

UNIVERSITE DE LA MEDITERRANEE AIX-MARSEILLE II

THESE

pour obtenir le grade de

DOCTEUR DE L'UNIVERSITE AIX-MARSEILLE II

Discipline : Mécanique

Option : ACOUSTIQUE

présentée et soutenue publiquement

par

Mathieu LAVANDIER

le 19 Décembre 2005

Titre :

**DIFFERENCES ENTRE ENCEINTES
ACOUSTIQUES : UNE EVALUATION PHYSIQUE
ET PERCEPTIVE**

Directeurs de thèse :

Philippe HERZOG – Sabine MEUNIER

JURY

Pr. Hugo FASTL	Lehrstuhl für Mensch-Maschine-Kommunikation	Rapporteur
Pr. Jean-Dominique POLACK	Laboratoire d'Acoustique Musicale	Rapporteur
Pr. Étienne PARIZET	Laboratoire Vibrations et Acoustique	Examineur
Pr. Mario ROSSI	Ecole Polytechnique Fédérale de Lausanne	Examineur
Dr. Philippe HERZOG	Laboratoire de Mécanique et d'Acoustique	Co-directeur de thèse
Dr. Sabine MEUNIER	Laboratoire de Mécanique et d'Acoustique	Co-directrice de thèse

UNIVERSITE DE LA MEDITERRANEE AIX-MARSEILLE II

THESE

pour obtenir le grade de

DOCTEUR DE L'UNIVERSITE AIX-MARSEILLE II

Discipline : Mécanique

Option : ACOUSTIQUE

présentée et soutenue publiquement

par

Mathieu LAVANDIER

le 19 Décembre 2005

Titre :

**DIFFERENCES ENTRE ENCEINTES
ACOUSTIQUES : UNE EVALUATION PHYSIQUE
ET PERCEPTIVE**

Directeurs de thèse :

Philippe HERZOG – Sabine MEUNIER

JURY

Pr. Hugo FASTL	Lehrstuhl für Mensch-Maschine-Kommunikation	Rapporteur
Pr. Jean-Dominique POLACK	Laboratoire d'Acoustique Musicale	Rapporteur
Pr. Étienne PARIZET	Laboratoire Vibrations et Acoustique	Examineur
Pr. Mario ROSSI	Ecole Polytechnique Fédérale de Lausanne	Examineur
Dr. Philippe HERZOG	Laboratoire de Mécanique et d'Acoustique	Co-directeur de thèse
Dr. Sabine MEUNIER	Laboratoire de Mécanique et d'Acoustique	Co-directrice de thèse

*Aux Bawlers pour la musique,
et à Jean-Claude pour les enceintes*

Remerciements

Tout d'abord un grand merci à Philippe et Sabine qui m'ont fait confiance pour démarrer ce sujet passionnant. J'ai pris un immense plaisir à travailler avec eux, et ils m'ont beaucoup appris. J'espère que le futur nous réserve de nouveaux G3.

Je remercie chaleureusement l'entreprise Mosquito qui a soutenu ce projet et m'a accueilli à de nombreuses reprises, en me prêtant enceintes et salle d'écoute. Merci à Philippe et Marc pour avoir rendu cela possible, et à Mickael pour son aide durant les séances d'enregistrements.

Je remercie Hugo Fastl, Jean-Dominique Polack, Etienne Parizet et Mario Rossi pour avoir accepté d'évaluer mon travail, et pour leurs commentaires enrichissants.

Merci à Jeremy qui a décidé un jour de venir rédiger sa thèse dans le bureau d'à côté. Il en a profité pour me donner ses programmes et m'apprendre Matlab et la MDS. Pour couronner le tout, il m'a fait découvrir Miles, Coltrane et Charlie... un immense merci pour tout ça.

Merci à Benjamin d'être venu travailler avec moi pendant son stage, ce fut extrêmement sympathique de partager la deuxième voie de l'ampli avec lui.

Je remercie les fabricants d'enceintes, les professionnels de l'audio et les laboratoires qui m'ont prêté leur matériel : BC Acoustique, Cabasse, le Conservatoire National des Arts et Métiers et Manuel, Copper et Cobalt, France Telecom R&D et Greg, Genesis et toute son équipe, