pas envisageable d'effectuer un tel calcul sur une matrice contenant plusieurs milliers de coefficients, comme c'est le cas en acoustique des salles, par exemple. En pratique, il existe une méthode astucieuse permettant d'effectuer tout de même le calcul.

Reprenons l'équation (D.10) et posons $M = N$. L'idée est de rendre circulante la matrice de Toeplitz \mathbf{H} (qui devient une matrice carrée de taille *minimale* $(2M - 1) \times (2M - 1)$), et de compléter le vecteur \mathbf{x} au moins par $M - 1$ zéros afin de ne rien changer à l'opération de convolution. Dans ce cas, la matrice $\mathbf{H}^* \mathbf{H}$ devient elle aussi circulante.

Mais il s'agit d'une *approximation* : sous réserve que les dimensions du système soit très grande et que la réponse impulsionnelle soit à support étroit, alors on peut montrer que les matrices \mathbf{T} et \mathbf{C} ci-dessous sont *asymptotiquement* équivalentes.

Une matrice \mathbf{T} de dimensions $L \times M$ est dite de Toeplitz si

$$(T_{ij}) = t_{i-j}, \quad \text{pour } i = 0, 1, \dots, L - 1 \text{ et } j = 0, 1, \dots, M - 1 ; \quad (\text{D.38})$$

en d'autres termes, tous les éléments se trouvant sur des lignes parallèles à la première diagonale sont identiques. Si de plus \mathbf{T} est carrée et symétrique de taille $L \times L$, on a $(T_{ij}) = t_{|i-j|}$, pour $i = 0, 1, \dots, L - 1$, c.-à-d.

$$\mathbf{T} = \begin{pmatrix} t_0 & t_1 & \cdot & \cdot & t_{L-2} & t_{L-1} \\ t_1 & t_0 & \cdot & \cdot & \cdot & t_{L-2} \\ \cdot & \cdot & t_0 & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & t_0 & \cdot & \cdot \\ t_{L-2} & \cdot & \cdot & \cdot & t_0 & t_1 \\ t_{L-1} & t_{L-2} & \cdot & \cdot & t_1 & t_0 \end{pmatrix}; \quad (\text{D.39})$$

une telle matrice est entièrement déterminée par sa première ligne (ou sa première colonne).

Une matrice carrée \mathbf{C} de dimensions $L \times L$ est dite circulante à droite (ou en haut) si

$$(C_{ij}) = \begin{cases} c_{j-i} & \text{si } j - i \geq 0 \\ c_{L-j+1} & \text{si } j - i < 0 \end{cases}, \quad (\text{D.40})$$

c.-à-d.

$$\mathbf{C} = \begin{pmatrix} c_0 & c_1 & c_2 & \cdots & \cdots & c_{L-1} \\ c_{L-1} & c_0 & c_1 & \cdots & \cdots & c_{L-2} \\ c_{L-2} & c_{L-1} & c_0 & \cdots & \cdots & \cdot \\ \vdots & c_{L-2} & c_{L-1} & c_0 & \cdots & \vdots \\ \cdot & \cdots & \ddots & \ddots & c_0 & c_1 \\ c_1 & c_2 & \cdots & c_{L-2} & c_{L-1} & c_0 \end{pmatrix}. \quad (\text{D.41})$$

On peut montrer [114, chap. 7] que les valeurs propres d'une matrice circulante sont les coefficients de la transformée de Fourier discrète (TFD) de sa première ligne, et ses vecteurs propres sont les vecteurs de base de la TFD.

Appelons \mathbf{F} la matrice des vecteurs de base de la TFD, de la forme

$$\mathbf{F} = (\mathbf{p}_0 \mathbf{p}_1 \cdots \mathbf{p}_{L-1}), \quad (\text{D.42})$$

avec

$$(\mathbf{p}_k)_n = \frac{1}{\sqrt{L}} \exp\left(-2j\pi \frac{kn}{L}\right), \quad \text{pour } k = 0, 1, \dots, L-1 \text{ et } n = 0, 1, \dots, L-1; \quad (\text{D.43})$$

\mathbf{F} est une matrice qui, multipliée par un vecteur \mathbf{x} , donne un vecteur \mathbf{X} qui contient les valeurs de sa TFD. Le lecteur vérifiera que $\mathbf{F} = \mathbf{F}^*$ et que $\mathbf{F}\mathbf{F}^* = \mathbf{F}^*\mathbf{F} = \mathbf{I}$.

Appelons Λ une matrice diagonale contenant les coefficients de la TFD de la première ligne de la matrice, c.-à-d.

$$(\Lambda_{kk}) = \sum_{n=0}^{L-1} c(n) \exp\left(-2j\pi \frac{kn}{L}\right), \quad \text{pour } k = 0, 1, \dots, L-1. \quad (\text{D.44})$$

Si \mathbf{C} est une matrice circulante, alors

$$\mathbf{C} = \mathbf{F}\Lambda\mathbf{F}^{-1} = \mathbf{F}\Lambda\mathbf{F}^*. \quad (\text{D.45})$$

Dans le cas d'une opération de convolution, la matrice \mathbf{H} est normale, c.-à-d. que l'on a $\mathbf{H}\mathbf{H}^* = \mathbf{H}^*\mathbf{H}$, notée \mathbf{R} matrice de corrélation de la réponse impulsionnelle discrétisée $h(n)$, de forme Toeplitz symétrique.

Lorsque \mathbf{H} a été rendue circulante, $\mathbf{H}_c\mathbf{H}_c^*$ notée \mathbf{G} devient

$$\begin{pmatrix} r(0) & r(1) & \cdots & r(M-1) & r(M-1) & \cdots & r(1) \\ r(1) & r(0) & & r(M-2) & & & \\ \vdots & & & \vdots & & & \\ r(M-1) & & & r(0) & r(0) & \cdots & r(M-2) \\ r(M-1) & & & & & & \\ \vdots & & & & & & \\ r(1) & & & & & & r(0) \end{pmatrix}, \quad (\text{D.46})$$

avec $r(p) = \sum_{n=0}^{M-1} h(n)h(n+p)$, $p = 0, 1, \dots, M-1$; \mathbf{c} 'est une matrice circulante. On diagonalise alors cette matrice dans la base de ses vecteurs propres, c.-à-d. $\mathbf{G}\mathbf{V} = \mathbf{V}\mathbf{D}$ avec \mathbf{V} la matrice des vecteurs propres (vecteurs de base de la TFD) et \mathbf{D} la matrice diagonale des valeurs propres réelles (TFD de la première ligne de \mathbf{G}); puis on calcule $\Lambda = \sqrt{\mathbf{D}}$, et on définit un seuil en dessous duquel les valeurs propres seront mises à 0. Enfin comme on veut $\mathbf{H}^+ = \mathbf{V}\Lambda^+\mathbf{U}^*$, on calcule $\mathbf{U}^* = \Lambda^{-1}(\mathbf{H}_c\mathbf{V})^*$ puisque l'on a $\mathbf{H}_c = \mathbf{U}\Lambda\mathbf{V}^*$, avec $\mathbf{V}^{-1} = \mathbf{V}^*$ car \mathbf{V} est orthonormale.

Comme \mathbf{H} a été rendue circulante, de taille minimale $(2M - 1) \times (2M - 1)$, \mathbf{H}^+ est aussi de taille minimale $(2M - 1) \times (2M - 1)$, et par conséquent la longueur minimale du filtre inverse est de $2M - 1$.

Notons qu'il n'est pas nécessaire de calculer entièrement la matrice \mathbf{H}^+ ; seule la première ligne suffit, puisque c'est aussi une matrice circulante. D'où on peut se contenter de calculer la première ligne de \mathbf{V} et la première colonne de \mathbf{U}^* puisque Λ^+ est diagonale. Plus en détail encore, le calcul de la première colonne de \mathbf{U}^* peut être effectué facilement puisque \mathbf{H}_c est définie par sa première ligne, et d'autre part \mathbf{V} est constitué de racines complexes de l'unité, calculables *ex nihilo*. À aucun moment, il ne sera procédé au stockage en mémoire d'une matrice complète, ce qui permet d'envisager l'inversion de réponses impulsionnelles très longues.

On vérifie que l'on a bien $\mathbf{G}\mathbf{V} = \mathbf{V}\mathbf{D}$, $\mathbf{G}\mathbf{U} = \mathbf{U}\mathbf{D}$, $\mathbf{H}_c\mathbf{V} = \mathbf{U}\sqrt{\mathbf{D}}$ et $\mathbf{H}_c^*\mathbf{U} = \mathbf{V}\sqrt{\mathbf{D}}$. Au passage, on remarque que l'on a $\mathbf{H}_c\mathbf{H}_c^* = \mathbf{U}\Lambda\mathbf{U}^*$ et $\mathbf{H}_c^*\mathbf{H}_c = \mathbf{V}\Lambda\mathbf{V}^*$; or la matrice \mathbf{H} est normale, donc deux bases de vecteurs propres associées à un seul ensemble de valeurs propres conduisent à la même diagonalisation. D'ailleurs, les valeurs propres de \mathbf{G} sont dégénérées : elles sont doubles. Ceci est cohérent avec la forme de \mathbf{G} : c'est une matrice circulante dans les deux sens (à droite et à gauche), elle est donc de Toeplitz et de Hankel, ou encore centrosymétrique ; sa première ligne étant réelle et paire, sa TFD sera également réelle et paire, ce qui explique la dégénérescence des valeurs propres de \mathbf{G} . Chaque valeur propre définit ainsi un plan propre, complètement caractérisé par deux vecteurs.

AMORTISSEMENT DES VALEURS SINGULIÈRES

Au lieu de diminuer le nombre de valeurs singulières, on peut les amortir. Dans le cas de la déconvolution, la matrice \mathbf{H} est de Toeplitz ; on la suppose de plus carrée de taille N (sinon le raisonnement s'applique à l'équation normale). Soit \mathbf{P} la matrice carrée de taille N dont tous les éléments sont nuls sauf ceux de la diagonale secondaire formée de 1. Résoudre $\mathbf{y} = \mathbf{H}\mathbf{x}$ est donc équivalent à résoudre

$$\mathbf{y} = \mathbf{A}\mathbf{z} \quad \text{où } \mathbf{A} = \mathbf{H}\mathbf{P} \text{ et } \mathbf{z} = \mathbf{P}\mathbf{x}, \quad (\text{D.47})$$

puisque $\mathbf{P}^2 = \mathbf{I}$; ceci revient en fait à retourner l'ordre des inconnues dans \mathbf{x} . Comme \mathbf{H} est de Toeplitz, \mathbf{A} est symétrique et possède donc un système propre orthogonal complet \mathbf{v}_j , $j = 1, 2, \dots, N$. Si les valeurs propres λ_j sont distinctes et non nulles, on peut écrire la solution avec des données bruitées

$$\mathbf{z}_\epsilon = \mathbf{V}\Lambda\mathbf{V}^t(\mathbf{y} + \mathbf{b}) = \sum_{j=1}^N \beta_j \mathbf{v}_j \quad \text{où } \beta_j = \frac{\mathbf{y}^t \mathbf{v}_j + \mathbf{b}^t \mathbf{v}_j}{\lambda_j}, j = 1, 2, \dots, N. \quad (\text{D.48})$$

Il est clair que l'erreur dans la solution dépend de $\mathbf{b}^t \mathbf{v}_j$ et de λ_j . Pour réduire cette erreur, on introduit les facteurs d'amortissement $w_j(\alpha)$ tels que

$$\mathbf{z}_{\epsilon, \alpha} = \sum_{j=1}^N w_j(\alpha) \beta_j \mathbf{v}_j \quad \text{avec } w_j(\alpha) = \frac{|\lambda_j|}{|\lambda_j| + \alpha_j}, \quad (\text{D.49})$$

ce qui revient à choisir

$$\mathbf{z}_{\epsilon, \alpha} = \sum_{j=1}^N \left(\frac{\mathbf{v}_j \mathbf{u}_j^t(\mathbf{y} + \mathbf{b})}{\sigma_j + \alpha_j} \right) \quad \text{où } (\sigma_j) = \Sigma = \Lambda \cdot \text{signe}(\Lambda). \quad (\text{D.50})$$

On effectue un amortissement des valeurs singulières, plutôt qu'une troncature brutale. On peut choisir les coefficients d'amortissement α_j tels que $\alpha_j = \alpha = \text{constante}$; ceci revient à perturber la matrice \mathbf{A} par $\alpha\mathbf{I}$, mais si α est faible, cela risque de ne pas améliorer sensiblement le mauvais conditionnement du

problème, et si α est élevé, on peut avoir une mauvaise approximation. Il est préférable de choisir, pour les α_j des multiples des variations premières des composantes des vecteurs propres \mathbf{j} , c.-à-d.

$$\alpha_j = \sum_{i=1}^{N-1} \left(\frac{\mathbf{v}_j(i+1) - \mathbf{v}_j(i)}{N} \right)^2 ; \quad (\text{D.51})$$

cette expression ressemble à une norme.

Cette approche est intéressante car elle permet d'étudier directement les effets du bruit dans le domaine fréquentiel ; mais, elle ne fournit pas à coup sûr un problème numériquement bien posé comme avec la TSVD.

RÉGULARISATION PAR MÉTHODES ITÉRATIVES

Les méthodes d'approximation successives, ou du gradient, peuvent être utilisées pour construire des méthodes de résolution approchée de problèmes mal posés, tels que les équations intégrales de première espèce. À chaque fois qu'on utilise une méthode itérative pour la résolution numérique d'un problème, une règle d'arrêt doit être fixée pour terminer la procédure. Dans le cas d'un problème mal posé, la terminaison est compliquée par l'instabilité numérique du problème. Si les mesures sont contaminées par du bruit, c.-à-d. $y_\epsilon = y + b$ avec $\|b\| < \epsilon$, il est naturel d'établir une règle d'arrêt fondée sur l'évolution du résidu $\|\mathbf{H}\hat{\mathbf{x}}_i - \mathbf{y}_\epsilon\|$. Le seuil d'arrêt $\delta > 0$ étant choisi au début des itérations, le rang k de l'itération finale est déterminé par les conditions

$$\|\mathbf{H}\hat{\mathbf{x}}_i - \mathbf{y}_\epsilon\| > \delta \quad \text{pour } i < k, \text{ et } \|\mathbf{H}\hat{\mathbf{x}}_k - \mathbf{y}_\epsilon\| \leq \delta ; \quad (\text{D.52})$$

si le seuil $\delta = \delta(\epsilon)$ est choisi tel que $\delta(\epsilon) > \epsilon$ et $\delta(\epsilon) \rightarrow 0$ quand $\epsilon \rightarrow 0$, alors on établit que

$$\lim_{\epsilon \rightarrow 0, \|b\| < \epsilon} \sup \|\hat{\mathbf{x}}_k - \mathbf{H}^+ \mathbf{y}\| = 0, \quad (\text{D.53})$$

où $\hat{\mathbf{x}}_k = \hat{\mathbf{x}}_k(\delta, b, \hat{\mathbf{x}}_0)$ est la solution approchée obtenue avec le seuil δ . Ceci montre que, dans ces conditions, la procédure itérative est une méthode de régularisation au sens défini au début de ce paragraphe ; dans ce cas, bien sûr, $1/k$ joue le rôle du paramètre de régularisation.

Algorithme de régularisation La méthode présentée ici permet de calculer une solution régularisée au problème de la déconvolution avec un nombre d'opérations élémentaires et un encombrement mémoire raisonnables. La solution a la forme générale

$$\hat{\mathbf{x}} = (\mathbf{H}^t \mathbf{H} + \alpha \mathbf{D})^{-1} \mathbf{H}^t \mathbf{y} ; \quad (\text{D.54})$$

l'inversion directe de la matrice $(\mathbf{H}^t \mathbf{H} + \alpha \mathbf{D})$ devient très rapidement impraticable lorsque les dimensions du problème augmentent, ce qui est le cas dans notre étude, puisque les réponses impulsionnelles à inverser peuvent atteindre plusieurs milliers de points ; la taille mémoire nécessaire à l'inversion d'une telle matrice rend le processus totalement irréalisable.

On contourne cette difficulté en utilisant les propriétés des valeurs propres et des vecteurs propres des matrices circulantes ; il s'agit de la méthode de Hunt qui est une méthode de moindres carrés sous contraintes. Il faut déterminer $\hat{\mathbf{x}}$ qui minimise

$$J_\alpha(\mathbf{x}) = (\mathbf{y} - \mathbf{H}\mathbf{x})^t (\mathbf{y} - \mathbf{H}\mathbf{x}) + \alpha (\mathbf{C}\mathbf{x})^t (\mathbf{C}\mathbf{x}) \quad (\text{D.55})$$

sous la contrainte résiduelle

$$(\mathbf{y} - \mathbf{H}\mathbf{x})^t (\mathbf{y} - \mathbf{H}\mathbf{x}) = \epsilon^2 \quad \text{minimal}, \quad (\text{D.56})$$

où \mathbf{C} est une matrice correspondant à une opération de filtrage linéaire, donc de convolution par une réponse impulsionnelle. La solution régularisée a pour expression

$$\hat{\mathbf{x}} = (\mathbf{H}^t \mathbf{H} + \alpha \mathbf{C}^t \mathbf{C})^{-1} \mathbf{H}^t \mathbf{y}; \quad (\text{D.57})$$

la valeur du coefficient de régularisation α doit être recherchée par itérations jusqu'à ce que la condition (D.56) soit satisfaite. Ceci est une raison supplémentaire de disposer d'un algorithme rapide de résolution de (D.57). La matrice $(\mathbf{H}^t \mathbf{H} + \alpha \mathbf{C}^t \mathbf{C})$ est symétrique mais, dans le cas général, elle n'a pas d'autres propriétés remarquables. Dans le cas particulier où \mathbf{H} et \mathbf{C} sont des matrices de Toeplitz (ce qui est le cas lors d'une opération de convolution), cette matrice est aussi de Toeplitz, de plus elle est symétrique. L'idée est alors d'étendre les vecteurs \mathbf{x} , \mathbf{h} , \mathbf{c} et \mathbf{y} par des zéros, afin d'obtenir des matrices \mathbf{H} et \mathbf{C} *circulantes* que l'on peut diagonaliser par transformée de Fourier discrète (cf. éq. (D.45)).

L'équation (D.57) peut alors se mettre sous la forme

$$\mathbf{F}^{-1} \hat{\mathbf{x}} = (\mathbf{\Lambda}_H^* \mathbf{\Lambda}_H + \alpha \mathbf{\Lambda}_C^* \mathbf{\Lambda}_C)^{-1} \mathbf{\Lambda}_H^* \mathbf{F}^{-1} \mathbf{y}; \quad (\text{D.58})$$

comme la matrice $(\mathbf{\Lambda}_H^* \mathbf{\Lambda}_H + \alpha \mathbf{\Lambda}_C^* \mathbf{\Lambda}_C)$ est diagonale et que les produits $\mathbf{F}^{-1} \hat{\mathbf{x}}$ et $\mathbf{\Lambda}_H^* \mathbf{F}^{-1} \mathbf{y}$ sont des vecteurs, cette équation peut être considérée comme le résultat de produits termes à termes de séquences

$$\widehat{X}(k) = \frac{H^*(k)}{|H(k)|^2 + \alpha |C(k)|^2} Y(k), \quad k = 0, 1, \dots, L-1, \quad (\text{D.59})$$

où $H(k)$, $C(k)$ et $Y(k)$ sont les TFD des séquences $h(n)$, $y(n)$ et $c(n)$ habilement prolongées par des zéros, et l'exposant $*$ représente le complexe conjugué. La solution $\widehat{x(n)}$ est enfin obtenue par TFD inverse.

D.1.2 Application à l'inversion de réponses impulsionnelles acoustiques

D.1.2.1 Approche directe

On considère que la réponse impulsionnelle du filtre à inverser peut se mettre sous la forme $h(nT)$, avec $0 \leq n < N$ et T la période d'échantillonnage ; dans la suite on suppose que $T = 1$.

La transformée en z de la réponse impulsionnelle h s'écrit

$$H(z) = h(0) + h(1)z^{-1} + h(2)z^{-2} + \dots + h(N-1)z^{N-1}. \quad (\text{D.60})$$

Comme $h(n)$ est une réponse impulsionnelle finie, $H(z)$ peut être décrite par ses zéros dans le plan complexe (tous ses pôles sont à l'origine). Soit les zéros sont tous à l'intérieur du cercle unité (et le filtre est à phase minimale), soit il existe des zéros sur le cercle unité ou à l'extérieur du cercle unité. Lorsqu'il existe un zéro sur le cercle unité, le filtre inverse n'est pas défini à la fréquence correspondante ; nous excluons ce cas de notre réflexion. Le filtre inverse de $h(n)$ est le signal $g(n)$ tel que

$$h(n) * g(n) = \delta(n), \quad (\text{D.61})$$

où $*$ représente l'opération de convolution, et $\delta(n)$ le signal impulsion-unité. La transformée en z de $g(n)$ s'écrit

$$G(z) = \frac{1}{H(z)}; \quad (\text{D.62})$$

les zéros de $H(z)$ deviennent les pôles de $G(z)$. Pour avoir un filtre stable, il est nécessaire que le domaine de convergence de la transformée en z inclue le cercle unité ; les pôles à l'intérieur du cercle unité correspondent à une solution exponentielle causale dans le domaine temporel, tandis que les pôles à l'extérieur du cercle unité correspondent à une solution exponentielle non causale. Donc le filtre inverse

$g(n)$ peut être rendu stable, mais alors infiniment long et son support temporel dépend de la distribution des zéros de $H(z)$.

Pour obtenir un filtre de longueur finie qui approxime le véritable filtre inverse, on peut effectuer une troncature à l'aide d'une fenêtre $w(n)$, c.-à-d.

$$f(n) = g(n) \cdot w(n), \text{ avec } w(n) = \begin{cases} 1 & \text{pour } M_1 \leq n < M_2 \\ 0 & \text{autrement} \end{cases} ; \quad (\text{D.63})$$

pour une longueur donnée du filtre inverse M égale à $M_2 - M_1 + 1$, on choisit M_1 de sorte que l'énergie du signal $g(n)$ à l'intérieur de la fenêtre soit maximal. En général, on utilise une fenêtre rectangulaire, ce qui permet de réaliser une minimisation au sens des moindres carrés de l'erreur entre $F(e^{2j\pi f})$ et $G(e^{2j\pi f})$.

Si $M_1 < 0$, le filtre $f(n)$ n'est pas causal, et dans le cadre d'une implantation en temps réel, il est nécessaire d'introduire un retard de M_1 échantillons.

Décomposons le phénomène : toute transformée en z peut être scindée en une partie à phase minimale et une partie passe-tout (cf. § C.1.3), c.-à-d.

$$H(z) = H_{\min}(z)H_{\text{pt}}(z). \quad (\text{D.64})$$

On rappelle que $|H(z)| = |H_{\min}(z)|$, puisque $|H_{\text{pt}}(z)| = 1$.

Si on appelle a_i les zéros de $H(z)$ tels que

$$\begin{cases} |a_i| < 1 & \text{pour } i = 1, 2, \dots, k \\ |a_i| > 1 & \text{pour } i = k + 1, k + 2, \dots, N \end{cases}, \quad (\text{D.65})$$

alors on peut écrire

$$\begin{cases} H_{\min}(z) &= \prod_{i=1}^k (1 - a_i z^{-i}) \\ H_{\text{pt}}(z) &= \prod_{i=k+1}^N (1 - a_i z^{-i}) \end{cases}. \quad (\text{D.66})$$

La partie à phase minimale peut être inversée directement en prenant

$$G_{\min}(z) = \frac{1}{H_{\min}(z)} = \frac{1}{\prod_{i=1}^k (1 - a_i z^{-i})}, \quad (\text{D.67})$$

puisque cette fonction est stable (tous ses pôles sont à l'intérieur du cercle unité). Si on s'en tient là, on a

$$|G(z)| = \frac{1}{|H_{\min}(z)|}; \quad (\text{D.68})$$

mais il est clair que cette égalisation ne peut réaliser une déconvolution parfaite dans le domaine temporel, en accord avec l'équation (D.62) ; il restera un résidu correspond à la partie passe-tout, qui a –certes– un spectre uniforme, mais contient la plupart de l'énergie temporelle de la réverbération. La composante passe-tout d'une réponse impulsionnelle acoustique est perçue comme une réverbération et un « étallement » des composantes transitoires d'un signal passé au travers, mais sans la coloration souvent associée à ces distorsions [263].

La partie à phase maximale ne peut être inversée directement, puisque cela conduit à une fonction instable. Il faut donc recourir à une approximation, consistant en une troncature et un décalage, comme expliqué précédemment.

TECHNIQUE CEPSTRALE

On peut effectuer une décomposition homomorphique² de la transformée en z de la réponse impulsionnelle pondérée $h(n) \exp -\lambda n$ en une partie à phase minimale et une partie à phase maximale [225, 242] ; une fois la décomposition effectuée, la partie à phase minimale est inversée de manière exacte, et la partie à phase maximale est inversée de manière approchée. Cependant, il est apparu que cette méthode était très sensible à la valeur de λ , et pire encore à la réponse impulsionnelle elle-même [221], bien qu'elle permette d'obtenir directement une approximation de durée finie du filtre inverse, et surtout une valeur optimale du retard à l'aide d'un algorithme récursif [293, chap. 2].

TECHNIQUE DES MOINDRES CARRÉS

Une autre approche s'est révélée plus efficace et plus robuste, basée sur la technique des moindres carrés [223, 225]. Brièvement, on désire créer un filtre causal $g(n)$ de longueur finie qui, convolué avec $h(n)$, produira un signal de sortie $y(n)$ qui sera la meilleure approximation au sens des moindres carrés du signal impulsion-unité ; il s'agit donc de minimiser la fonction d'erreur I telle que

$$I = \frac{1}{L} \sum_{n=0}^L e^2(n) = \sum_{n=0}^L [\delta(n) - \hat{\delta}(n)]^2, \quad (\text{D.69})$$

où $\hat{\delta}(n) = \sum_{i=0}^{M-1} f(i)h(n-i)$ avec N la longueur du filtre direct, M la longueur du filtre inverse et $L = N + M - 1$ la longueur du signal résultat. L'avantage de cette approche est qu'elle permet de prendre en compte les propriétés de la phase non minimale de $h(n)$ pour introduire un retard de p échantillons en sortie ; la fonction d'erreur devient dans ce cas I_p telle que

$$I_p = \frac{1}{L} \sum_{n=0}^L [\delta(n-p) - \hat{\delta}(n)]^2. \quad (\text{D.70})$$

On peut également écrire dans le domaine fréquentiel [223]

$$\mathcal{I} = \left[\frac{1}{L} \sum_{n=0}^L \left(10 \log |\hat{\Delta}(k)| - \frac{1}{L} \sum_{n=0}^L 10 \log |\hat{\Delta}(k)| \right)^2 \right]^{\frac{1}{2}}, \quad (\text{D.71})$$

où $\hat{\Delta}(k)$ représente la TFD de $\hat{\delta}(n)$ pour les fréquences discrètes k . Plus la fonction d'erreur est petite, meilleure est l'inversion. En pratique, deux facteurs déterminent le succès de l'inversion :

1. la longueur M du filtre inverse, mais on a toujours tendance à la choisir la plus faible possible pour garantir une implantation en temps réel la plus efficace possible.
2. le désaccord entre la réponse impulsionnelle utilisée pour calculer le filtre inverse, et la véritable réponse impulsionnelle du système étudié ; ce facteur est surtout sensible dans le cas de la déréverbération d'un signal, car la réponse impulsionnelle d'une salle varie avec les positions et les orientations de la source et du récepteur [222].

Lorsqu'un signal de longueur fini $f(n)$ est l'inverse d'un filtre $h(n)$ possédant une composante à phase non minimale, on peut écrire

$$h(n) * f(n) = \delta(n-p), \quad (\text{D.72})$$

2. transformation d'un produit en somme, traditionnellement réalisé à l'aide de la fonction log

où $*$ représente l'opération de convolution, et $\delta(n)$ le signal impulsion-unité. Matriciellement, cette équation peut être réécrite

$$\begin{pmatrix} 0 \\ \vdots \\ 0 \\ 1 \\ 0 \\ \vdots \\ \vdots \\ 0 \end{pmatrix} = \begin{pmatrix} h(0) & 0 & \dots & \dots & 0 & 0 \\ h(1) & h(0) & 0 & & & \\ \vdots & h(1) & \ddots & & & \\ \vdots & \vdots & & \ddots & 0 & \\ h(N-1) & h(N-2) & \dots & \dots & h(0) & 0 \\ 0 & h(N-1) & \dots & \dots & h(1) & h(0) \\ \vdots & \vdots & \ddots & & \vdots & \vdots \\ \vdots & \vdots & & \ddots & \vdots & \vdots \\ 0 & 0 & \dots & 0 & h(N-1) & h(N-2) \end{pmatrix} \begin{pmatrix} f(0) \\ f(1) \\ \vdots \\ \vdots \\ \vdots \\ f(M-1) \end{pmatrix}, \quad (\text{D.73})$$

ou de manière plus concise $\mathbf{d} = \mathbf{H}\mathbf{f}$, avec \mathbf{d} de taille $L \times 1$, \mathbf{H} de taille $L \times M$, et \mathbf{f} de taille $M \times 1$; on suppose implicitement que $M \geq N$, et l'on a pris $M = N + 1$ ici. Comme le nombre de colonnes de \mathbf{H} est inférieur à son nombre de lignes, l'équation (D.73) n'admet pas de solution. L'idée est donc de former la fonction d'erreur

$$\mathbf{e} = \mathbf{d} - \mathbf{H}\mathbf{f}, \quad (\text{D.74})$$

et de calculer \mathbf{H} telle que le carré de la norme de la fonction d'erreur, c.-à-d. $\|\mathbf{e}\|^2 = \mathbf{e}^t \mathbf{e}$, soit minimal; cela conduit à l'expression

$$\mathbf{f} = (\mathbf{H}^t \mathbf{H})^{-1} \mathbf{H}^t \mathbf{d}, \quad (\text{D.75})$$

qui n'est autre que l'expression de la matrice inverse généralisée de \mathbf{H} au sens de Moore & Penrose. On remarquera que :

- $\mathbf{H}^t \mathbf{H}$ est la matrice d'autocorrélation de $h(n)$, de forme Toeplitz symétrique, de taille $M \times M$, et notée \mathbf{R} ;
- $\mathbf{H}^t \mathbf{d}$ est l'intercorrélacion entre le filtre à inverser et le signal désiré, ou encore la réponse impulsionnelle du filtre original retourné et décalé, notée $\tilde{\mathbf{h}}$.

On peut montrer [202] que, quel que soit la longueur M du filtre inverse, le carré de la norme de la fonction d'erreur ne tend plus vers 0 dès lors que le filtre direct possède une composante à phase non minimale, mais plutôt

$$\lim_{L \rightarrow \infty} \|\mathbf{e}_L\|^2 = 1 - \left(\frac{h_{N-1}}{h_0} \right)^2; \quad (\text{D.76})$$

ce qui veut dire que l'approximation au sens des moindres carrés connaît bien une limite.

L'équation (D.75) peut se mettre sous la forme $\mathbf{R}\mathbf{f} = \tilde{\mathbf{h}}$, et se résout de manière efficace grâce à l'algorithme bien connu de Levinson [99, 272] [264, chap. 2]. Cet algorithme est récursif; seule la première ligne de la matrice \mathbf{R} doit être stockée en mémoire: il est donc possible d'effectuer le calcul pour des filtres très long sur un ordinateur standard. Le coût de calcul est de $4M^2$ opérations pour un filtre inverse de longueur M . Cet algorithme a été implanté sous forme d'un objet externe pour Matlab[®] programmé en C.

La figure D.2 présente le résultat d'une telle démarche effectuée sur une réponse impulsionnelle d'oreille mesurée en chambre sourde par Gardner sur le mannequin KEMAR [92]; la longueur du filtre direct est de 128 points. Nous avons fait varier la longueur du filtre inverse, et nous avons calculé le carré de la norme de l'erreur; les résultats sont rassemblés dans le tableau D.1. On constate que, même pour une longueur du filtre inverse égale à celle du filtre direct, la réjection est de l'ordre de 30 dB.

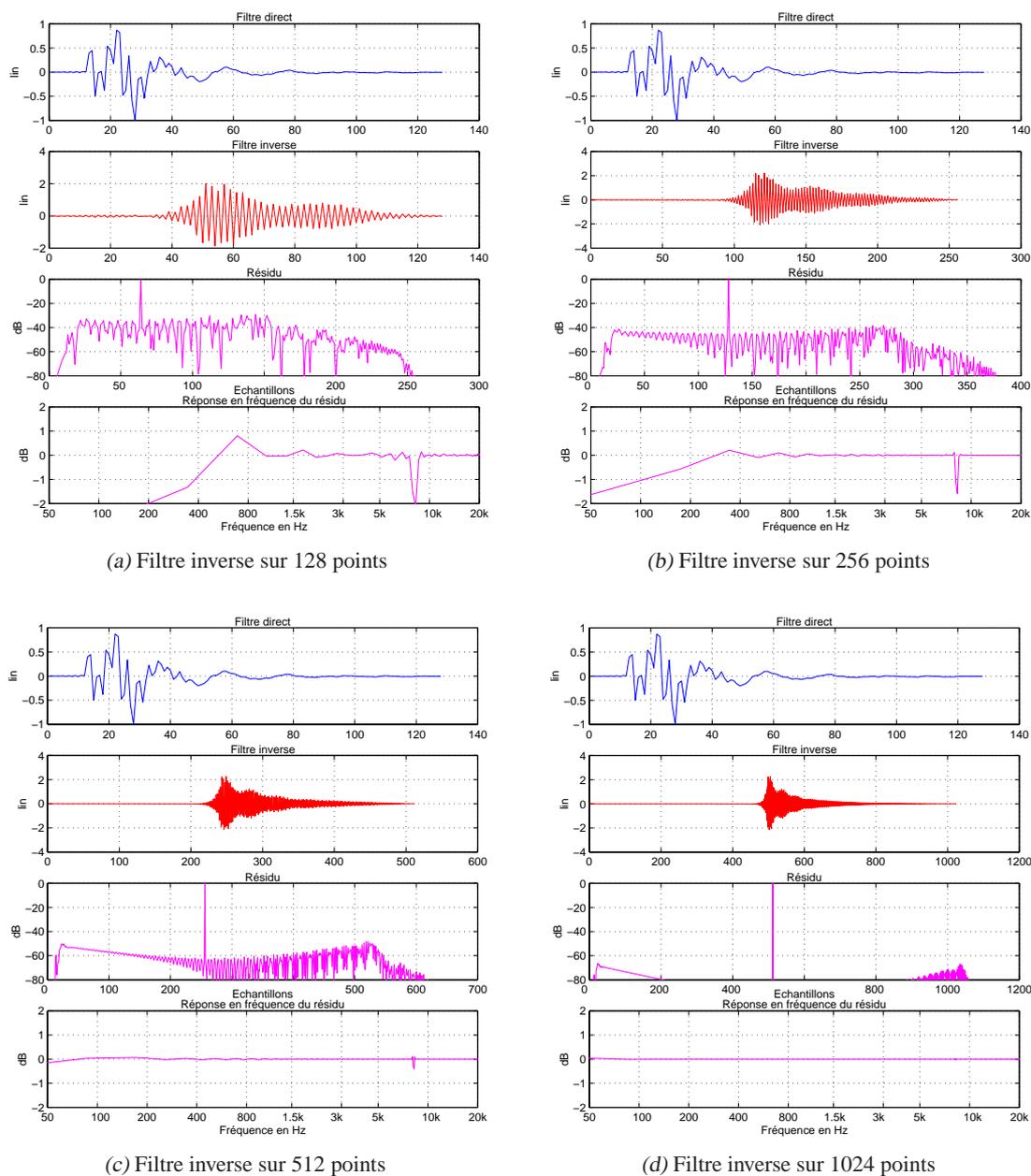


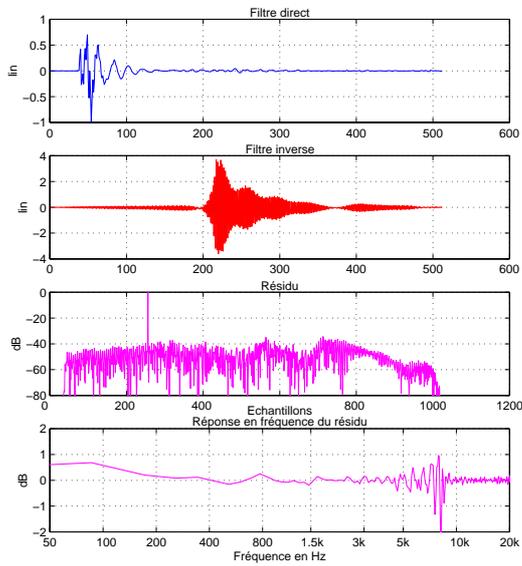
FIGURE D.2 – Inversion approximée au sens des moindres carrés d'une réponse impulsionnelle d'oreille mesurée sur 128 échantillons

TABLEAU D.1 – Précision de l'approximation dans le cas de l'inversion présentée figure D.2

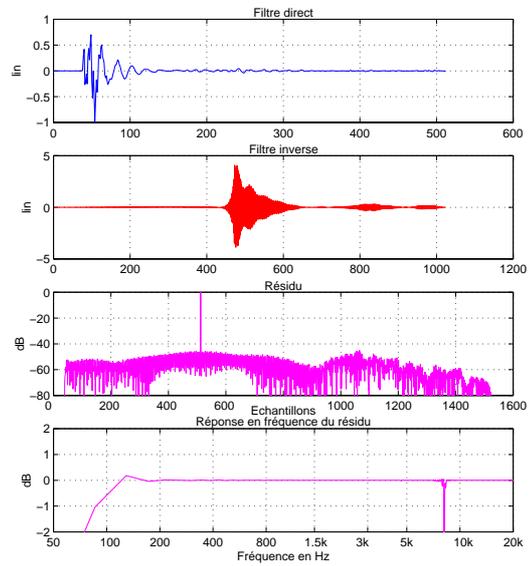
Longueur du filtre inverse	128	256	512	1024
Carré de la norme de l'erreur	0,029160	0,006174	0,000790	0,000015

TABLEAU D.2 – Précision de l'approximation dans le cas de l'inversion présentée figure D.3

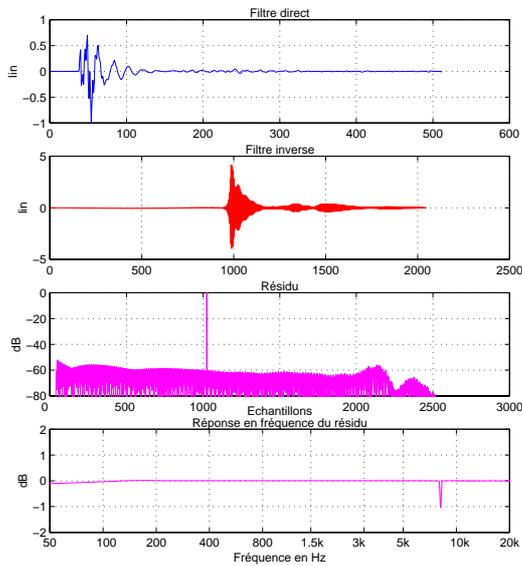
Longueur du filtre inverse	512	1024	2048	4096
Carré de la norme de l'erreur	0,022585	0,005418	0,001030	0,000357



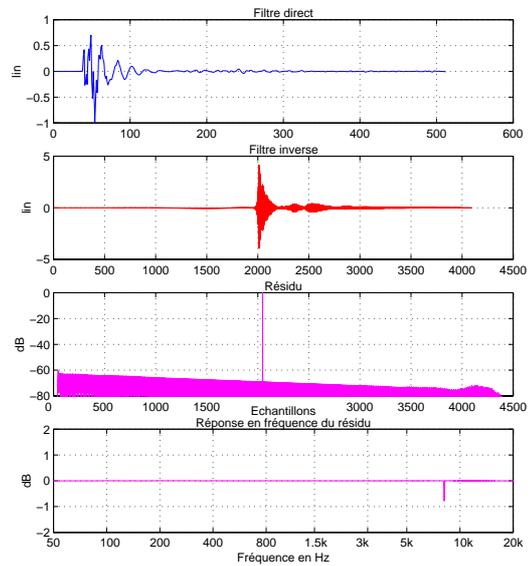
(a) Filtre inverse sur 512 points



(b) Filtre inverse sur 1024 points



(c) Filtre inverse sur 2048 points



(d) Filtre inverse sur 4096 points

FIGURE D.3 – Inversion approximée au sens des moindres carrés d'une réponse impulsionnelle d'oreille mesurée sur 512 échantillons

TABLEAU D.3 – Précision de l'approximation dans le cas de l'inversion présentée figure D.4

Longueur du filtre inverse	2048	4096	8192	16384
Carré de la norme de l'erreur	0,035617	0,008843	0,003071	0,001015

La figure D.3 présente le même genre de résultats pour une réponse impulsionnelle d'oreille mesurée par nos soins en chambre sourde, sur 512 points ; le carré de la norme de l'erreur est présenté dans le tableau D.2. La diminution de l'erreur est moins sensible que précédemment.

La figure D.4 présente toujours le même genre de résultats pour une réponse impulsionnelle mesurée dans un véhicule de type Safrane entre le haut-parleur avant gauche et un micro omnidirectionnel placé approximativement au centre de la tête du conducteur. Nous avons conservé les 2048 premiers points, ce qui correspond approximativement à 50 ms, durée au-delà de laquelle le bruit de mesure commence à se faire sentir. Le carré de la norme de l'erreur est présenté dans le tableau D.3. La diminution de l'erreur est encore moins sensible que précédemment, ce qui montrerait que plus le filtre direct est long, plus le filtre inverse doit comporter de points pour atteindre un certain résultat.

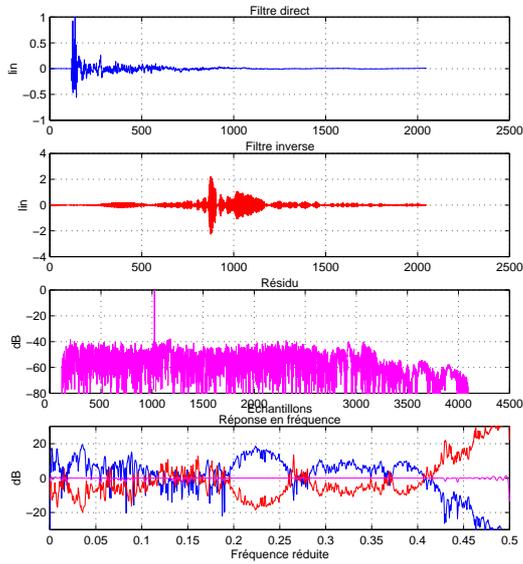
L'équation (D.75) peut aussi être résolue dans le domaine fréquentiel [157], c.-à-d.

$$F(e^{2j\pi f}) = \frac{H^*(e^{2j\pi f})}{|H(e^{2j\pi f})|^2}, \text{ où} \quad (\text{D.77})$$

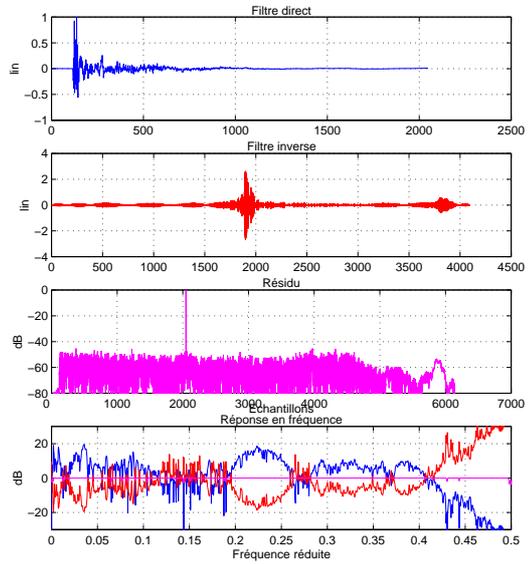
$H(e^{2j\pi f})$ représente la transformée de Fourier à temps discret de la séquence $h(n)$, et l'exposant * le passage au complexe conjugué. Le coût de calcul de l'opération s'en trouve considérablement réduit, mais l'opération n'est théoriquement exacte que dans le domaine fréquentiel continu. Or ce domaine n'est pas accessible sur un calculateur numérique, et l'échantillonnage inévitable du domaine fréquentiel provoque une erreur d'autant plus grande que le pas d'échantillonnage est élevé. Pour une longueur donnée du filtre inverse, son efficacité est considérablement réduite. D'autre part, nous subissons ici de manière *inévitabile* les effets du repliement temporel, associé au calcul d'un filtre dans le domaine fréquentiel discret ; il est possible de le minimiser en mettant en œuvre une technique de régularisation par filtrage (voir ci-après).

Reprenons notre réponse impulsionnelle d'oreille mesurée sur 512 points, et appliquons la formule ci-dessus (cf. fig. D.5). Pour une longueur du filtre inverse de 1024 points, le carré de l'erreur résiduelle est de 0,335786, valeur qui doit être comparée avec celle du tableau D.2 dans le domaine temporel : l'erreur dans le domaine fréquentiel tombe à une valeur comparable à celle du domaine temporel pour une longueur quadruple du filtre inverse.

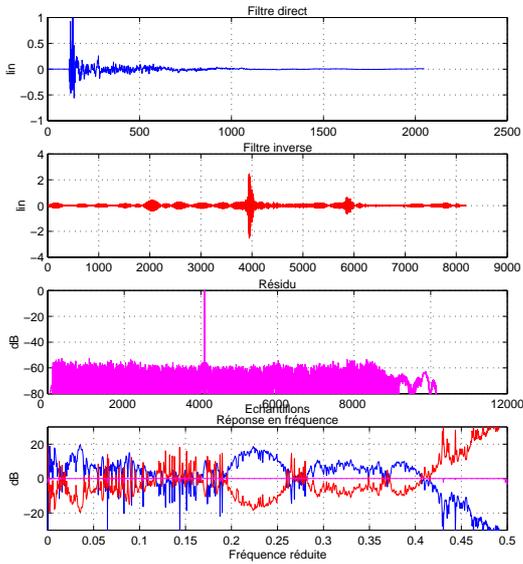
Aussi précise que soit l'inversion, d'un point de vue mathématique, elle n'en est pas pour le moins satisfaisante, d'un point de vue perceptif. En effet, si la réponse en fréquence du filtre original présente des anti-résonances trop marquées, celles-ci se transforment en des résonances excessives sur le filtre inverse (cf. fig. D.6). De plus, si la partie du spectre située en hautes fréquences présente une énergie faible (ce qui est souvent le cas à cause des filtres anti-aliasing), le filtre inverse aura une amplitude maximale dans cette zone de fréquences inaudibles, et une amplitude très faible dans les fréquences inférieures ; cela peut conduire à des problèmes de dynamique de quantification du signal filtré, puisque celui-ci peut être masqué par le bruit de quantification. Pour ces raisons, il est nécessaire que la réponse impulsionnelle originale subisse un traitement avant le processus d'inversion, consistant à « limiter » celui-ci : on parle de *régularisation*. Ce traitement est délicat à mettre en œuvre, et requiert une certaine expérience, car il est fortement dépendant du filtre original. Néanmoins, il est relativement facile d'écouter le résultat de l'inversion par le biais d'une convolution rapide, notre oreille restant le meilleur (et définitif) juge.



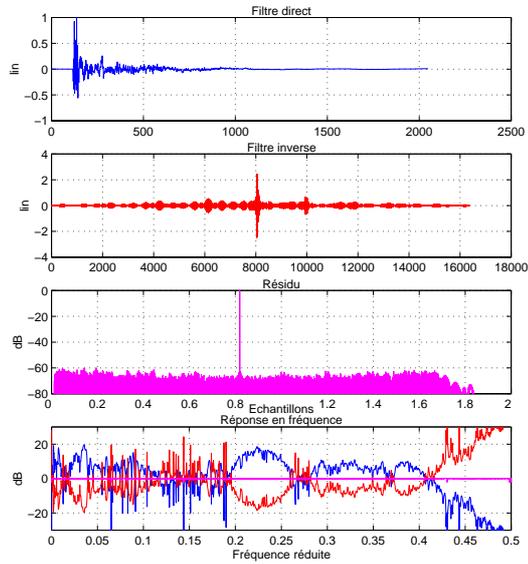
(a) Filtre inverse sur 2048 points



(b) Filtre inverse sur 4096 points

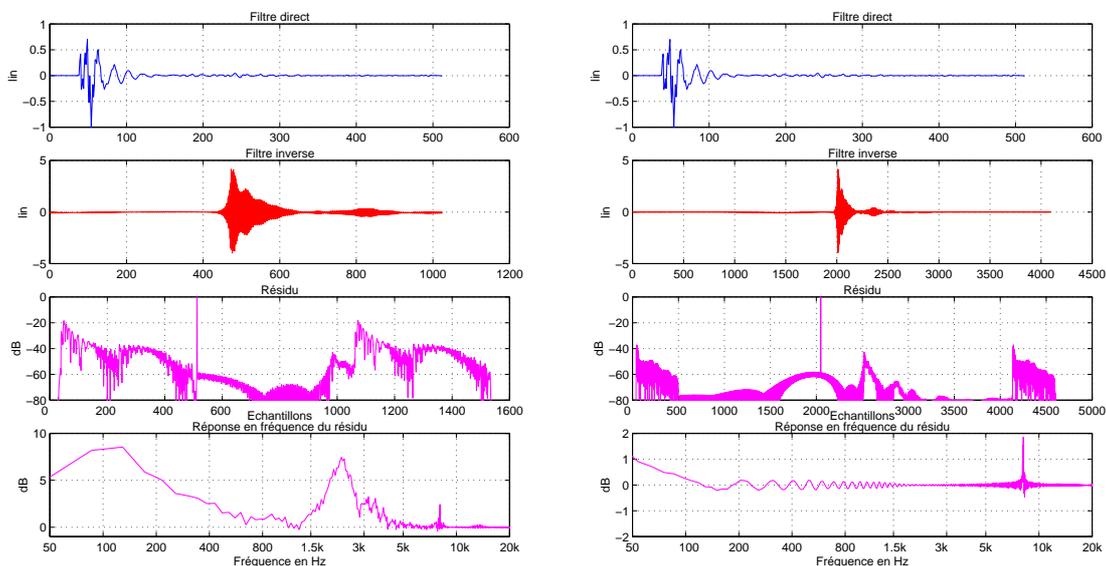


(c) Filtre inverse sur 8192 points



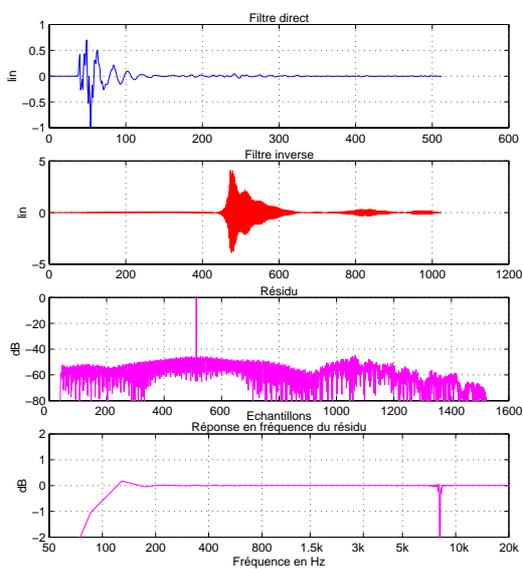
(d) Filtre inverse sur 16384 points

FIGURE D.4 – Inversion approximée au sens des moindres carrés d'une réponse impulsionnelle d'habitable mesurée sur 16384 échantillons dont on a conservé que les 2048 premiers



(a) Calcul effectué dans le domaine fréquentiel, filtre inverse sur 1024 points

(b) Calcul effectué dans le domaine fréquentiel, filtre inverse sur 4096 points



(c) Calcul effectué dans le domaine temporel, filtre inverse sur 1024 points

FIGURE D.5 – Inversion d'une réponse impulsionnelle d'oreille mesurée sur 512 points, comparée dans le domaine temporel et dans le domaine fréquentiel

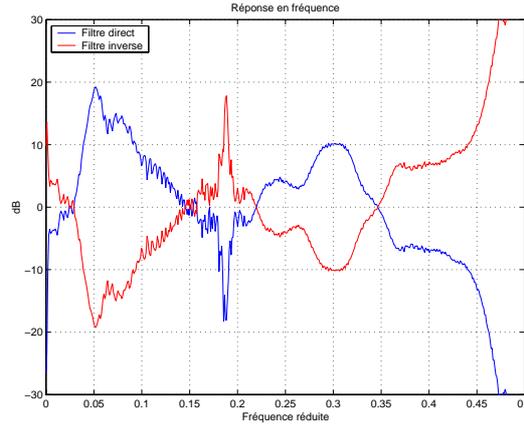


FIGURE D.6 – Réponses en fréquence du filtre direct et du filtre inverse, sans régularisation

D.1.2.2 Techniques de régularisation

Dans cette section, nous étudierons différentes méthodes permettant d’effectuer une régularisation, c.-à-d. une mise en forme du filtre direct effectuée avant la procédure d’inversion. Nous nous intéresserons à la technique bien classique consistant à limiter la dynamique du module ; puis nous présenterons une autre technique consistant à appliquer un filtre spécifique pendant la procédure d’inversion ; enfin nous parlerons de la troncature de la décomposition en valeurs singulières du filtre original.

Limitation de la dynamique du module La transformée de Fourier discrète (TFD) de la série $h(n)$ s’écrit

$$H(k) = \sum_{k=0}^{N-1} h(n) \exp(-2j\pi kn/N) = H(e^{2j\pi f}) \Big|_{f=k/N} ; \quad (D.78)$$

elle peut être vue comme étant un échantillonnage de la transformée de Fourier à temps discret de $h(n)$. L’inverse de ces valeurs constitue la TFD de $g(n)$, c.-à-d.

$$\frac{1}{H(k)} = G(e^{2j\pi f}) \Big|_{f=k/N} . \quad (D.79)$$

Comme $g(n)$ est à support infini, le spectre discret défini ci-dessus est sous-échantillonné, et si on calcule la TFD inverse, le résultat sera entaché par du recouvrement temporel, ce qui s’écrit

$$\tilde{g}(n) = \sum_{r=-\infty}^{+\infty} g(n + rN) = \text{TFDI} \left(\frac{1}{H(k)} \right) . \quad (D.80)$$

Mais comme $g(n)$ décroît exponentiellement lorsque $|n|$ augmente, lorsqu’on augmente la taille de la TFD, le recouvrement temporel diminue, et $\tilde{g}(n)$ approxime $g(n)$. Lorsque $H(z)$ a des zéros proches du cercle unité, la décroissance des pôles correspondant de $G(z)$ sera lente, ce qui nécessite une taille plus importante de la TFD pour combattre le recouvrement temporel.

Il est cependant possible de limiter la dynamique de $|1/H(z)|$ avant le calcul de la TFD inverse, ce qui permet de calculer $\tilde{g}(n)$ en utilisant une taille plus petite de la TFD. Les opérations à réaliser sont les suivantes :

1. Décider du nombre de points M du filtre inverse.
2. Calculer la TFD de $h(n)$ sur M points en utilisant l’algorithme de FFT.
3. Calculer $1/H(k)$ et limiter la dynamique du module.
4. Calculer la TFD inverse du spectre complexe à l’aide de la FFT inverse.

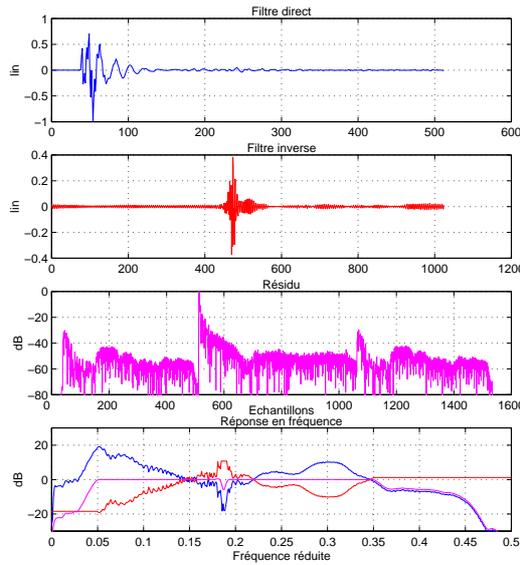


FIGURE D.7 – Régularisation par limitation de la dynamique et prolongement par une constante dans les fréquences extrêmes

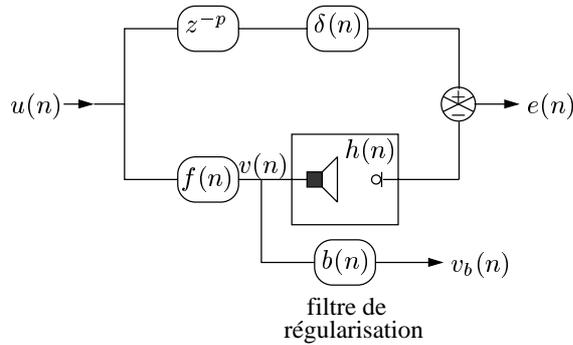


FIGURE D.8 – Problème modifié de déconvolution dans le domaine temporel discret

5. Fenêtrer le signal temporel résultant, et lui adjoindre un retard pour le rendre causal.

Le résultat d’un tel processus est présenté sur la figure D.7. Le filtre direct est toujours la même réponse impulsionnelle d’oreille mesurée sur 512 points ; le filtre inverse est calculé sur 1024 points. La dynamique a été limitée à 30 dB, et surtout aucune inversion n’a été effectuée en basses et hautes fréquences. L’erreur est très importante, mais cette fois le filtre inverse est audible.

Régularisation par filtrage Il s’agit de modifier la fonction d’erreur introduite à l’équation (D.69). Pour cela, intéressons-nous au schéma de la figure D.8. Le retard de modélisation p est nécessaire pour compenser l’inversion de la partie à phase non minimale de $h(n)$; on suppose toujours que le filtre direct $h(n)$ est de longueur N et que le filtre inverse $f(n)$ est de longueur $M \geq N$. L’idée de cette technique [163, 232] est de minimiser une fonction de coût donnée par

$$J = E + \beta V, \tag{D.81}$$

où E est une mesure de l’erreur de performance (adéquation du filtre inverse au signal désiré) et V est une mesure de l’effort déployé pour atteindre le but ; β , nombre réel positif, permet de déterminer le poids accordé au terme d’effort [70, 230]. En variant β de 0 à l’infini, le filtre inverse calculé varie selon que

l'on minimise seulement E ou seulement V . Le principe est le suivant : si le système est mal conditionné à une fréquence particulière (c.-à-d. si il présente une anti-résonance très marquée), le filtre inverse ne réussira pas à déconvoluer le système à cette fréquence parce que le terme d'effort sera très important comparé à l'erreur de performance.

Comme une grande valeur de β signifie que la solution favorisera une faible puissance de sortie du filtre inverse au dépend d'une erreur de performance importante, il est évident que β peut être utilisé pour réguler la puissance de sortie du filtre inverse. De plus, β peut être utilisé pour contrôler la « durée » du filtre inverse, et donc fournit un moyen de minimiser les effets de repliement habituellement rencontrés lors de la conception de filtres par des méthodes basées sur l'échantillonnage du domaine fréquentiel (et dont nous parlerons un peu plus loin). Plus précisément, la régularisation permet de contrôler la plus grande constante de temps du filtre inverse [161]³ ; pour faire en sorte que la valeur de cette constante de temps ne soit ni trop faible si trop élevée, le paramètre de régularisation β doit être réglé séparément. Si β est trop petit, il y aura des résonances très marquées dans la réponse en fréquence du filtre inverse, et si β est trop grand, l'inversion ne sera pas très précise. Heureusement, d'après Kirkeby [162] la valeur exacte de β n'est pas critique, puisqu'une tolérance de 20 % sur la valeur optimale peut être admise sans impact nocif sur la qualité perceptive du filtre inverse.

Plaçons-nous dans un espace de Hilbert. Ecrivons $E = \mathbf{e}^t \mathbf{e}$; traditionnellement, le terme d'effort V est défini comme l'énergie totale à l'entrée du haut-parleur, c.-à-d. $V = \mathbf{v}^t \mathbf{v}$ [163]. Cependant, il est possible d'introduire une régularisation dépendant de la fréquence [159], et de récrire $V = \mathbf{v}_b^t \mathbf{v}_b$, où \mathbf{v}_b est la version filtrée de \mathbf{v} . Dans ce cas, le vecteur \mathbf{f} qui minimise la fonction de coût J est donné par

$$\mathbf{f} = (\mathbf{H}^t \mathbf{H} + \beta \mathbf{B}^t \mathbf{B})^{-1} \mathbf{H}^t \delta_p, \quad (\text{D.82})$$

où δ_p représente le vecteur $\delta(n - p)$, $n = 0, 1, \dots, M + N - 2$.

Lorsque β et \mathbf{b} sont correctement choisies, la régularisation assure que le filtre inverse \mathbf{f} n'amplifiera pas de manière excessive certaines zones fréquentielles ; il est donc possible d'atténuer de manière sélective les résonances trop marquées du spectre du signal passé au travers de ce filtre.

En d'autres termes, le degré de conditionnement du système peut être évalué en calculant les pôles de la solution inverse « exacte » au sens des moindres carrés (plus ils sont proches du cercle unité, plus le mauvais conditionnement est fort), et il est démontré que la régularisation s'effectue en éloignant ces pôles du cercle unité [164].

Le filtre $b(n)$ doit répondre au principe suivant : il doit avoir un gain $|B(e^{2j\pi f})|$ élevé dans les régions fréquentielles où on veut limiter l'inversion du filtre direct $h(n)$; ailleurs, le gain doit être très faible [166]. Il est donc possible de contrôler l'inversion de manière précise un grand nombre de régions fréquentielles, moyennant un nombre de coefficients plus ou moins élevé pour le filtre RIF $b(n)$ (ce qui n'est pas un inconvénient tant qu'il ne dépasse pas M , nombre de points désiré pour le filtre inverse, puisque la matrice $\mathbf{H}^t \mathbf{H}$ est de taille $M \times M$). Notons qu'il peut être avantageux de spécifier le gabarit de $|B(e^{2j\pi f})|$ sur une échelle fréquentielle logarithmique, afin de se rapprocher des propriétés perceptives de l'oreille.

Une fois le filtre $b(n)$ calculé, il faut déterminer β . Ce coefficient agit de manière globale sur toute la réponse en fréquence du filtre à inverser ; une valeur élevée entraîne tout simplement une inversion moins efficace. On peut spécifier un critère pour β : par exemple, on peut faire en sorte qu'une certaine valeur maximale du gain du filtre inverse soit atteinte. D'après nos investigations, il est apparu que

$$\beta = \frac{1 - \min |H(e^{2j\pi f})|^2}{|B(e^{2j\pi f})|^2} \Rightarrow |F(e^{2j\pi f})| \geq \frac{1}{2\sqrt{\beta}|B(e^{2j\pi f})|}. \quad (\text{D.83})$$

3. nous déconseillons la lecture de ce rapport très complet mais obscur

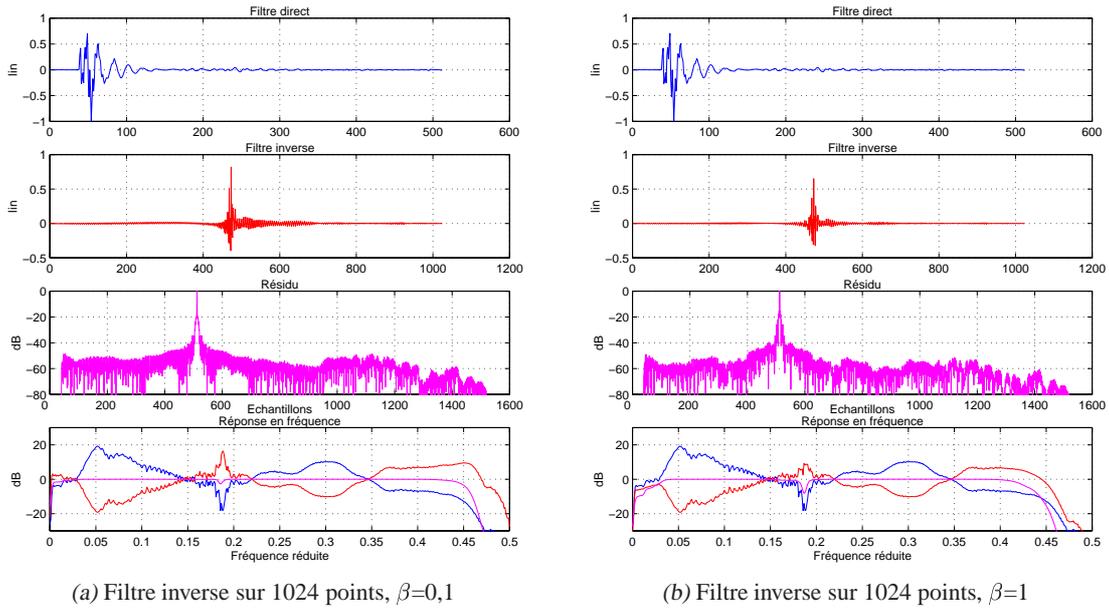


FIGURE D.9 – Inversion d'une réponse impulsionnelle d'oreille mesurée sur 512 points, régularisée par filtrage et calculée dans le domaine temporel

À titre d'information, mentionnons aussi que β peut être fixé par des critères de masquage fréquentiel sur le signal qui va *a priori* passer dans le filtre inverse [183].

Cette méthode peut être utilisée dans le domaine temporel grâce à un algorithme efficace de résolution de l'équation (D.82), dans laquelle la matrice $(\mathbf{H}^t \mathbf{H} + \beta \mathbf{B}^t \mathbf{B})$, est de forme Toeplitz symétrique, de taille $L = M + N - 1$. L'algorithme que nous avons choisi est celui de Levinson, car il nécessite une très faible occupation en mémoire. Pour un filtre direct de taille N , les opérations à réaliser sont les suivantes :

1. Décider du nombre de points M du filtre inverse.
2. Spécifier le gabarit de $|B(e^{2j\pi f})|$ et la valeur de β .
3. Calculer la première ligne des matrices d'autocorrélation $\mathbf{H}^t \mathbf{H}$ et $\mathbf{B}^t \mathbf{B}$, de longueur M .
4. Utiliser l'algorithme de Levinson pour résoudre le système linéaire $(\mathbf{H}^t \mathbf{H} + \beta \mathbf{B}^t \mathbf{B})\mathbf{f} = \mathbf{H}^t \delta_p$.

Sur la figure D.9, on peut voir l'influence du facteur de régularisation β lorsque le calcul est effectué conformément à la procédure décrite ci-dessus. Le filtre direct est toujours la même réponse impulsionnelle d'oreille mesurée sur 512 points ; le filtre inverse est calculé sur 1024 points.

Mais on peut aussi résoudre cette équation dans le domaine fréquentiel (cf. méthode de Hunt, décrite à la fin du § D.1.1.3), qui devient

$$F(e^{2j\pi f}) = \frac{H^*(e^{2j\pi f})}{|H(e^{2j\pi f})|^2 + \beta |B(e^{2j\pi f})|^2}, \text{ où} \quad (\text{D.84})$$

$H(e^{2j\pi f})$ et $B(e^{2j\pi f})$ représentent les transformées de Fourier à temps discret discret des séquences $h(n)$ et $b(n)$, et l'exposant $*$ le passage au complexe conjugué. L'avantage d'un coût de calcul ridicule étant compensé par la précision moindre de l'inversion, puisqu'il faut procéder à un échantillonnage du domaine fréquentiel (l'équation précédente n'est en fait résolue que pour un nombre fini de points) ; plus l'échantillonnage est fin meilleure est l'inversion, mais plus long est le filtre inverse !

Pour un filtre direct de taille N , les opérations à réaliser sont les suivantes :

1. Décider du nombre de points M du filtre inverse.

2. Spécifier le gabarit de $|B(e^{2j\pi f})|$ et la valeur de β .
3. Calculer les TFD de $h(n)$ et $b(n)$ sur M points en utilisant l'algorithme de FFT.
4. Pour chaque échantillon k de la TFD, calculer $F(k)$, $k = 0, 1, \dots, N - 1$ selon l'équation (D.84).
5. Calculer une période de $f(n)$ en prenant la TFD inverse de $F(e^{2j\pi f})$.
6. Rendre effectif le retard de modélisation, en effectuant un décalage circulaire de $f(n)$ de p échantillons.

Gardons bien présent à l'esprit que $f(n)$ est entaché d'une erreur correspondant au phénomène de repliement temporel, phénomène absolument *inévitabile*, et associé au calcul d'un filtre dans le domaine fréquentiel discret (l'équation (D.84) réalise implicitement une convolution circulaire entre le numérateur et le dénominateur). En pratique, c'est le facteur de régularisation β qui fait en sorte que le filtre inverse ait une décroissance suffisamment rapide pour minimiser les effets de la convolution circulaire. Plus la valeur de β est importante, plus le repliement sera minimisé, mais moins précise sera l'inversion !

Sur la figure D.10, on peut voir l'influence du facteur de régularisation β lorsque le calcul est effectué dans le domaine fréquentiel ; en particulier, l'influence est double : elle a lieu sur la réponse en fréquence du filtre inverse, en empêchant une amplification trop prononcée, et elle a lieu également sur sa réponse temporelle, en minimisant l'influence de la convolution circulaire.

Régularisation par troncature de la décomposition en valeurs singulières Cette méthode a été décrite au § D.1.1.3. Il s'agit de calculer l'inverse généralisée au sens de Moore & Penrose de la matrice de convolution \mathbf{H} , c.-à-d.

$$\mathbf{H}^+ = \mathbf{V}\mathbf{\Lambda}^+\mathbf{U}^*, \quad (\text{D.85})$$

avec

$$\mathbf{\Lambda}^+ = \text{diag}(\alpha_i), \text{ avec } \begin{cases} \alpha_i = \frac{1}{\lambda_i} & \text{si } \lambda_i \neq 0 \text{ significatif} \\ \alpha_i = 0 & \text{si } \lambda_i \approx 0 \text{ non significatif} \end{cases}, \quad (\text{D.86})$$

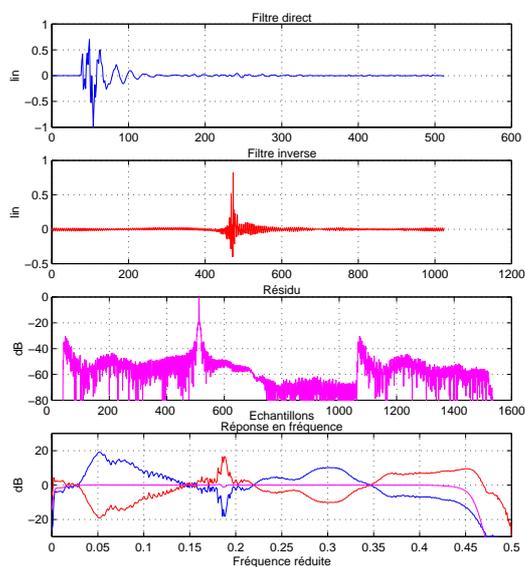
et \mathbf{U} la matrice unitaire dont les colonnes sont les vecteurs propres de $\mathbf{H}^*\mathbf{H}$, \mathbf{V} la matrice unitaire dont les colonnes sont les vecteurs propres de $\mathbf{H}\mathbf{H}^*$, et $\mathbf{\Lambda}$ la matrice diagonale formée des racines carrées des valeurs propres correspondantes, appelées valeurs singulières de \mathbf{H} .

En pratique, cette méthode ne peut être utilisée dans le cadre de l'inversion de réponses impulsionnelles comportant plusieurs milliers de points *que* si on effectue l'*approximation* de la matrice de convolution, de forme Toeplitz, sous forme circulante ; le calcul peut alors être effectué dans le domaine fréquentiel discret.

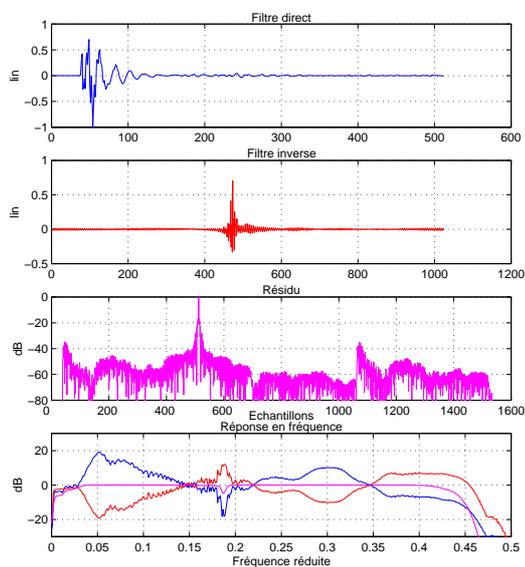
Il est seulement nécessaire de calculer la première ligne de \mathbf{H}^+ , puisque cette matrice est également circulante. De plus, la matrice $\mathbf{\Lambda}$ est obtenue par TFD et les matrices \mathbf{U} et \mathbf{V} contiennent les bases d'exponentielles complexes de la TFD, facilement calculables. Par conséquent, il n'est pas nécessaire de stocker la moindre matrice, et l'occupation mémoire est très faible. Le nombre d'opérations à réaliser étant important, cet algorithme a été implanté sous forme d'un objet externe pour Matlab[®] programmé en C.

Il faut bien sûr appliquer un décalage circulaire des éléments de la première ligne de \mathbf{H}^+ pour rendre effectif le retard de modélisation.

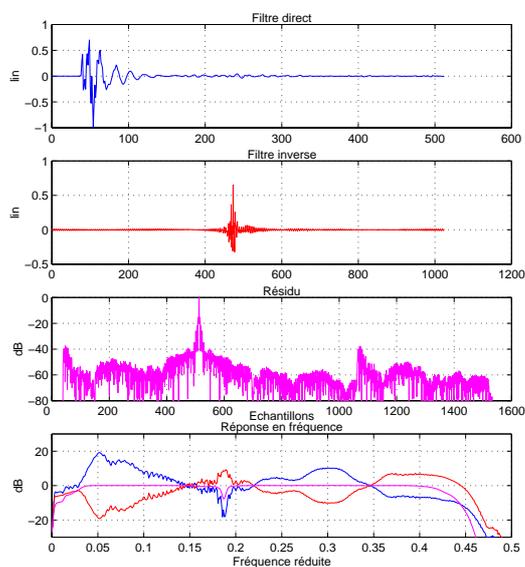
Gardons à l'esprit que la décomposition en valeurs singulières ne peut être calculée que moyennant l'approximation de la matrice de convolution de Toeplitz en matrice circulante. De plus, le filtre inverse est également entaché d'une erreur de repliement due à sa spécification dans le domaine fréquentiel discret. Enfin, il n'est pas facile de définir le seuil en dessous duquel les valeurs singulières seront considérées comme nulles, car elles varient continûment (cf. fig. D.11). Remarquons que le rapport $\lambda_{\max}/\lambda_{\min}$ est ici très élevé, et témoigne donc d'un problème largement mal posé ; ce rapport mesure le conditionnement de la matrice, ou en d'autres termes l'immunité de la solution vis-à-vis de faibles perturbations du filtre direct (cf. § D.1.1.1). Cette approche a été utilisée par Ordua [244] pour tester la robustesse d'un système d'annulation des trajets croisés.



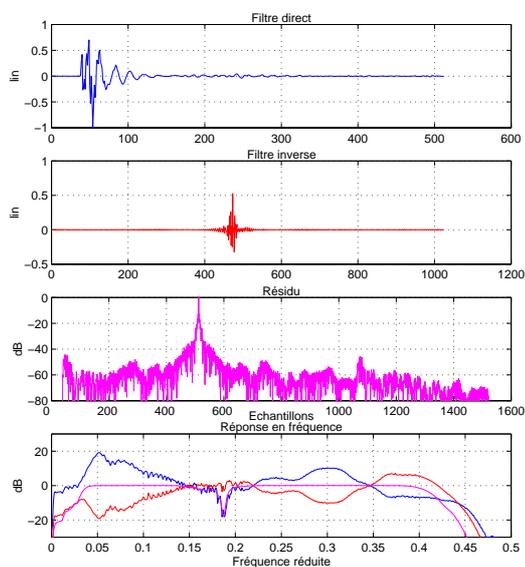
(a) Filtre inverse sur 1024 points, $\beta=0,1$



(b) Filtre inverse sur 1024 points, $\beta=0,5$



(c) Filtre inverse sur 1024 points, $\beta=1$



(d) Filtre inverse sur 1024 points, $\beta=5$

FIGURE D.10 – Inversion d'une réponse impulsionnelle d'oreille mesurée sur 512 points, régularisée par filtrage et calculée dans le domaine fréquentiel

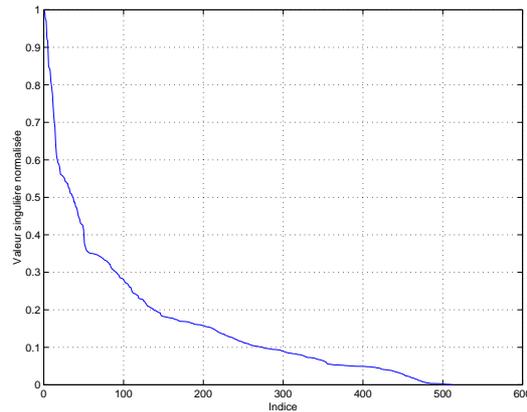


FIGURE D.11 – Valeurs singulières de la réponse impulsionnelle directe, classées par ordre décroissant

Ici encore, bien qu'une vue objective de la situation permette d'optimiser ce seuil, c'est bien souvent l'écoute qui permettra de trouver les meilleurs paramètres.

Sur la figure D.12, on peut voir l'influence de la troncature plus ou moins importante de la décomposition en valeurs singulières ; pratiquement, on fixe le seuil de manière relative par rapport à la plus grande valeur singulière. Il apparaît cependant deux erreurs : celle due au repliement temporel du filtre inverse (qui disparaît avec l'importance de la troncature), et celle due à l'approximation circulante de la matrice de convolution (cette dernière est visible sur la figure D.12(a) où on a gardé toutes les valeurs singulières).

Régularisation par amortissement des valeurs singulières Cette méthode a été décrite au § D.1.1.3. Comme précédemment, on calcule effectue la décomposition en valeurs singulières de la matrice \mathbf{H} , mais au lieu de définir un seuil en dessous duquel les valeurs singulières seront considérées comme nulles, on effectue un amortissement de ces valeurs singulières. Pratiquement, on ajoute une constante aux valeurs singulières, procédé qui aura une influence d'autant plus significative que les valeurs singulières sont faibles. Contrairement à la troncature de la décomposition, il n'y a pas réduction de la dimension du sous-espace de projection.

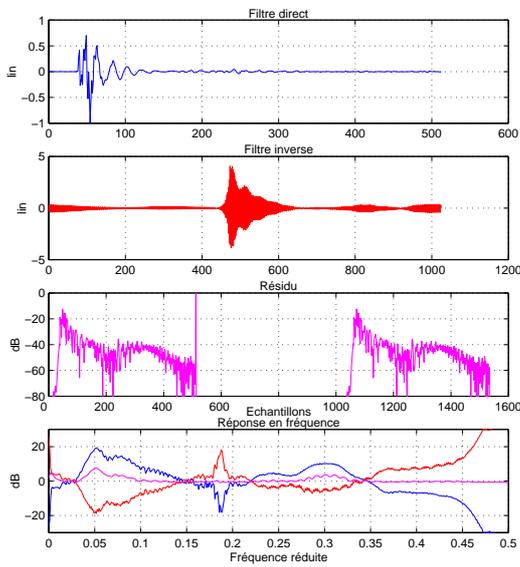
Cet algorithme a également été implanté sous forme d'un objet externe pour Matlab[®] programmé en C.

Sur la figure D.13, on peut voir l'influence de la constante ajoutée aux valeurs singulières, et qui a pour effet de limiter l'inversion des faibles valeurs. Il apparaît que, contrairement à la troncature, l'amortissement des valeurs singulières permet de réduire de manière moins significative l'influence du repliement temporel du filtre inverse.

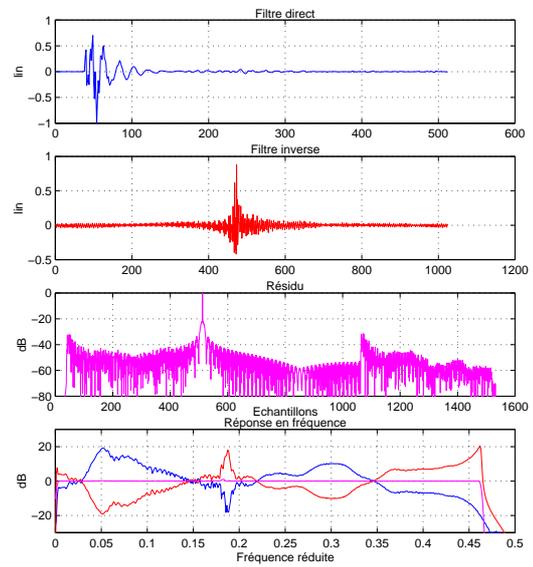
D.1.2.3 Conclusion

Toutes les méthodes illustrées précédemment ont un point commun : elles permettent de calculer une inversion approchée d'une réponse impulsionnelle comportant une composante à phase maximale ; et par le biais de techniques de régularisation, elles fournissent des filtres inverses perceptivement convaincants, c.-à-d. comportant des résonances limitées et des régions basses et hautes fréquences adoucies.

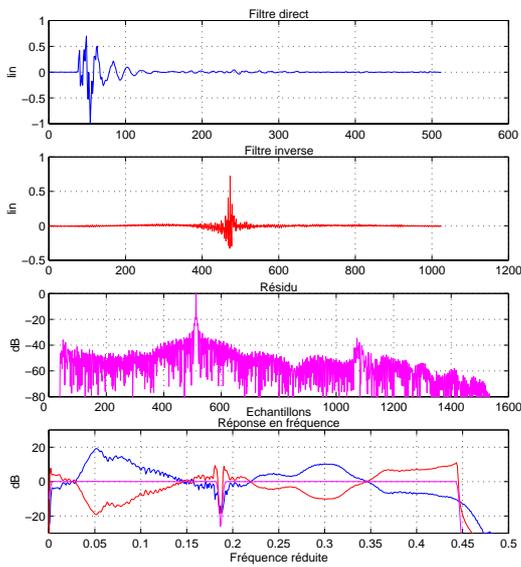
L'approximation au sens des moindres carrés est plus robuste que l'utilisation de la technique cepstrale, et sa mise en œuvre dans le domaine temporel permet de s'affranchir des problèmes d'échantillonnage du domaine fréquentiel discret (nombre d'échantillons élevé pour garantir une précision suffisante,



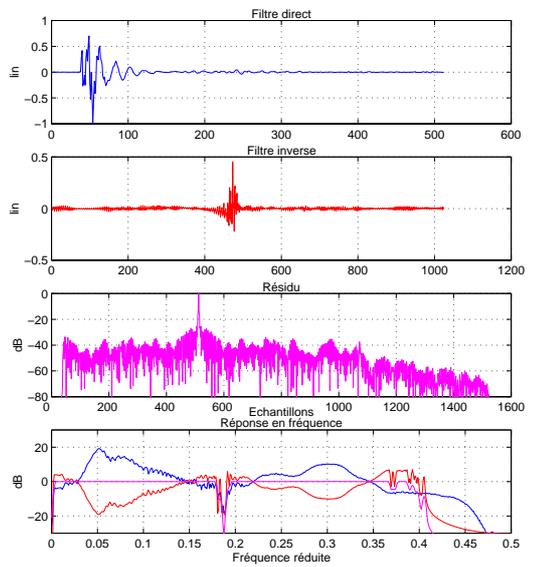
(a) Filtre inverse sur 1024 points, seuil=0%



(b) Filtre inverse sur 1024 points, seuil=1%

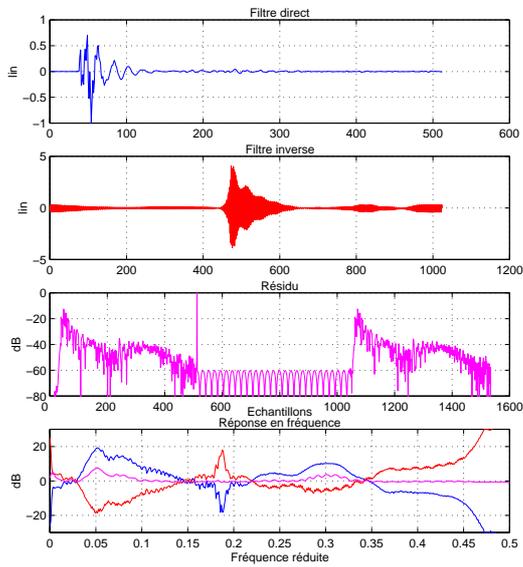


(c) Filtre inverse sur 1024 points, seuil=3%

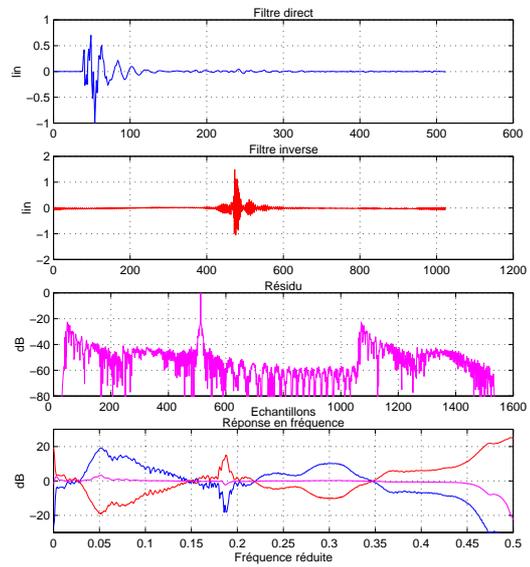


(d) Filtre inverse sur 1024 points, seuil=5%

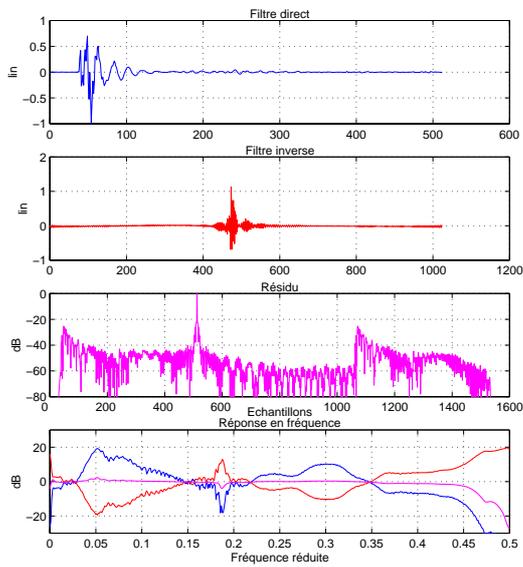
FIGURE D.12 – Inversion d'une réponse impulsionnelle d'oreille mesurée sur 512 points, régularisée par troncature de sa décomposition en valeurs singulières, réalisée dans le domaine fréquentiel



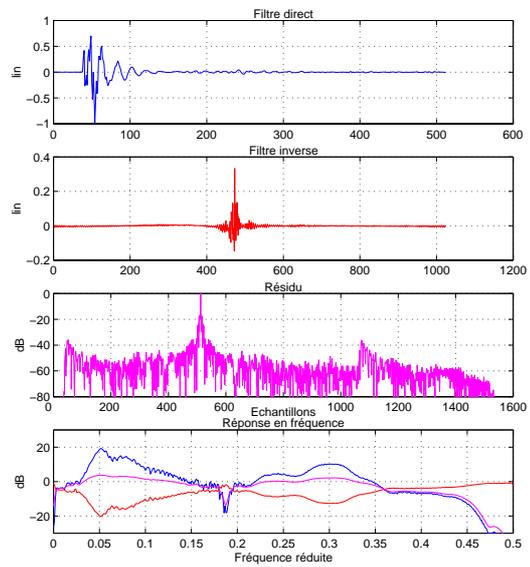
(a) Filtre inverse sur 1024 points, constante=0



(b) Filtre inverse sur 1024 points, constante=0,05



(c) Filtre inverse sur 1024 points, constante=0,10



(d) Filtre inverse sur 1024 points, constante=1,00

FIGURE D.13 – Inversion d'une réponse impulsionnelle d'oreille mesurée sur 512 points, régularisée par amortissement de ses valeurs singulières, réalisée dans le domaine fréquentiel

synonyme d'une longueur de filtre inverse importante, et phénomène inévitable de repliement lors du retour dans le domaine temporel) ; mais le coût de calcul d'une implantation dans le domaine temporel est très élevé, puisqu'il faut inverser un système linéaire... heureusement réalisable rapidement, avec une occupation mémoire, limitée grâce à l'algorithme de Levinson.

Concernant la régularisation, toutes les méthodes ne sont pas équivalentes.

La régularisation par limitation de la dynamique du module laisse néanmoins l'excès de phase intact, et lors du retour dans le domaine temporel, la correspondance entre module et phase est partiellement perdue. Même si le filtre est perceptivement valide, il comporte une erreur sur la phase.

La régularisation par filtrage, calculée dans le domaine temporel, semble à nos yeux la meilleure méthode. D'une part, le filtre de régularisation $b(n)$ peut être conçu pour empêcher l'apparition de nombreuses résonances, et d'autre part, le facteur de régularisation β permet de contrôler globalement l'importance de la régularisation et donc la puissance du filtre inverse. La mise en œuvre dans le domaine temporel permet de s'affranchir du problème de repliement temporel inhérent au passage dans le domaine fréquentiel discret, au prix d'un coût de calcul beaucoup plus élevé, mais non rédhibitoire grâce à l'algorithme de Levinson.

La décomposition d'une réponse impulsionnelle acoustique en valeurs singulières est une méthode nouvelle et originale, dont nous n'avons pas trouvé trace dans la littérature. La limitation du nombre de ses valeurs singulières avant le calcul de l'inverse est une technique pertinente ; mais elle se heurte à deux difficultés : d'une part il n'est pas évident de définir un seuil en dessous duquel les valeurs singulières seront forcées à zéro, et d'autre part le filtre inverse comporte un repliement temporel (néanmoins contrôlable par la valeur du seuil). L'amortissement des valeurs singulières permet de limiter l'inversion des plus faibles valeurs, sans limiter la dimension du sous-espace de projection, et de ce fait ne fournit pas véritablement une solution stable ; le repliement temporel inhérent à la méthode est, de plus, difficilement contrôlable par le taux d'amortissement.

ANNEXE E

Réverbération artificielle

E.1 Le Spatialisateur®

Le choix de la simulation d'un effet de salle par réseau de retards récursifs s'est imposé à l'IRCAM, s'appuyant en cela sur les nombreuses études qui y ont été menées sur la perception de la qualité acoustique [22, 85, 145, 150, 181, 314]. Un système temps-réel de réverbération, appelé Spatialisateur® (© IRCAM), fonctionne sous forme d'une bibliothèque d'objets compatibles avec le logiciel *jMax*®¹ (© IRCAM) sous Linux, ou avec le logiciel Max/MSP®² (© OPCODE/CYCLING74/IRCAM) sous MacOS. Étant conçu suivant une structure modulaire, il peut être considéré comme une application complète, ou comme une librairie d'objets élémentaires pour le traitement spatial des sons en temps réel. L'algorithme développé apporte les avantages suivants [140, 141, 143, 144] :

- le contrôle en temps réel de la distribution des réflexions précoces, et de la réverbération tardive ;
- la possibilité d'augmenter indéfiniment le temps de réverbération sans augmenter le coût de calcul ;
- la possibilité d'obtenir plusieurs canaux de sortie sans augmenter le coût de calcul.

L'absence de coloration dans les filtres en peigne récursifs est apporté par des matrices de bouclages unitaires, ainsi que par une densité d'échos et de modes suffisants [138, 141].

Le Spatialisateur® reçoit les signaux provenant de sources sonores instrumentales ou synthétiques (supposées dénuées de réverbération), et fournit les signaux prêts à alimenter chaque haut-parleur du dispositif de reproduction. L'approche générale adoptée dans ce projet peut se caractériser par le fait qu'elle permet à l'utilisateur de spécifier l'effet désiré du point de vue de l'auditeur plutôt que du point de vue du traitement réalisé ou du dispositif utilisé. Cela se traduit par trois caractéristiques essentielles :

1. L'intégration des traitements directionnels et temporels dans le même processeur.
2. L'adaptation du traitement au contexte de reproduction et de prise de son.
3. L'existence d'une interface de commande fondée sur une approche perceptive.

Le but visé par cet environnement de spatialisation n'est donc pas l'exactitude physique, mais la *pertinence perceptive*. L'avantage que cette approche remporte sur la convolution concerne le contrôle des effets perceptifs, qui sont représentés par neuf paramètres indépendants (cf. fig. E.1). L'inconvénient de cette approche est qu'elle est totalement dépendante de la puissance du processeur hôte, en terme de nombre de sources et de qualité du rendu de la réverbération.

Bien qu'il puisse être réglé par des considérations perceptives relativement intuitives, le Spatialisateur® peut aussi être réglé par une procédure objective fondée sur une analyse temps-fréquence de réponse impulsionnelle de salle (temps de réverbération et spectres énergétiques de quatre zones temporelles de la

1. <http://www.ircam.fr/produits/logiciels/log-forum/jmax-e.html>

2. <http://www.cycling74.com/products/maxmsp.html>

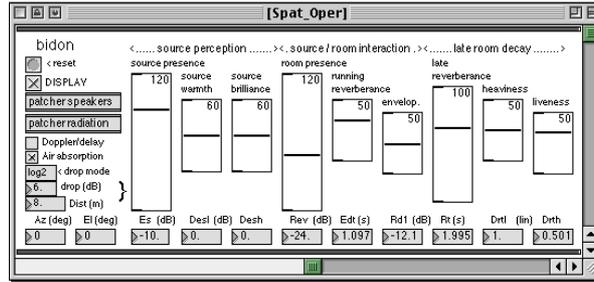


FIGURE E.1 – Interface perceptuelle de haut niveau permettant le réglage du Spatialisateur®

réponse). Il est donc également possible de pénétrer plus au cœur de l’algorithme, pour régler des critères objectifs de bas niveau plus proches du traitement du signal effectué.

E.1.1 Description détaillée de l’algorithme

La structure du processeur de réverbération numérique temps-réel est présentée figure E.2. Il intègre dans une même entité des effets :

- *temporels*, à savoir le contrôle de la distribution temporelle des réflexions précoces, des délimitations temporelles des différentes zones de la réponse impulsionnelle, et du temps de réverbération ;
- *directionnels*, à savoir le contrôle de l’effet binaural sur le son direct, de la latéralisation des premières réflexions, et de la corrélation interaurale des réflexions tardives diffuses ;
- *spectraux*, à savoir le contrôle du rayonnement de la source, et de l’atténuation spectrale à apporter à chaque zone temporelle de la réponse impulsionnelle.

EFFETS TEMPORELS

Le *découpage temporel* de la réponse est fixé arbitrairement sachant qu’aucune réflexion n’apparaîtra avant le début de la section R1 et que la densité d’échos doit être suffisante dès le début de la section R3 afin de garantir l’hypothèse de champ diffus (fig. E.3).

Les *retards* appliqués aux réflexions précoces sont calculés par la formule (E.1) où t_i correspond effectivement à la date d’arrivée de la réflexion mesurée sur la réponse impulsionnelle captée avec un micro omnidirectionnel, et ITD_i correspond au retard interaural de la réflexion en fonction de sa direction d’arrivée.

$$t_{li} = t_i - \frac{1}{2}ITD_i \quad \text{et} \quad t_{ri} = t_i + \frac{1}{2}ITD_i \quad (\text{E.1})$$

Le plus petit retard t est supérieur ou égal à 20 ms, et le plus grand inférieur ou égal à 40 ms, en vertu du découpage temporel montré figure E.3. Les retards τ' sont calculés de telle sorte que la somme des deux plus petits retards t et τ' soit égale à 40 ms et la somme des deux plus grands soit égale à 100 ms. Enfin, le plus petit des retards absorbants τ défini ci-après doit prendre une valeur telle que la somme de tous les plus petits retards t , τ et τ' soit égale à 100 ms.

Le contrôle du *temps de réverbération* en fonction de la fréquence est assuré par les filtres absorbants de la figure E.2. Ces filtres, dont une représentation schématique est donnée figure E.4, sont des filtres en peigne modifiés de telle sorte que le gain de boucle est remplacé par un filtre, noté $h_i(z)$ sur la figure E.4. Le filtre en peigne modifié est stable si la condition $|h_i(e^{2j\pi f})| < 1$ est remplie pour toute fréquence f . Ce principe de réglage repose sur l’hypothèse que le filtre réverbérant va se comporter localement, au voisinage d’une fréquence f donnée, comme si chaque filtre absorbant était remplacé par un gain pur k_i indépendant de la fréquence et tel que $k_i = |h_i(e^{2j\pi f})|$. Cela conduit, d’après [138], à spécifier la

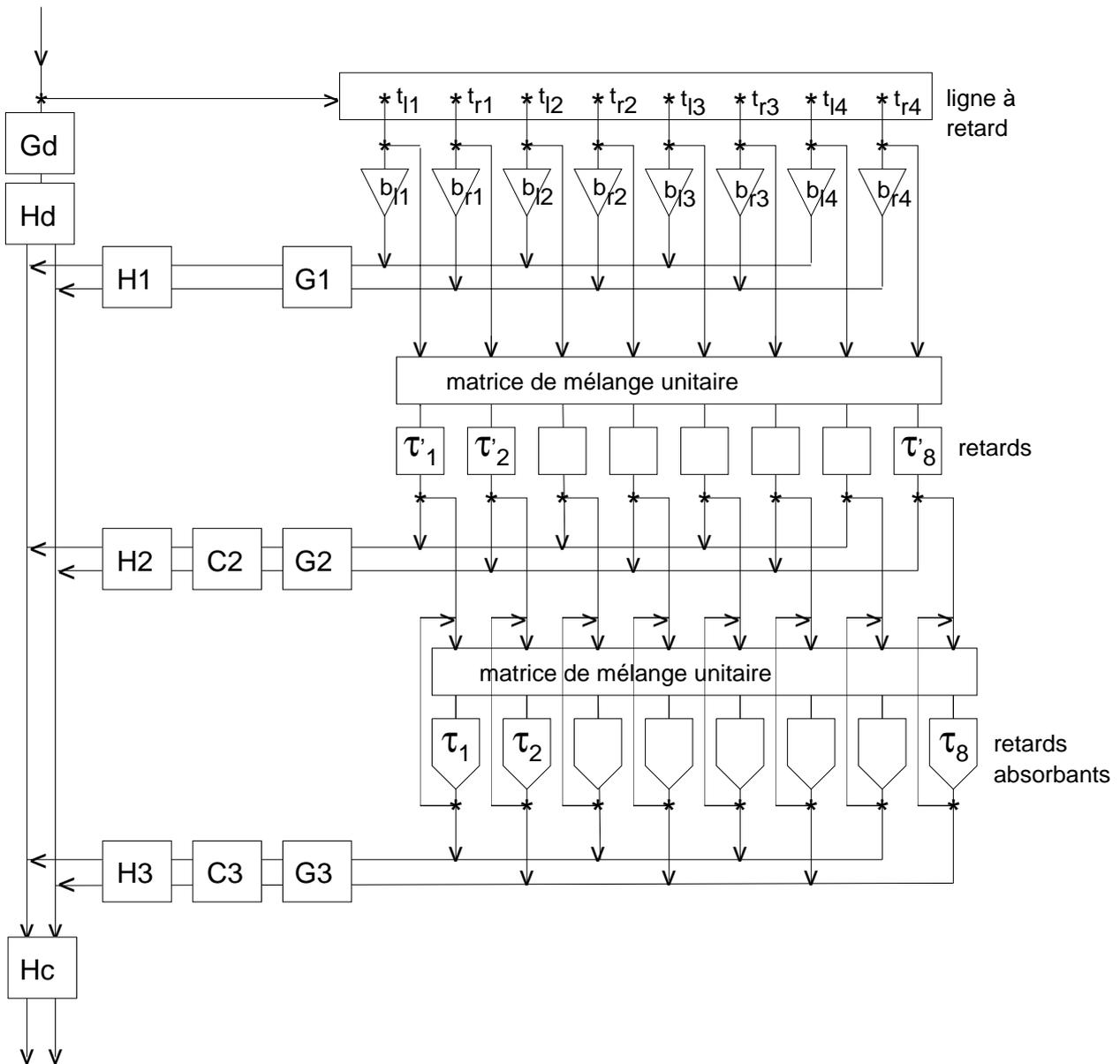


FIGURE E.2 – Structure de l’algorithme du Spatialisateur® pour une reproduction sur deux canaux. *G* : corrections spectrales mono, *C* : corrélations interaurales, *H* : corrections spectrales binaurales

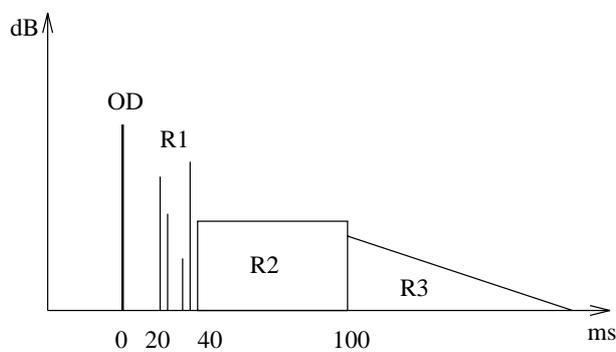


FIGURE E.3 – Décomposition temporelle par défaut de l’effet de salle par le Spatialisateur®

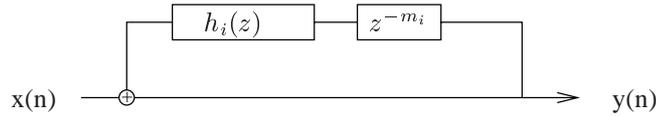


FIGURE E.4 – *Filtre en peigne modifié, ou «retard absorbant»*

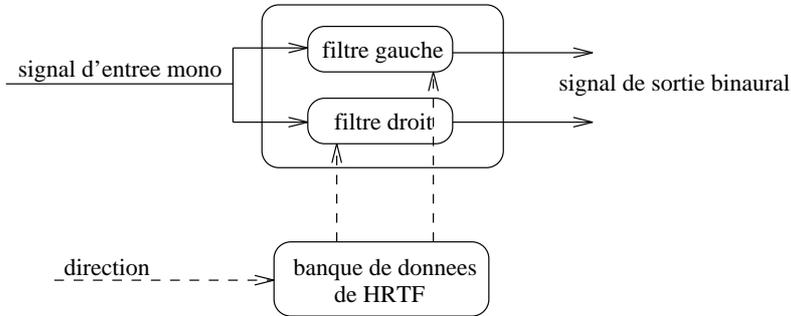


FIGURE E.5 – *Principe de la synthèse binaurale*

réponse en amplitude de chaque filtre absorbant en fonction de la courbe $T_r(f)$ par la relation

$$20 \log_{10} |h_i(e^{2j\pi f})| = -\frac{60 T_{ech}}{T_r(f)} m_i, \tag{E.2}$$

où T_{ech} est la période d'échantillonnage et m_i le retard en échantillons. Dans cette expression, on a négligé la réponse en phase du filtre absorbant.

EFFETS DIRECTIONNELS

Le contrôle de l'effet directionnel sur le son direct est assuré par le filtre H_d qui n'est autre que la fonction de transfert binaurale (Head Related Transfer Function, ou HRTF) correspondant à la direction que l'on veut donner au son direct. Le principe de fonctionnement de ce filtre est décrit figure E.5.

On notera la présence de gains énergétiques b_l et b_r sur les réflexions précoces. Ceux-ci sont calculés à l'aide des équations

$$b_{li} = b_i \times g_{li} \quad \text{et} \quad b_{ri} = b_i \times g_{ri}, \tag{E.3}$$

où b_i représente le gain énergétique de la réflexion sur la réponse impulsionnelle mesurée avec un micro omnidirectionnel, et g_{li} (resp. g_{ri}) représente la gain de la fonction de transfert binaurale gauche (resp. droite) pour la direction de la réflexion i .

La latéralisation des réflexions précoces est uniquement effectuée à l'aide des retards t et des gains b . Mais on peut aussi rajouter le filtre binaural H_1 . Deux choix sont possibles pour calculer ce filtre : le premier consiste à faire une moyenne des filtres binauraux correspondants à chacune des directions d'arrivée des échos (filtre moyen), et le deuxième est un filtre diffus calculé à partir des fonctions de transfert binaurales de toutes les directions d'arrivées possibles [141].

Enfin, la latéralisation des réflexions tardives diffuses est assurée par le coefficient de corrélation interaurale, ou IACC (Inter Aural Cross Correlation). Ce coefficient mesure l'intercorrélation des signaux parvenant aux deux oreilles (cf. § III.4.1). Il est modélisé par des filtres récursifs d'ordre huit.

Le contrôle de ce coefficient s'effectue simplement en mélangeant les signaux gauche et droit de sortie du filtre réverbérant (ceux-ci sont non corrélés). Martin [191] donne les relations suivantes pour corréler les signaux : si $y_1(n)$ et $y_2(n)$ sont des bruits blancs non corrélés, les signaux droit et gauche

$y_d(n)$ et $y_g(n)$ sont donnés par la relation

$$\begin{pmatrix} y_r(n) \\ y_l(n) \end{pmatrix} = \begin{pmatrix} \cos \theta & \sin \theta \\ \sin \theta & \cos \theta \end{pmatrix} \begin{pmatrix} y_1(n) \\ y_2(n) \end{pmatrix}. \quad (\text{E.4})$$

Le calcul de la fonction d'intercorrélation entre les signaux $y_d(n)$ et $y_g(n)$ montre alors que

$$\theta = \frac{1}{2} \arcsin(\text{IACC}). \quad (\text{E.5})$$

EFFETS SPECTRAUX

Les effets spectraux sur les quatre zones temporelles, tels que mesurés par le micro omnidirectionnel, sont représentés par les filtres G_d , G_1 , G_2 et G_3 . Le filtre G_d provient directement du calcul de l'énergie de la section OD. Le filtre G_1 provient du calcul de l'énergie de R1, corrigée par les gains b_i grâce à la relation $G_1 = \mathcal{E}(R1) / \sum_{i=1}^4 b_i$. Les gains G_2 et G_3 proviennent du calcul des énergies des sections R2 et R3. Le calcul des énergies est effectué dans le domaine temporel.

Les filtres H_2 et H_3 correspondent à la fonction de transfert en champ diffus du récepteur simulé (ici l'oreille), puisque l'on suppose que le champ sonore est diffus à partir de la section temporelle R2 [138].

Enfin, on notera la présence d'un filtre H_c d'égalisation en sortie, qui est utilisé soit pour l'annulation des trajets croisés dans le cas d'une simulation transaurale [138], soit pour l'égalisation du casque dans le cas d'une simulation binaurale. La présence de ce filtre est facultative dans le cadre d'une expérience de comparaison avec le résultat d'une convolution, car on peut tolérer une certaine « coloration » sur la réverbération, pourvu qu'elle soit identique dans les deux cas.

E.1.2 Normalisation

Lors de la mesure des fonctions de transfert d'oreille (HRTF), on inclus l'effet du haut-parleur et des micros. Et lors d'une synthèse binaurale (restitution au casque d'une direction d'incidence à l'aide des HRTF), on inclus également l'effet du casque. Afin d'éliminer ces effets, on peut déconvoluer les HRTF par la réponse du haut-parleur et celle de l'ensemble casque-micro. Cette déconvolution est effectuée dans le domaine fréquentiel, en supposant que toutes les fonctions de transfert sont à phase minimale [141]. Malheureusement, cette normalisation est loin d'être idéale, car elle lie irrémédiablement la banque de HRTF mesurée à une série particulière de transducteurs électroacoustiques. Pire même, elle suppose que le couplage casque-oreille soit identique à tout moment, et que la position du micro à l'intérieur du conduit auditif soit reproductible, ce qui n'est pas le cas [305]. Pour remédier à cette situation, il existe deux solutions :

- une normalisation par une HRTF particulière (dite normalisation *en champ libre*), en l'occurrence celle correspondant à la direction frontale dans le plan horizontal, puisque la plupart des casques sont égalisés en référence à cette direction. Cette normalisation élimine tous les effets indépendants de la direction, ainsi que la résonance du conduit auditif (donc l'influence de la position du micro dans le conduit) ;
- une normalisation par la HRTF en champ diffus (dite normalisation *en champ diffus*), qui n'est autre que la racine carrée de la puissance moyenne des HRTF [174, 197]; cette normalisation procure les mêmes avantages que la normalisation en champ libre, sans compter le fait qu'elle ne privilégie plus aucune direction particulière.

Le choix s'est porté sur la normalisation en champ diffus [141] qui offre en plus une compatibilité avec des mesures de HRTF réalisées sur tête artificielle normalisée par rapport au champ diffus, ainsi qu'une compatibilité avec une reproduction sur casques normalisés par rapport au champ diffus [305] pour lesquels une égalisation du couplage casque-oreille n'est pas indispensable.

De plus, cette normalisation permet de réduire le coût de calcul du processeur. En effet, si on décide de normaliser tous les filtres de corrections binaurales par rapport à la fonction de transfert binaurale en champ diffus, les filtres H2 et H3 disparaissent. Dans ce cas, le filtre d'égalisation du casque devra lui aussi être normalisé par la HRTF en champ diffus, à moins que le casque ne soit déjà égalisé en référence au champ diffus [305].

E.1.3 Simplifications

L'algorithme de simulation implanté dans le Spatialisateur[®] repose sur un certain nombre de simplifications.

Les *réflexions précoces* sont considérées comme discrètes, et sont présentes en petit nombre: huit au maximum (quatre pour une reproduction mono). Leur latéralisation est reproduite par simple effet stéréophonique : on applique seulement entre les canaux de sortie une différence de temps d'arrivée et de gain, correspondant à la direction de provenance de la réflexion. Ces différences de temps et de gain sont indépendantes de la fréquence. Alors qu'en réalité les réflexions précoces sont aussi filtrées par les fonctions de transfert de la tête.

La *réverbération tardive* est réalisée par un réseau de retards récursifs. On peut se demander si cette méthode produit une réverbération naturelle, et surtout si la densité de modes est suffisante.

La *modélisation temps-fréquence* de la réponse impulsionnelle fait apparaître quatre zones temporelles. Elles sont appelées OD, R1, R2 et R3, et correspondent à l'onde directe, les premières réflexions, le début de la réverbération tardive, et la réverbération tardive par elle-même. L'énergie de chaque section temporelle est réglable en trois bandes de fréquences, grâce à des filtres « double shelving » implantés sous forme d'égaliseurs paramétriques [66, 228] ; ces filtres peuvent être vus comme la mise en série d'un filtre shelving passe-bas et d'un filtre shelving passe-haut (ou vice versa) implantés sous forme de filtres RII d'ordre deux [266]. On peut ainsi contrôler l'énergie en trois bandes de fréquences dont les limites sont réglables (cf. fig. E.6). Chaque filtre est donc défini par cinq paramètres :

- deux fréquences de transition : f_1 et f_2 ;
- trois niveaux énergétiques : en basses fréquences, en moyennes fréquences et en hautes fréquences.

Le niveau en basses fréquences correspond à celui du continu, le niveau en moyennes fréquences correspond à celui de la fréquence moyenne (sur une échelle logarithmique) des deux fréquences de transition, et le niveau en hautes fréquences correspond à celui de la demi-fréquence d'échantillonnage. On remarquera que le niveau énergétique à chaque fréquence de transition vaut la moitié de l'excursion énergétique entre la bande centrale et chaque bande extrême. Par défaut dans le Spatialisateur[®] les fréquences de transition sont fixées à $f_1=250$ Hz et $f_2=8000$ Hz ; le niveau en moyennes fréquences est donc fixé pour une fréquence de 1414 Hz.

Le *temps de réverbération* est lui aussi réglable uniquement dans trois bandes de fréquences. Ce réglage est toujours effectué grâce à un filtre implanté sous forme d'égaliseur paramétrique.

L'*effet directionnel sur le son direct* est réalisé par filtrage à l'aide des fonctions de transfert de la tête (HRTF). Ces fonctions de transfert sont modélisées par des filtres RII d'ordre douze [174, 189, 260]. Mais il faut s'assurer que l'on garde suffisamment d'information.

Enfin le *coefficient de corrélation interaurale* est modélisé à l'aide d'un filtre RII d'ordre huit, dans sept bandes d'octaves uniquement. Il en est de même pour les filtres en $\sin \theta(f)$ et $\cos \theta(f)$.

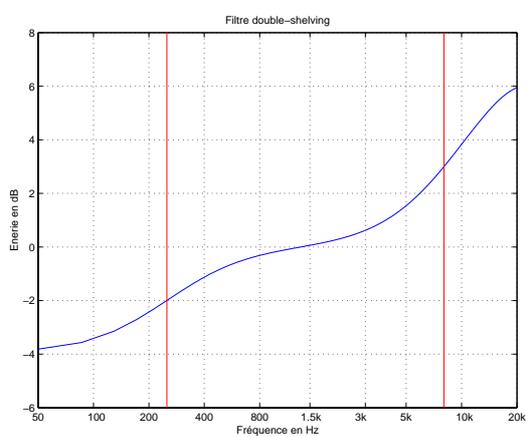


FIGURE E.6 – Exemple de fonction de transfert d'un filtre « double shelving » utilisé dans le Spatialisateur®, tracée sur une échelle fréquentielle logarithmique. On a spécifié une énergie de -4 dB en basses fréquences, 0 dB en moyennes fréquences et +6 dB en hautes fréquences ; les fréquences de coupure sont réglées à 250 Hz et 8000 Hz, ce qui correspond aux valeurs par défaut

Page blanche

BIBLIOGRAPHIE

- [1] M. ABRAMOWITZ et I.A. STEGUN. *Handbook of mathematical functions*. Dover, New-York, juin 1974.
- [2] F.T. AGERKVIST. A time-frequency auditory model using wavelet packets. *J. Audio Eng. Soc.*, 44(1-2):37–50, janvier-février 1996.
- [3] H. AIKAKE. A new look at the statistical model. *IEEE transactions on automatic control*, 48:575–595, 1974.
- [4] J.B. ALLEN et L.R. RABINER. A unified approach to short-time Fourier analysis and synthesis. Dans *Proceedings*, volume 65(11), pages 1558–1564. IEEE, 1977.
- [5] H. ALRUTZ et M.R. SCHROEDER. A fast Hadamard transform method for the evaluation of measurements using pseudorandom test signals. Dans *11th international conference*, volume 6, pages 235–238, Paris, 1983. ICA.
- [6] Y. ANDO. *Concert hall acoustics*, chapitre Appendix. Springer-Verlag, Berlin, 1985.
- [7] J.A. ANGUS. Distributed mode loudspeaker radiation mechanisms. Dans *108th convention*, n° 5164 dans Transducers, part two, Paris, février 2000. Audio Eng. Soc.
- [8] F. ASANO et D.C. SWANSON. Sound equalisation in enclosures using modal reconstruction. *J. Acoust. Soc. Am.*, 98(4):2062–2069, octobre 1995.
- [9] B.S. ATAL et M.R. SCHROEDER. Apparent sound source translator. United States Patent 3 236 949, février 1966.
- [10] M. BARRON. Subjective study of british symphony concert halls. *Acustica*, 66(1):1–14, 1988.
- [11] M. BARRON et A.H. MARSHALL. Spatial impression due to early lateral reflections in concert halls: the derivation of a physical measure. *J. Sound and Vibration*, 77(2):211–232, 1981.
- [12] A. BASKIND. Étude de la puissance acoustique rayonnée en champ diffus. Rapport de stage du DEA ATIAM, Université Pierre & Marie CURIE, Laboratoire d’acoustique musicale, juin 1999.
- [13] A. BASKIND et J.D. POLACK. Sound power radiated by sources in diffuse field. Dans *108th convention*, n° 5146 dans Studio, control room and concert hall acoustics, Paris, février 2000. Audio Eng. Soc.
- [14] J.L. BAUCK. A simple loudspeaker array and associated crosstalk canceler for improved 3d audio. *J. Audio Eng. Soc.*, 49(1-2):3–13, janvier-février 2001.
- [15] J.L. BAUCK et D.H. COOPER. Generalized transaural stereo and applications. *J. Audio Eng. Soc.*, 44(9):683–704, septembre 1996.
- [16] A. BELLINI, A. CANNIZZARO, E. UGOLOTTI, A. FARINA, et MORANDI C. Experimental validation of stereo dipole systems inside car cockpits. Dans *International Conference on Signal Processing Applications and Technology*, Orlando, novembre 1999. ICSPAT.

- [17] A. BELLINI, G. CIBELLI, E. UGOLOTTI, F. BRUSCHI, et A. FARINA. Experimental validation of equalizing filters for car cockpits designed with warping techniques. Dans *109th convention*, n° 5278 dans Automotive audio, Los Angeles, septembre 2000. Audio Eng. Soc.
- [18] J.S. BENDAT et A.G. PIERSOL. *Random data - Analysis and measurement procedures*, chapitre 4. Wiley Interscience, New York, 1986.
- [19] J.M. BERMAN et L.R. FINCHAM. The application of digital techniques to the measurement of loudspeakers. *JAES*, 25(6):370–384, juin 1977.
- [20] J. BLAUERT. *Spatial hearing, the psychophysics of human sound localization*. MIT Press, 1983.
- [21] J. BLAUERT et W. LINDEMANN. Auditory spaciousness: some further psychoacoustics analyses. *J. Acoust. Soc. Am.*, 80(2):533–542, 1986.
- [22] G. BLOCH et J.P. JULLIEN. Commande perceptive d’un système d’acoustique variable, application à la salle modulable de l’ircam. Dans *1^{er} Congrès Français d’Acoustique*, volume 2, pages 841–844, Lyon, 1990.
- [23] M.A. BLOMMER et G.H. WAKEFIELD. On the design of a pole-zero approximations using a logarithmic error measure. *IEEE transactions on signal processing*, 42(11):3245–3248, 1994.
- [24] J. BORISH. Self-contained cross-correlation for measuring the impulse response using pseudo-random noise. *JAES*, 33(1):1827–1839, novembre 1985.
- [25] J. BORISH et J.B. ANGELL. An efficient algorithm for measuring the impulse response using pseudo-random noise. *JAES*, 31(7):478–488, juillet-août 1983.
- [26] J.M. BOUROCHE et G. SAPORTA. *Analyse des données*. Que sais-je. PUF, Paris, 1994.
- [27] J.S. BRADLEY. Comparison of concert hall measurements of spatial impression. *J. Acoust. Soc. Am.*, 96(6):3225–3235, 1994.
- [28] J.S. BRADLEY et G.A. SOULODRE. The influence of late arriving energy on spatial impression. *J. Acoust. Soc. Am.*, 97(4):2263–2271, 1995.
- [29] J.S. BRADLEY et G.A. SOULODRE. Objective measures of listener envelopment. *J. Acoust. Soc. Am.*, 98(5):2590–2597, 1995.
- [30] M. BRUNEAU. *Manuel d’acoustique fondamentale*. Hermès, Paris, 1998.
- [31] R. BRUNO. Implantation d’un algorithme de convolution rapide sans retard perceptible dans un environnement de programmation temps-réel. Rapport technique, IRCAM, Acoustique des salles, juin 2000.
- [32] O. CAPPE. *Techniques de réduction de bruit pour la restauration d’enregistrements musicaux*. Thèse de doctorat, École Nationale Supérieure des Télécommunications, Paris, septembre 1993.
- [33] J.D. CAROLL. Degenerate solutions in the nonmetric fitting of a wide class of models for proximity data. *Unpublished manuscript*, 1988. Bell Laboratories, Murray Hill, NJ.
- [34] J.D. CAROLL et J.J. CHANG. Analysis of individual differences in multidimensional scaling via an n-way generalization of “Eckart-Young” decomposition. *Psychometrika*, 35(3), septembre 1970.
- [35] M. CHARBIT. *Éléments de théorie du signal : les signaux aléatoires*. Ellipses, 1990.
- [36] D. CHU. Phase digitizing sharpens timing measurements. *IEEE spectrum*, pages 28–32, 1998.
- [37] W.T. CHU. Impulse response and reverberation decay measurements made by using a periodic pseudorandom sequence. *Applied Acoustics*, 29:193–205, 1990.
- [38] G. CIBELLI, A. BELLINI, E. UGOLOTTI, A. FARINA, et C. MORANDI. Experimental validation of loudspeakers equalization inside car cockpits. Dans *106th convention*, n° 4898 dans Signal processing, part one, Munich, mai 1999. Audio Eng. Soc.
- [39] T. CLAASEN et W. MECKLENBRAUKER. The Wigner distribution : a tool for time-frequency analysis - Part I. *Philips Journal Research*, 35(3):217–250, 1980.
- [40] T. CLAASEN et W. MECKLENBRAUKER. The Wigner distribution : a tool for time-frequency analysis - Part II. *Philips Journal Research*, 35(4/5):276–300, 1980.

- [41] T. CLAASEN et W. MECKLENBRAUKER. The Wigner distribution : a tool for time-frequency analysis - Part III. *Philips Journal Research*, 35(6):372–383, 1980.
- [42] J.J. CLARK, M.R. PALMER, et P.D. LAWRENCE. A transformation method for the reconstruction of functions from non uniformly spaced samples. *IEEE transactions on acoustics, speech, and signal processing*, 33(4):1151–1165, octobre 1985.
- [43] P.M. CLARKSON, J.N. MOURJOPOULOS, et J.K. HAMMOND. Spectral, phase, and transient equalization for audio systems. *J. Audio Eng. Soc.*, 33(3):127–132, mars 1985.
- [44] L. COHEN. Time-frequency distributions - A review. Dans *Proceedings*, volume 77(7), pages 941–981. IEEE, juillet 1989.
- [45] D. COLTON, H.W. ENGL, et A.K. LOUIS. *Surveys on solution methods for inverse problems*. Springer Verlag, Vienne, mai 2000.
- [46] D.H. COOPER et J.L. BAUCK. Prospects for transaural recording. *J. Audio Eng. Soc.*, 37(1-2):3–19, janvier-février 1989.
- [47] D.H. COOPER et J.L. BAUCK. Head diffraction compensated stereo systems. United States Patent 4 893 342, 1990.
- [48] L.D. COSART, L. PEREGRINO, et A. TAMBE. Time domain analysis and its practical application to the measurement of phase noise and jitter. *IEEE transactions on instrumentation and measurement*, 46(4):1016–1019, août 1997.
- [49] L. CREMER et H.A. MÜLLER. *Principles and applications of room acoustics*, volume 1. Applied science, Londres, 1982.
- [50] L. CREMER et H.A. MÜLLER. *Principles and applications of room acoustics*, volume 2. Applied science, Londres, 1982.
- [51] R.E. CROCHIERE et L.R. RABINER. *Multirate digital signal processing*. Prentice-Hall, 1996.
- [52] F. CRUZ. *Évaluation des performances d'un environnement informatique d'acoustique prévisionnelle*. Thèse de doctorat, Université du Maine, Le Mans, décembre 1999.
- [53] P. DAMASKE. Head related two channel stereophony with loudspeaker reproduction. *J. Acoust. Soc. Am.*, 50:1109–1115, 1971.
- [54] J. DANIEL. *Représentation de champs acoustiques, application à la transmission et à la reproduction de scènes sonores complexes dans un contexte multimédia*. Thèse de doctorat, Université Pierre & Marie CURIE, Paris, juin 2000. URL : <http://gyronymo.free.fr>.
- [55] P. D'ANTONIO et J.H. KONNERT. Complex time response measurements using time delay spectrometry. *JAES*, 37(9):674–690, septembre 1989.
- [56] J. DATTORO. The implementation of recursive digital filters for high-fidelity audio. *J. Audio Eng. Soc.*, 36(11):851–877, novembre 1988.
- [57] J. DATTORO. Effect design : reverberator and other filters. *J. Audio Eng. Soc.*, 45(9):660–684, septembre 1997.
- [58] H.A. DAVID. *The method of paired comparisons*, chapitre 1,4. Griffin, second édition, 1988.
- [59] M.F. DAVIS. Loudspeaker systems with optimized wide listening area imaging. *J. Audio Eng. Soc.*, 35(11):888–896, novembre 1987.
- [60] J. DE LEEUW et J. MEULMAN. A special jackknife for multidimensional scaling. *Journal of classification*, 3:97–112, 1986.
- [61] D. DE VRIES, E. HULSEBOS, et J. BAAN. Spatial fluctuations of spaciousness measures in auditoria. Dans *108th convention*, n° 5147 dans Studio, control room and concert hall acoustics, Paris, février 2000. Audio Eng. Soc.
- [62] J.P. DELMAS. *Éléments de théorie du signal : les signaux déterministes*. Ellipses, 1991.
- [63] G. DEMOMENT. Problèmes inverses en traitement du signal et de l'image. Cours du magistère « Electronique, électrotechnique et automatique », Université de Paris-Sud & École Normale Supérieure de Cachan, 1997.

- [64] P. DÉROGIS. *Analyse des vibrations et du rayonnement de la table d'harmonie d'un piano droit et conception d'un système de reproduction du champ acoustique*. Thèse de doctorat, Université du Maine, Le Mans, octobre 1997.
- [65] C. DUNN et M.O. HAWKSFIRD. Distorsion immunity of MLS derived impulse response measurements. *JAES*, 41(5):314–335, mai 1993.
- [66] P. DUTILLEUX. Simple to operate digital time-varying filters. Dans *86th convention*, n° 2757. Audio Eng. Soc., mars 1989.
- [67] K.J. EBELLING, K. FREUDENSTEIN, et A. ALRUTZ. Experimental investigations of statistical properties of diffuse sound fields in reverberation rooms. *Acustica*, 51(3):145–153, 1982.
- [68] B. EFRON et R.J. TIBSHIRANI. *An introduction to the bootstrap*. Chapman & Hall, 1993.
- [69] S.J. ELLIOTT, L.P. BHATIA, F.S. DEGHAN, A.H. FU, M.S. STEWART, et D.W. WILSON. Practical implementation of low-frequency equalization using adaptative digital filters. *J. Audio Eng. Soc.*, 42(12):988–998, décembre 1994.
- [70] S.J. ELLIOTT, C.C. BOUCHER, et P.A. NELSON. The behavior of a multiple channel active control system. *IEEE transactions on signal processing*, 40:1041–1052, 1992.
- [71] S.J. ELLIOTT et P.A. NELSON. Multiple point equalization in a room using adaptative digital filters. *J. Audio Eng. Soc.*, 37(11):899–907, novembre 1989.
- [72] M. EMERIT. *Simulation binaurale de l'acoustique des salles de concert*. Thèse de doctorat, Institut National Polytechnique de Grenoble & Centre Scientifique et Technique du Bâtiment, Grenoble, septembre 1995.
- [73] L.R. ERIKSSON, M.C. ALLIE, et R.A. GREINER. The selection and application of an IIR adaptive filter for use in active sound attenuation. *IEEE transactions on acoustics, speech, and signal processing*, 35:433–437, 1986.
- [74] A. FARINA. Simultaneous measurement of impulse response and distortion with a swept-sine technique. Dans *108th convention*, n° 5093 dans Audio measurement and instrumentation, Paris, février 2000. Audio Eng. Soc.
- [75] A. FARINA, A. BELLINI, et E. ARMELLONI. Implementation of cross-talk cancelling filters with warped structures. Dans *International DSP conference*, Boston, septembre 2000. SHARC.
- [76] A. FARINA, A. BELLINI, et E. ARMELLONI. Non-linear convolution : a new approach for the auralization of distorting systems. Dans *110th convention*, n° 5359 dans Signal processing for audio, part 1, Amsterdam, mai 2001. AES.
- [77] A. FARINA, A. LANGHOFF, et L. TRONCHIN. Acoustic characterisation of virtual musical instruments using MLS technique on ancient violins. *J. New Music Research*, 27(4):359–379, décembre 1998.
- [78] A. FARINA et E. UGOLOTTI. Spatial equalisation of sound systems in cars by digital inverse filtering. Dans *4th international conference and exhibition*, Audio in automotive spaces, Bologne, octobre 1997. Comfort in the automotive industry.
- [79] A. FARINA et E. UGOLOTTI. Use of digital inverse filtering techniques for improving car audio systems. Dans *103rd convention*, n° 4569 dans 3D binaural audio technology, New-York, septembre 1997. Audio Eng. Soc.
- [80] A. FARINA et E. UGOLOTTI. Automatic measurement system for car audio applications. Dans *104^{ème} convention*, n° 4692 dans Spatial perception and processing, part two, Amsterdam, mai 1998. AES. URL : <http://www.ramsete.com/aurora>.
- [81] A. FARINA et E. UGOLOTTI. Numerical model of the sound field inside cars for the creation of virtual audible reconstructions. Dans *DAFX*, n° 98 dans Poster session 2, Barcelone, novembre 1998. Digital audio effects.
- [82] A. FARINA et E. UGOLOTTI. Spatial equalization of sound systems in cars. Dans *15th international conference*, Audio in automotive spaces, Copenhague, octobre 1998. Audio Eng. Soc.

- [83] A. FARINA, E. UGOLOTTI, A. BELLINI, G. CIBELLI, et C. MORANDI. Inverse numerical filters for linearisation of loudspeaker's response. Dans *DAFX*, n° 98 dans Filtering, delay and modulation, Barcelone, novembre 1998. Digital audio effects.
- [84] A. FARINA, E. UGOLOTTI, A. BELLINI, G. CIBELLI, et C. MORANDI. Non-linear digital audio processor for dedicated loudspeaker systems. *IEEE transactions on consumer electronics*, 44(3), août 1998.
- [85] A. FISCETTI. *Caractérisation perceptive d'une salle à géométrie variable : insuffisance des critères traditionnels*. Thèse de doctorat, Université du Maine, Le Mans, février 1990.
- [86] P. FLANDRIN. Principe et mise en œuvre de l'analyse temps-fréquence par transformation de Wigner-Ville. Dans *Traitement du signal*, volume 2, pages 143–151, 1985.
- [87] P. FLANDRIN et J.P. SESSAREGO. Méthodes temps-fréquence en acoustique. Dans *1^{er} Congrès Français d'Acoustique*, volume 2, pages 707–716, Lyon, 1990.
- [88] K.C. FOO, M.O. HAWKSFORD, et M.P. HOLLIER. Optimization of virtual sound reproduced using two loudspeakers. Dans *16th international conference*, Binaural technology : theory and implementation, part two, Rovaniemi, avril 1999. Audio Eng. Soc.
- [89] B. FRIEDLANDER et B. PORAT. The modified Yule-Walk method of ARMA spectral estimation. *IEEE transactions on aerospace electronic systems*, 20(2):158–173, mars 1984.
- [90] J. GARAS et P.C. SOMMEN. Warped linear time invariant systems and their application in audio signal processing. Dans *ICASSP*, Phoenix, mars 1999. IEEE.
- [91] B. GARDNER. *3D audio using loudspeakers*. Thèse de doctorat, Massachusetts Institute of Technology, Cambridge, septembre 1997.
- [92] B. GARDNER et K. MARTIN. HRTF measurements of a KEMAR dummy-head microphone. Rapport technique 280, Massachusetts Institute of Technology, Media Laboratory, Cambridge, 1991. Mesures disponibles à l'URL <http://sound.media.mit.edu/KEMAR.html>.
- [93] W.G. GARDNER. Efficient convolution without input/output delay. *J. Audio Eng. Soc.*, 43(3):127–136, 1995.
- [94] W.G. GARDNER. Efficient convolution without input/output delay. *J. Audio Eng. Soc.*, 43(3):127–136, 1995.
- [95] M. GERZON. Practical periphony: the reproduction of full sphere sound. Dans *65^{ème} convention*, n° 1571, Londres, février 1980. Audio Eng. Soc.
- [96] A. GILLOIRE et J.P. JULLIEN. L'acoustique des salles dans les télécommunications. *L'écho des recherches*, 127:43–54, 1987.
- [97] B.R. GLASBERG et B.C.J. MOORE. Derivation of auditory filter shapes from notched-noise data. *Hearing Research*, 47:103–138, 1990.
- [98] M.A. GOLD. Subjective evaluation of spatial impression: the importance of lateralization. Dans *International conference*, Lauderdale, Florida, Mai 1994. NOISE-CON.
- [99] G.H. GOLUB et C.F. VAN LOAN. *Matrix computation*, chapitre 4. Johns Hopkins, Londres, 3^{ème} édition, 1996.
- [100] E. GRANIER. Comparing and optimizing audio systems in cars. Dans *100th convention*, n° 4283 dans Measurements and instrumentation, part 2, Copenhague, mai 1996. Audio Eng. Soc.
- [101] E. GRANIER. Limites du mannequin acoustique. Note de service n° H26/97.57, Renault - Direction de la Recherche, Rueil-Malmaison, avril 1997.
- [102] E. GRANIER, M. KLEINER, B.E. DALENBÄCK, et P. SVENSSON. Experimental auralization of car audio installations. *J. Audio Eng. Soc.*, 44(10):835–849, octobre 1996.
- [103] R.M. GRAY. On the asymptotic eigenvalue distribution of Toeplitz matrices. *IEEE transactions on information theory*, 18:725–730, 1972.
- [104] U. GRENANDER et G. SZEGÖ. *Toeplitz forms and their applications*. AMS Chelsea publishing, Providence, 2001.

- [105] D. GRIESINGER. Spaciousness and localization in listening rooms and their effects on the recording technique. *J. Audio Eng. Soc.*, 34(4):255–268, 1986.
- [106] D. GRIESINGER. Equalization and spatial equalization of dummy-head recordings for loudspeaker reproduction. *J. Audio Eng. Soc.*, 37(1-2):20–29, janvier-février 1989.
- [107] D. GRIESINGER. The psychoacoustics of apparent source width, spaciousness, and envelopment in performance spaces. *Acustica*, 83(4):721–731, 1997.
- [108] D. GRIESINGER. Spaciousness and envelopment in musical acoustics. Dans *101st convention*, n° 4401 dans Room acoustics and simulation techniques, Los Angeles, novembre 1997. Audio Eng. Soc.
- [109] D. GRIESINGER. Spatial impression en envelopment in small rooms. Dans *103rd convention*, n° 4638 dans Room acoustics and simulation techniques, New-York, septembre 1997. Audio Eng. Soc.
- [110] D. GRIESINGER. General overview of spatial impression, envelopment, localization, and externalization. Dans *15th international conference*, Perception of sound in small spaces, Copenhagen, octobre 1998. Audio Eng. Soc.
- [111] D. GRIESINGER. Speaker placement, externalization, and envelopment in home listening rooms. Dans *105th convention*, n° 4860 dans Multichannel sound, San Francisco, septembre 1998. Audio Eng. Soc.
- [112] D. GRIESINGER. Objective measure of spaciousness and envelopment. Dans *16th international conference*, Invited papers, Rovaniemi, avril 1999. Audio Eng. Soc.
- [113] V. GRILLON. *Auralisation dans les maquettes, traitement des réponses impulsionnelles*. Thèse de doctorat, Université du Maine, Le Mans, avril 1995.
- [114] J. GUÉGAND et T. DUGARDIN. *Algèbre en classes préparatoires, MP*/MP*. Ellipses, Paris, octobre 1999.
- [115] Y. HANEDA. Common acoustical pole and zero modeling of room transfer functions. *IEEE transactions on speech and audio processing*, 2(2):320–328, 1994.
- [116] Y. HANEDA, S. MAKINO, Y. KANEDA, et N. KOIZUMI. Arma modeling of a room transfer function at low frequencies. *J. Acoust. Soc. Jpn.*, 15(5):353–355, 1994.
- [117] C. HANSEN. *Rank-deficient and discret ill-posed problems*. SIAM, Philadelphia, juin 1998.
- [118] T. HANYU et S. KIMURA. Effects of spatial information of a sound field on listener envelopment. Dans *135th meeting*, n° 4pAAb4 dans Performance spaces, pages 2123–2124, Seattle, june 1998. ASA.
- [119] F.J. HARRIS. On the use of windows for harmonic analysis with the discrete Fourier transform. Dans *proc. IEEE*, volume 66(1), pages 51–83, 1978.
- [120] W. HARTMANN. *Signals, sound and sensation*. Springer-Verlag, New-York, 1998.
- [121] P.D. HATZIANTONIOU et J.N. MOURJOPOULOS. Generalized fractional-octave smoothing of audio and acoustic responses. *J. Audio Eng. Soc.*, 48(4):259–279, 2000.
- [122] P. HERZOG, A. SOTO-NICOLAS, et F. GUÉRY. Passive and active control of the low-frequency modes in a small room. Dans *98^{ème} convention*, n° 3951 dans Architectural acoustics, part II, Paris, février 1995. Audio Eng. Soc.
- [123] R.C. HEYSER. Acoustical measurements by time delay spectrometry. *JAES*, 15(4):370–382, octobre 1967.
- [124] T. HIDAKA, L. BERANEK, et T. OKANO. Interaural cross-correlation as a measure of spatial impression. Dans *124th meeting*, n° 5pAA3 dans Architectural acoustics, New-Orleans, octobre 1992. ASA.
- [125] T. HIDAKA, L. BERANEK, et T. OKANO. Interaural cross-correlation, lateral fraction, and low and high-frequency sound levels as measures of acoustical quality in concert halls. *J. Acoust. Soc. Am.*, 98(2):998–1007, 1995.

- [126] F. HLAWATSCH et G.F. BOUDREAUX-BARTELS. Linear and quadratic time-frequency signal representations. *IEEE signal processing magazine*, pages 21–67, avril 1992.
- [127] A.C. HOPE. A simplified Monte Carlo significance test procedure. *Journal of the Royal Statistical Society*, B(30):582–598, 1968.
- [128] A. HÄRMA, M. KARJALAINEN, L. SAVIOJA, V. VÄLIMÄKI, U.K. LAINE, et J. HUPANIEMI. Frequency-warped signal processing for audio applications. *J. Audio Eng. Soc.*, 48(11):1011–1029, novembre 2000.
- [129] T. IRINO et H. KAWAHARA. Signal reconstruction from modified auditory wavelet transform. *IEEE transactions on signal processing*, 41(12):3549–3554, 1993.
- [130] T. IRINO et R.D. PATTERSON. A time-domain level dependant auditory filter : the gammachirp. *J. Acoust. Soc. Am.*, 101:412–419, 1997.
- [131] ISO. Measurement of the reverberation time of rooms with reference to other acoustical parameters. Acoustics in building 3382:1997, ISO, Genève, juin 1997. URL: <http://www.iso.ch>.
- [132] M. IWAHARA et T. MORI. Stereophonic sound reproduction system. United States Patent 4 118 599, 1978.
- [133] D.S. JAGGER. Recent developments and improvements in SOUNDFIELD microphone technology. Dans *75^{ème} convention*, n° 2064, Paris, mars 1984. Audio Eng. Soc.
- [134] H. JEONG et J.G. IH. Implementation of a new algorithm using the STFT with variable frequency resolution for the time-frequency auditory model. *J. Audio Eng. Soc.*, 47(4):240–251, avril 1999.
- [135] D. JOHNSTON. *CoolEditPro v1.0 application programmer's interface*. Syntrillium Software Corporation, 1998.
- [136] D. JOHNSTON. *CoolEditPro v1.1 user's guide*. Syntrillium Software Corporation, 1998.
- [137] V.L. JORDAN. A group of objective acoustical criteria for concert halls. *Applied Acoustics*, 14:253–266, 1981.
- [138] J.M. JOT. *Étude d'un spatialisateur de sons par modèles physiques et perceptifs*. Thèse de doctorat, École Nationale Supérieure des Télécommunications, Paris, septembre 1992.
- [139] J.M. JOT, L. CERVEAU, et O. WARUSFEL. Analysis and synthesis of room reverberation based on a statistical time-frequency model. Dans *103th convention*, n° 4629 dans Room acoustics and simulation techniques, New-York, septembre 1997. Audio Eng. Soc.
- [140] J.M. JOT, J.P. JULLIEN, et O. WARUSFEL. Method to simulate the acoustical quality of a room and associated audio-digital processor. United States Patent 5 812 674, septembre 1998.
- [141] J.M. JOT, V. LARCHER, et O. WARUSFEL. Digital signal processing issues in the context of binaural and transaural stereophony. Dans *98th convention*, n° 3980 dans I, Paris, février 1995. Audio Eng. Soc.
- [142] J.M. JOT, S. WARDLE, et V. LARCHER. Approaches to binaural synthesis. Dans *105th convention*, n° 4861 dans Spatial perception and processing, part two, San Francisco, septembre 1998. Audio Eng. Soc.
- [143] J.M. JOT et O. WARUSFEL. Le spatialisateur. Dans *Actes du colloque Son et Espace*, Lyon, avril 1995.
- [144] J.M. JOT et O. WARUSFEL. A real-time spatial sound processor for music and virtual reality applications. Dans *International Computer Music Congress*, Banff, septembre 1995.
- [145] J.P. JULLIEN. Correlations among objective criteria of room acoustic quality. Dans *12th international congress*, volume 2 de *Architectural and building acoustics*, pages E4–9, Toronto, juillet 1986. ICA.
- [146] J.P. JULLIEN, A. GILLOIRE, et A. SALIOU. Caractérisation d'une méthode de mesure de réponse impulsionnelle en acoustique des salles. Dans *11th international conference*, volume 6, pages 217–220, Paris, 1983. ICA.

- [147] Y. KAHANA, P.A. NELSON, et O. KIRKEBY. Multichannel sound reproduction using a four-ear dummy head. Dans *102nd convention*, n° 4465 dans Multichannel sound, Munich, mars 1997. Audio Eng. Soc.
- [148] Y. KAHANA, P.A. NELSON, et O. KIRKEBY. Objective and subjective assessment of systems for the production of virtual acoustic images for multiple listeners. Dans *103rd convention*, n° 4573 dans 3D binaural audio technology, New-York, septembre 1997. Audio Eng. Soc.
- [149] Y. KAHANA, P.A. NELSON, et S. YOON. Experiments on the synthesis of virtual acoustic sources in automotive interiors. Dans *16th international conference*, Spatial sound reproduction and application, Rovaniemi, avril 1999. Audio Eng. Soc.
- [150] E. KAHLE. *Validation d'un modèle objectif de la perception de la qualité acoustique sur un ensemble de salles de concert et d'opéras*. Thèse de doctorat, Université du Maine, Le Mans, mai 1995.
- [151] Y. KANEDA. A study of non linear effect on acoustic impulse response measurement. *J. Acoust. Soc. Jp.*, 16(3):193–195, 1995.
- [152] L.J. KARAM et J.H. MACCLELLAN. Complex Chebyshev approximation for FIR filter design. *IEEE transactions on circuits and systems II*, mars 1995.
- [153] M. KARJALAINEN, E. PIIRILÄ, A. JÄRVINEN, et J. HUOPANIEMI. Realizable warped IIR filters and their properties. Dans *ICASSP*, Munich, avril 1997. IEEE.
- [154] M. KARJALAINEN, E. PIIRILÄ, A. JÄRVINEN, et J. HUOPANIEMI. Comparison of loudspeaker equalization methods based on DSP techniques. *J. Audio Eng. Soc.*, 47(1-2):14–30, janvier-février 1999.
- [155] F. KAWAKAMI et K. YAMAGUCHI. Space-ensemble average of reverberation decay curves. *J. Acoust. Soc. Am.*, 70(4):1071–1082, octobre 1981.
- [156] G.S. KENDALL. The decorrelation of audio signals and its impact on spatial imagery. *Computer Music Journal*, 19(4):71–87, 1995.
- [157] O. KIRKEBY et P.A. NELSON. Reproduction of plane wave sound fields. *J. Acoust. Soc. Am.*, 94:2992–3000, 1993.
- [158] O. KIRKEBY et P.A. NELSON. Virtual source imaging using the stereo dipole. Dans *103rd convention*, n° 4574 dans 3D binaural audio technology, New-York, septembre 1997. Audio Eng. Soc.
- [159] O. KIRKEBY et P.A. NELSON. Digital filter design for inversion problems in sound reproduction. *J. Audio Eng. Soc.*, 47(7-8):583–595, juillet-août 1999.
- [160] O. KIRKEBY, P.A. NELSON, et H. HAMADA. The stereo dipole : a virtual source imaging system using two closely spaced loudspeakers. *J. Audio Eng. Soc.*, 46(5):387–395, mai 1998.
- [161] O. KIRKEBY, P.A. NELSON, H. HAMADA, et F. ORDUÑA BUSTAMENTE. Fast deconvolution of multichannel systems using regularization. Rapport technique 255, ISVR, Southampton, 1996.
- [162] O. KIRKEBY, P.A. NELSON, H. HAMADA, et F. ORDUÑA BUSTAMENTE. Fast deconvolution of multichannel systems using regularization. *IEEE transactions on speech and audio processing*, 6(2):189–194, mars 1998.
- [163] O. KIRKEBY, P.A. NELSON, F. ORDUÑA BUSTAMENTE, et H. HAMADA. Local sound field reproduction using digital signal processing. *J. Acoust. Soc. Am.*, 100(3):1584–1593, septembre 1996.
- [164] O. KIRKEBY, P. RUBAK, et A. FARINA. Analysis of ill-conditioning of multichannel deconvolution problems. Dans *ASSP Workshop*, Mohonk, octobre 1999. IEEE.
- [165] O. KIRKEBY, P. RUBAK, L.G. JOHANSEN, et P.A. NELSON. Implementation of cross-talk cancellation networks using warped fir filters. Dans *16th international conference*, Binaural technology : theory and implementation, part two, Rovaniemi, avril 1999. Audio Eng. Soc.

- [166] O. KIRKEBY, P. RUBAK, P.A. NELSON, et A. FARINA. Design of crosstalk cancellation networks by using fast deconvolution. Dans *106th convention*, n° 4916 dans Signal processing, part two, Munich, may 1999. Audio Eng. Soc.
- [167] A. KIRSCH. *An introduction to the mathematical theory of inverse problems*. Springer Verlag, Vienne, octobre 1996.
- [168] M. KLEINER. A new way of measuring the lateral energy fraction. *Applied Acoustics*, 27:321–327, 1989.
- [169] J.B. KRUSKAL et M. WISH. *Multidimensional scaling*. Sage, Beverly Hills, Calif, 1978.
- [170] K. KUROZUMI et K. OHGUSI. The relationship between the cross-correlation coefficient of two channel acoustic signals and sound image quality. *J. Acoust. Soc. Am.*, 74(6):1726–1733, 1983.
- [171] H. KUTTRUFF. Sound field in small rooms. Dans *15th international conference*, Overview of the physics and modeling of small spaces, Copenhague, octobre 1998. AES.
- [172] H. KUTTRUFF. *Room acoustics*. Spon Press, Londres, 4^{ème} édition, octobre 2000.
- [173] A. LABORIE. Capture, échantillonnage, et manipulations de fonctions de directivité - Application à l'enregistrement de scènes sonores 3d avec un micro SOUNDFIELD. Rapport de stage du Mastère TDS, École Nationale Supérieure de Télécommunications, IRCAM - Équipe acoustique des salles, octobre 2000.
- [174] V. LARCHER. Interpolation de filtres audio-numériques pour application à la reproduction spatiale des sons sur écouteurs. Rapport technique, École Nationale Supérieure des Télécommunications, Département Signal, décembre 1994.
- [175] V. LARCHER. Correction individuelle pour la reproduction binaurale d'enregistrements effectués dans l'habitacle d'un véhicule automobile. Rapport technique, IRCAM, mai 1996.
- [176] V. LARCHER. *Techniques de spatialisation des sons pour la réalité virtuelle*. Thèse de doctorat, Université Pierre & Marie CURIE, Paris, mai 2001.
- [177] V. LARCHER, J.M. JOT, J. GUYARD, et O. WARUSFEL. Study and comparison of efficient methods for 3D audio spatialization based on linear decomposition of HRTF data. Dans *108th convention*, n° 5097 dans Sound reproduction systems, Paris, février 2000. Audio Eng. Soc.
- [178] V. LARCHER, G. VANDERNOOT, et J.M. JOT. Equalization methods in binaural technology. Dans *105th convention*, n° 4858 dans Spatial perception and processing, part two, San Francisco, septembre 1998. Audio Eng. Soc.
- [179] O. LARTILLOT. Implantation de méthodes de mesures asynchrones de réponses impulsionnelles. Rapport technique, École Supérieure d'Électricité & Renault, Section signaux, images et formes, Metz, juillet 1999. Rapport de stage.
- [180] O. LARTILLOT. Mesure de la réponse impulsionnelle d'un canal acoustique par méthodes asynchrones. Rapport technique, École Supérieure d'Électricité & Renault, Section signaux, images et formes, Metz, mai 1999. Étude bibliographique.
- [181] C. LAVANDIER. *Validation perceptive d'un modèle objectif de caractérisation de la qualité acoustique des salles*. Thèse de doctorat, Université du Maine, Le Mans, juin 1989.
- [182] F. LAVILLE, M. SIDKI, et J. NICOLAS. Spherical acoustical holography using sound intensity measurements: theory and simulation. *Acustica*, 76:193–198, 1992.
- [183] Y.H. LIEW, J. YANG, S.E. TAN, et W.S. GAN. Power improvement in crosstalk cancellation using psychoacoustic frequency masking. Dans *109th convention*, n° 5203 dans Perception and psychoacoustics, part one, Los Angeles, septembre 2000. Audio Eng. Soc.
- [184] J.S. LIM et C. KYRIAKAKIS. Virtual loudspeaker rendering for multiple listeners. Dans *109th convention*, n° 5183 dans Digital signal processing, part two, Los Angeles, septembre 2000. Audio Eng. Soc.
- [185] L. LJUNG et T.L. SODERSTROM. The Steiglitz-MacBride algorithm revisited ; convergence, analysis and accuracy aspects. *IEEE transactions on automatic control*, 16(3):712–717, 1981.

- [186] J.J. LÓPEZ, F. ORDUÑA, et A. GONZÁLEZ. Equalization zones for crosstalk cancellation as a function of loudspeaker position and room acoustics. Dans *16th international conference*, Binaural technology : theory and implementation, part two, Rovaniemi, avril 1999. Audio Eng. Soc.
- [187] J.J. LÓPEZ, F. ORDUÑA, et A. GONZÁLEZ. Modeling and measurement of crosstalk cancellation zones for small displacements of the listener in transaural sound reproduction with different loudspeaker arrangements. Dans *109th convention*, n° 5267 dans Instrumentation and measurement, Los Angeles, septembre 2000. Audio Eng. Soc.
- [188] D. MACGRATH. Huron : a digital audio convolution workstation. Dans *98th convention*, n° 4023 dans Audio signal processing, part 1, Paris, février 1995. Audio Eng. Soc.
- [189] M. MARIN. Étude de la localisation en restitution des sons. Application à l'amélioration de la qualité des systèmes de téléconférence. Rapport technique, CNET, LAA/TSS/CMC, avril 1993.
- [190] W.L. MARTENS. The impact of decorrelated low-frequency reproduction on auditory spatial imagery: are two subwoofers better than one? Dans *16th international conference*, Perception of spatial sound, Rovaniemi, avril 1999. Audio Eng. Soc.
- [191] J. MARTIN. Binaural simulation of concert halls: a new approach for the binaural reverberation process. *J. Acoust. Soc. Am.*, 94(6):3255–3264, 1993.
- [192] N. MARTIN, J. MARS, J. MARTIN, et C. CHORIER. A Capons's time octave representation application in room acoustics. *IEEE transactions on signal processing*, 43:1842–1854, 1995.
- [193] N. MARTIN, J. MARS, J.C. PIERRON, et J. MARTIN. Représentation temps-fréquence et acoustique des salles. Dans *13^{ème} colloque sur le traitement du signal et ses applications*, volume 1, Nice, 1991. GRETSI.
- [194] D.C. MASSIE. An engineering study of the four-multiply normalized ladder filter. *J. Audio Eng. Soc.*, 41(7-8):564–582, juillet-août 1993.
- [195] J. MAX et M. DIOT. Comparaison de deux principales méthodes d'analyse spectrale : la méthode du périodogramme moyenné et la méthode du corrélogramme. Dans *10^{ème} colloque sur le traitement du signal et ses applications*, volume 1, pages 465–469, Nice, 1985. GRETSI.
- [196] S. MEHRGARDT et V. MELLER. Transformation characteristics of the external human hear. *J. Acoust. Soc. Am.*, 66:1567–1576, juin 1977.
- [197] M. MEIN. Perception de l'information binaurale liée aux réflexions précoces dans une salle. Application à la simulation de la qualité acoustique. Rapport de stage du DEA d'acoustique appliquée, Université du Maine, Le Mans, juin 1993.
- [198] D. MERCIER. *Le livre des techniques du son*, volume 3. Fréquences, Paris, 2^{ème} édition, 1998.
- [199] P. MINNAAR, S.K. OLESEN, F. CHRISTENSEN, et H. MØLLER. Localization with binaural recordings from artificial and human heads. *J. Audio Eng. Soc.*, 49(5):323–336, mai 2001.
- [200] M. MIRANDA et P. TILLI. Asymptotic spectra of Hermitian block Toeplitz matrices and preconditioning results. *SIAM J. Matrix Anal. Applic.*, 21(3):867–881, 2000.
- [201] S.K. MITRA. Structural subband decomposition : a new concept in digital signal processing. Dans *ICASSP*, Munich, avril 1997. IEEE.
- [202] M. MIYOSHI et Y. KANEDA. Inverse filtering of room acoustics. *IEEE transactions on acoustics, speech, and signal processing*, 36(2):145–152, 1988.
- [203] A. MOHAMAD-DJAFARI. Problèmes inverses. Cours du DEA « Automatique et traitement du signal » Document n° 11013, École Supérieure d'Electricité, Laboratoire signaux et systèmes, Gif-sur-Yvette, juillet 1999. URL : <http://www.supelec.fr/lss/perso/adjafari/index.htm>.
- [204] H. MØLLER. Reproduction of artificial head recordings through loudspeaker. *J. Audio Eng. Soc.*, 37(1-2):30–33, janvier-février 1989.
- [205] H. MØLLER. Fundamentals of binaural technology. *Applied Acoustics*, 36:171–218, 1992.
- [206] H. MØLLER, D. HAMMERSHØI, C. JENSEN, et M. SØRENSEN. Evaluation of artificial heads in listening tests. *J. Audio Eng. Soc.*, 47(3):83–99, mars 1999.

- [207] H. MØLLER, C. JENSEN, D. HAMMERSHØI, et M. SØRENSEN. Design criteria for headphones. *J. Audio Eng. Soc.*, 43(4):218–232, 1995.
- [208] H. MØLLER, C. JENSEN, D. HAMMERSHØI, et M. SØRENSEN. Using a typical human subject for binaural recording. Dans *100th convention*, n° 4157 dans Psychoacoustics, hearing and hearing instruments, Copenhague, mai 1996. Audio Eng. Soc.
- [209] H. MØLLER, M. SØRENSEN, C. JENSEN, et D. HAMMERSHØI. Binaural technique : do we need individual recordings? *J. Audio Eng. Soc.*, 44(6):451–469, 1996.
- [210] E. MOMMERTZ et S. MÜLLER. Applying the inverse fast Hadamard transform to improve MLS measurements. Dans *15th international conference*, volume 4, pages 127–130, Trondheim, juin 1995. ICA.
- [211] E. MOMMERTZ et S. MÜLLER. Measuring impulse responses with digitally pre-emphasized pseudorandom noise derived from maximum length sequences. *Applied Acoustics*, 44:195–214, 1995.
- [212] J.A. MOORER. About this reverberation business. *Computer Music Journal*, 3(2):13–18, 1979.
- [213] M. MORIMOTO. The role of rear loudspeakers in spatial impression. Dans *103rd convention*, n° 4554 dans Room acoustics and simulation techniques, New-York, septembre 1997. Audio Eng. Soc.
- [214] M. MORIMOTO et K. IIDA. Effects of front/back energy ratios of early and late reflections on listener envelopment. Dans *135th meeting*, n° 1aAA2 dans Sound field for the audience, part one, pages 7–8, Seattle, june 1998. ASA.
- [215] M. MORIMOTO et Z. MAEKAWA. Auditory spaciousness and envelopment. Dans *13th international conference*, pages 215–218, Belgrade, 1989. ICA.
- [216] M. MORIMOTO et C. PÖSSELT. Contribution of reverberation to auditory spaciousness in concert halls. *J. Acoust. Soc. Jp.*, 10(2):87–92, 1989.
- [217] M. MORIMOTO, S. SUGIURA, et K. IIDA. Relation between auditory source width in various sound fields and degree of interaural cross-correlation: confirmation by constant method. *Applied Acoustics*, 42:233–238, 1994.
- [218] M. MORIMOTO et M. WATANABE. Directional dependance of the change of auditory source width by very short time-delay reflections. Dans *135th meeting*, n° 4pAAb10 dans Performance spaces, pages 2135–2136, Seattle, june 1998. ASA.
- [219] P. MOROSINI. Méthodes d'estimation de l'écart entre deux fréquences d'échantillonnage. Rapport technique, École Supérieure d'Electricité & Renault, Section signaux, images et formes, Metz, mai 1999.
- [220] P.M. MORSE et U.K. INGARD. *Theoretical acoustics*. Mac Graw-Hill, New-York, 1968.
- [221] J. MOURJOPOULOS. *The removal of reverberation from signals*. Thèse de doctorat, Université de Southampton, Southampton, 1984.
- [222] J. MOURJOPOULOS. On the variability and invertibility of room impulse response functions. *J. Sound and Vibration*, 102(2):217–228, 1985.
- [223] J. MOURJOPOULOS. Digital equalization methods for audio systems. Dans *84th convention*, n° 2598, Paris, mars 1988. Audio Eng. Soc.
- [224] J. MOURJOPOULOS. Digital equalization of room acoustics. *J. Audio Eng. Soc.*, 42(11):884–900, novembre 1994.
- [225] J. MOURJOPOULOS, P.M. CLARKSON, et J.K. HAMMOND. A comparative study of least-squares and homomorphic techniques for the inversion of mixed-phase signals. Dans *ICASSP*, pages 1858–1861. IEEE, 1982.
- [226] J. MOURJOPOULOS et M.A. PARASKEVAS. Pole and zero modelinf og room transfer functions. *J. Sound and Vibration*, 146(2):281–302, 1991.

- [227] J. MOURJOPOULOS, A. TSOPANOGLIOU, et N. FAKOTAKIS. A vector quantization approach for room transfer function classification. Dans *ICASSP*, pages 3593–3596. IEEE, 1991.
- [228] J.N. MOURJOPOULOS, E.D. KYRIAKIS-BITZAROS, et C.E. GOUTIS. Theory and real-time implementation of time varying digital audio filters. *J. Audio Eng. Soc.*, 38(7):523–536, 1990.
- [229] S.T. NEELY et J.B. ALLEN. Invertibility of a room impulse response. *J. Acoust. Soc. Am.*, 66(1):165–169, 1979.
- [230] P.A. NELSON. Active control of acoustic fields and the reproduction of sound. *J. Sound Vib.*, 177:447–477, 1994.
- [231] P.A. NELSON et S.J. ELLIOTT. *Active control of sound*. Academic Press, Londres, juin 1997.
- [232] P.A. NELSON, H. HAMADA, et S.J. ELLIOTT. Adaptive inverse filters for stereophonic sound reproduction. *IEEE transactions on signal processing*, 40(7):1621–1632, juillet 1992.
- [233] P.A. NELSON, F. ORDUÑA BUSTAMANTE, D. ENGLER, et H. HAMADA. Virtual source imaging for in-car sound reproduction. Dans *Soc. Auto. Eng.*, n° 95A1017 dans Automobile audio systems, 1995.
- [234] P.A. NELSON, F. ORDUÑA BUSTAMANTE, D. ENGLER, et H. HAMADA. Experiments on a system for the synthesis of virtual acoustic sources. *J. Audio Eng. Soc.*, 44(11):990–1007, novembre 1996.
- [235] P.A. NELSON, F. ORDUÑA BUSTAMANTE, et H. HAMADA. Inverse filter design and equalization zones in multichannel sound reproduction. *IEEE transactions on speech and audio processing*, 3(3):185–192, 1995.
- [236] P.A. NELSON, F. ORDUÑA BUSTAMANTE, et H. HAMADA. Multichannel signal processing techniques in the reproduction of sound. *J. Audio Eng. Soc.*, 44(11):973–989, novembre 1996.
- [237] R. NICOL. *Étude de la restitution du son spatialisé dans une zone étendue - Application à la téléprésence*. Thèse de doctorat, Université du Maine, Le Mans, décembre 1999.
- [238] 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. Mémoire d'ingénieur, Conservatoire National des Arts et Métiers, Paris, septembre 1999.
- [239] A.H. NUTTALL. Some windows with very good sidelobe behavior. *IEEE transactions on acoustics, speech, and signal processing*, 29(1):84–91, février 1981.
- [240] T. OKANO, L. BERANEK, et T. HIDAKA. Relations among interaural cross-correlation coefficient (IACC), lateral fraction (LF), and apparent source width (ASW) in concert halls. *J. Acoust. Soc. Am.*, 104(1):255–265, 1998.
- [241] A.V. OPPENHEIM, D.H. JOHNSON, et K. STEIGLITZ. Computation of spectra with unequal resolution using the Fast Fourier Transform. *IEEE*, pages 299–301, 1971.
- [242] A.V. OPPENHEIM, G.E. KOPEC, et J.M. TRIBOLET. Signal analysis by homomorphic prediction. Dans *ICASSP*, pages 327–332. IEEE, 1976.
- [243] A.V. OPPENHEIM et R.W. SCHAFFER. *Discrete-time signal processing*. Prentice Hall, Englewood Cliffs, 2nd édition, 1999.
- [244] F. ORDUÑA, J.J. LÓPEZ, et A. GONZÁLEZ. Robustness of acoustic crosstalk cancellation as a function of frequency and loudspeaker separation. Dans *109th convention*, n° 5219 dans Perception and psychoacoustics, part one, Los Angeles, septembre 2000. Audio Eng. Soc.
- [245] S. ORFANIDIS. Digital parametric equalizer design with prescribed Nyquist frequency gain. *J. Audio Eng. Soc.*, 45(6):444–455, juin 1997.
- [246] A. OROZCO-SANTILLÁN. Experimental low-frequency sound equalisation in an extended region of an enclosure using adaptive filters. Dans *104^{ème} convention*, n° 4755 dans Acoustics and sound reinforcement, Amsterdam, mai 1998. Audio Eng. Soc.
- [247] N.C. OTTO. Listening tests methods for automotive sound quality. Dans *103rd convention*, n° 4586 dans Perception and subjective tests, New-York, septembre 1997. Audio Eng. Soc.

- [248] T. PAATERO et M. KARJALAINEN. Kautz filters and generalized frequency resolution. Theory and audio applications. Dans *110th convention*, n° 378 dans Signal processing for audio, part 2, Amsterdam, mai 2001. Audio Eng. Soc.
- [249] A. PAPOULIS. *Signal analysis*, chapitre 10. MacGraw-Hill, New-York, 1977.
- [250] E. PARIZET et N. CHOUARD. Sound quality of car radio: the relevance of interaural correlation coefficient. *Acustica*, 83(5):922–925, 1997.
- [251] T.W. PARKS et C.S. BURRUS. *Digital filter design*. Wiley & Sons, New-York, 1987.
- [252] R.D. PATTERSON. Auditory filter shapes derived with noise stimuli. *J. Acoust. Soc. Am.*, 72(3):640–654, 1982.
- [253] R.D. PATTERSON, M. ALLERHAND, et C. GIGUERE. Time-domain modelling of the peripheral auditory processing. *J. Acoust. Soc. Am.*, 98:1890–1894, 1995.
- [254] R.D. PATTERSON, I. NIMMO-SMITH, D.L. WEBER, et R. MILROY. The deterioration of hearing with age : frequency selectivity, the critical ratio, the audiogram, and speech threshold. *J. Acoust. Soc. Am.*, 72(6):1788–1803, 1982.
- [255] C.J. PLACK et B.C. MOORE. Temporal window shape as a function of frequency and level. *J. Acoust. Soc. Am.*, 87(5):2178–2187, mai 1990.
- [256] R. PLOMP et H.J. STEENEKEN. Effect of phase on the timbre of complex tones. *J. Acoust. Soc. Am.*, 46(2, part 2):409–421, 1969.
- [257] J.D. POLACK. *La transmission de l'énergie sonore dans les salles*. Thèse de doctorat, Université du Maine, Le Mans, 1988.
- [258] M.A. POLETTI. The application of linearly swept frequency measurements. *JASA*, 84(2):599–610, août 1988.
- [259] M.A. POLETTI. Linearly swept frequency measurements, time delay spectrometry, and the Wigner distribution. *JAES*, 36(6):457–468, juin 1988.
- [260] F. PONCET. Simulation de localisation de sources sonores dans l'espace. Rapport technique, École Nationale Supérieure des Télécommunications, Département Signal, mars 1992.
- [261] B. PORAT. *Digital processing of random signals : theory and methods*. Prentice Hall, Englewood Cliffs, 1994.
- [262] K. PRECODA et T.H. MENG. Listener differences in audio compression evaluations. *J. Audio Eng. Soc.*, 45(9):708–715, 1997.
- [263] D. PREIS. Phase distortion and phase equalization in audio signal processing - A tutorial review. *J. Audio Eng. Soc.*, 30(11):774–794, novembre 1982.
- [264] W.H. PRESS, B.P. FLANNERY, S.A. TEUKOLSKY, et W.T. VETTERLING. *Numerical recipes in C*. Cambridge university press, Cambridge, 2^{ème} édition, 1993.
- [265] T.A. RAMSTAD. Digital methods for conversion between arbitrary sampling frequencies. *IEEE transactions on acoustics, speech, and signal processing*, 32(3):577–591, juin 1984.
- [266] P.A. REGALIA et S.K. MITRA. Tunable digital frequency response equalization filters. *IEEE transactions on acoustics, speech, and signal processing*, 35(1):118–120, janvier 1987.
- [267] D. REILLY et D. MACGRATH. Convolution processing for realistic reverberation. Dans *98th convention*, n° 3977 dans Audio signal processing, part 1, Paris, février 1995. Audio Eng. Soc.
- [268] D. RIFE. Comments on "Distorsion immunity of MLS-derived impulse response measurements", and author's reply. *JAES*, 42(6):490–497, juin 1994.
- [269] D. RIFE. "Maximum-Length Sequence System Analyzer", *reference manual*, volume 10A. DRA Laboratories, 1994. URL : <http://www.mlssa.com>.
- [270] D. RIFE et J. VANDERKOOY. Transfer function measurement with maximum-length sequences. *J. Audio Eng. Soc.*, 37(6):419–443, juin 1989.

- [271] M. ROBERTS, J. GRIECO, et C. ELLIS. Diffus field radiators in automotive sound system design. Dans *108th convention*, n° 5163 dans Transducers, part two, Paris, février 2000. Audio Eng. Soc.
- [272] E.A. ROBINSON et S. TREITEL. *Geophysical signal analysis*, page 163 et suiv. Prentice Hall, Englewood Cliffs, 1980.
- [273] J. ROSE, P.A. NELSON, T. TAKEUCHI, et B. RAFAELY. Variance of sweet spot size with head location for virtual audio. Dans *110th convention*, n° 5391 dans Psychoacoustic, perception and listening tests, part one, Los Amsterdam, mai 2001. Audio Eng. Soc.
- [274] F. RUMSEY et T. MACCORMICK. *Son et enregistrement - Théorie et pratique*. Eyrolles, Paris, 1994. Traduit de l'anglais par Michel Sénéchal.
- [275] T. SALAVA. Acoustic load and transfer functions in rooms at low frequencies. *J. Audio Eng. Soc.*, 36(10):763–774, octobre 1988.
- [276] T. SALAVA. Subwoofers in small listening rooms. Dans *106th convention*, n° 4940 dans Room acoustics, Munich, mars 1999. Audio Eng. Soc.
- [277] G. SAPORTA. *Probabilités, analyse des données et statistique*. Technip, Paris, 1990.
- [278] R. SAUCEDO et E.E. SCHIRING. *Introduction to continuous and discrete control systems*. Mac-Millan, New-York, 1968.
- [279] S.S. SCHIFFMAN, M.L. REYNOLDS, et F.W. YOUNG. *Introduction to multidimensional scaling*. Academic Press, London, 1981.
- [280] G. SCHIFFRER et D. STANZIAL. Energetic properties of acoustic fields. *J. Acoust. Soc. Am.*, 96(6):3645–3653, Décembre 1994.
- [281] M.R. SCHROEDER. Frequency correlations functions of frequency responses in rooms. *J. Acoust. Soc. Am.*, 34(12):1819–1823, décembre 1962.
- [282] M.R. SCHROEDER. New method for measuring reverberation time. *J. Acoust. Soc. Am.*, 37(3):409–412, mars 1965.
- [283] M.R. SCHROEDER. Digital simulation of sound transmission in reverberant spaces. *J. Acoust. Soc. Am.*, 47(2):424–431, février 1970.
- [284] M.R. SCHROEDER. Computer models for concert hall acoustics. *Am. J. Phys.*, 41:461–471, avril 1973.
- [285] M.R. SCHROEDER. Integrated impulse method for measuring decay without using impulses. *JASA*, 66(2):497–500, août 1979.
- [286] M.R. SCHROEDER. Normal frequency and excitation statistics in rooms: model experiments with electric waves. *J. Audio Eng. Soc.*, 35(5):307–316, mai 1987.
- [287] M.R. SCHROEDER. Statistical parameters of the frequency response curves in large rooms. *J. Audio Eng. Soc.*, 35(5):299–306, mai 1987.
- [288] M.R. SCHROEDER, D. GOTTLOB, et K.F. SIEBRASSE. Comparative study of european concert halls. *J. Acoust. Soc. Am.*, 56:1195–1201, 1974.
- [289] M.R. SCHROEDER, D. GOTTLOB, et K.F. SIEBRASSE. Comparative study of European concert halls: correlation of subjective preference with geometric and acoustic parameters. *J. Acoust. Soc. Am.*, 56(4):1195–1201, octobre 1974.
- [290] M.R. SCHROEDER et H. KUTTRUFF. On frequency response curves in rooms. Compararison of experimental, theoretical and monte carlo results for average frequency spacing between maxima. *J. Acoust. Soc. Am.*, 34(1):76–80, janvier 1962.
- [291] G. SCHWARTZ. Estimating the dimensions of a model. *Annals of statistics*, 6:461–464, 1978.
- [292] W.R. SCOTT et J.S. SMITH. Error corrections for an automated time domain network analyser. *IEEE transactions on instrumentation and measurement*, 35(3):300–303, septembre 1986.
- [293] S.M. SIMPSON. Recursive scheme for normal equations of toeplitz form. Contract AF 19 (604) 7378 7, Massachusetts Institute of Technology, Cambridge, 1963. ARPA project.

- [294] M. SLANEY. An efficient implementation of the Patterson-Holdsworth auditory filter bank. Rapport technique 35, Apple Computer, Inc., Advanced Technology Group, 1993. URL : <http://www.slaney.org/malcolm/pubs.html>.
- [295] M. SLANEY. Auditory toolbox: a Matlab toolbox for auditory modeling work. Rapport technique 45, Apple Computer, Inc., Advanced Technology Group, 1998. URL : <http://www.slaney.org/malcolm/pubs.html>.
- [296] J. SMITH et P. GOSSET. A flexible sampling rate conversion method. Dans *ICASSP*, San Diego, mars 1984. IEEE.
- [297] J.O. SMITH. *Techniques for digital filter design and system identification with to the violin*. PhD thesis, Center for Computer Research in Music and Acoustics, Stanford University, Music Department, 1983.
- [298] J.O. SMITH et J. ABEL. The Bark bilinear transform. Dans *ASSP Workshop*, Mohonk, octobre 1995. IEEE.
- [299] J.O. SMITH et J. ABEL. Bark and ERB bilinear transforms. *IEEE transactions on speech and audio processing*, 1999.
- [300] ST250. *User's guide*. SOUNDFIELD, 1996. URL : <http://www.soundfield.com>.
- [301] K. STEIGLITZ et L.E. MACBRIDE. A technique for the identification of linear systems. *IEEE transactions on automatic control*, 10:461–464, octobre 1965.
- [302] G.P. SUCCI. The interior acoustic field of an automobile cabin. *J. Acoust. Soc. Am.*, 81(6):1688–1694, juin 1987.
- [303] P. SVENSSON et J.L. NIELSEN. Errors in MLS measurements caused by time variance in acoustic systems. Dans *100^{ème} convention*, n° 4268 dans Instrumentation and measurement, Copenhagen, mai 1996. AES.
- [304] S. TASAKI, K. ENAMI, et T. YANAGISAWA. Relationship between the physical and the psychoacoustics parameters of the soundfield in automobile. Dans *Soc. Auto. Eng.*, n° 890118 dans Automobile audio systems, 1989.
- [305] G. THEILE. On the standardization of the frequency response of high-quality studio headphones. *J. Audio Eng. Soc.*, 34(12):956–970, 1986.
- [306] W.S. TORGERSON. *Theory and methods of scaling*. Wiley & Sons, New-York, 1958.
- [307] M. TYRIL, J.A. PEDERSEN, et P. RUBAK. Digital filters for low frequency equalization. Dans *106th convention*, n° 4897 dans Signal processing, part 1. Audio Eng. Soc., mai 1999.
- [308] J. VANDERKOOY. Another approach to time-delay spectrometry. *JAES*, 34(7-8):523–538, juillet-août 1986.
- [309] J. VANDERKOOY. Aspects of MLS measuring systems. *JAES*, 42(4):219–231, avril 1994.
- [310] G. VANDERNOOT. Analyse temps-fréquence de réponses impulsionnelles de salles de concert. Rapport de stage du Mastère TDS, École Nationale Supérieure de Télécommunications, IRCAM - Équipe acoustique des salles, octobre 1995.
- [311] G. VANDERNOOT. Filtre numérique et mannequin acoustique. Note de service n° H26/98.264, Renault - Direction de la Recherche, Rueil-Malmaison, avril 1998.
- [312] J. VERSPECHT. Accurate spectral estimation based on measurements with a distorted timebase digitizer. *IEEE transactions on instrumentation and measurement*, 43(2):210–215, avril 1994.
- [313] D.B. WARD et G.W. ELKO. Optimum loudspeaker spacing for robust crosstalk cancellation. Dans *ICASSP*, pages 3541–3545. IEEE, 1998.
- [314] O. WARUSFEL. Étude des paramètres liés à la prise de son pour les applications d'acoustique virtuelle. Dans *1^{er} Congrès Français d'Acoustique*, volume 2, pages 877–880, Lyon, février 1990.
- [315] G. WEINRICH et E.B. ARNOLD. Method for measuring acoustic radiation fields. *J. Acoust. Soc. Am.*, 68(2):404–411, août 1980.

- [316] P.D. WELCH. The use of FFT for the estimation of power spectra: a method based on time averaging over short modified periodograms. Dans *IEEE transactions on audio and electroacoustics*, volume AU-15, pages 70–73, juin 1967.
- [317] E. WENZEL et F. WIGHTMAN. Localization with non individualized virtual acoustics display cues. *8th annual conference*, 1991.
- [318] B. WIDROW et S. STEARNS. *Adaptive signal processing*. Prentice Hall, Englewood Cliffs, 1986.
- [319] L. WIEGREBE et K. KRUMBHOLZ. Temporal resolution and temporal masking properties of transient stimuli: data and an auditory model. *J. Acoust. Soc. Am.*, 105(5):2746–2756, mai 1999.
- [320] M.D. WILDE. The psychacoustical effects of interaural cross-correlation. Master’s thesis, Northwest university, 1989.
- [321] M.D. WILDE, G. KENDALL, et W. MARTENS. Method for controlling the width and distance of an acoustical image. Unpublished patent application assigned to the former Auris Corporation, 1990.
- [322] L. WILKINSON. SYSTAT® 9, volume Statistics II, chapitre 2, pages 47–73. SPSS, Chicago, 1999.
- [323] F.J. WILLIAMS et N.J. SLOANE. Pseudo-random sequences and arrays. Dans *Proceedings*, volume 64(12), pages 1715–1728. IEEE, décembre 1976.
- [324] R. WILSON. Filter topologies. *J. Audio Eng. Soc.*, 41(9):667–678, septembre 1993.
- [325] S. WINSBERG et J.D. CAROLL. A quasi-nonmetric method for multidimensional scaling of multiway data via a restricted case of an extended INDSICAL model. Dans *Multiway data analysis*, pages 405–414. Elsevier, 1989.
- [326] S. WINSBERG et J.D. CAROLL. A quasi-nonmetric method for multidimensional scaling via an extended euclidian model. *Psychometrika*, 54:217–229, 1989.
- [327] S. WINSBERG et J.D. CAROLL. Fitting an extended INDSICAL model to three-way proximity data. *Journal of classification*, 1995.
- [328] S. WINSBERG et G. DE SOETE. A latent class approach to fitting the weighted euclidean model, CLASCAL. *Psychometrika*, 58(2):315–330, 1993.
- [329] T. WULFRANK. Aspects temporels du masquage auditif : expériences et modélisation. Rapport de stage du DEA ATIAM, Université Pierre & Marie CURIE, IRCAM - Équipe perception et cognition musicales, juin 2001.
- [330] N. XIANG et J. BLAUERT. Auditorium acoustics prediction using binaural tenth-scale modeling. Dans *Proceedings*, volume 14(2), pages 47–65, Birmingham, 1992. International congress on acoustics, architecture and auditoria.
- [331] N. ZACHAROV, S. BECH, et D. MEARES. The use of two subwoofers in the context of surround sound reproduction. *J. Audio Eng. Soc.*, 46(4):276–287, 1998.
- [332] E. ZWICKER et R. FELDTKELLER. *L’oreille récepteur d’information*. Masson, Paris, 1981.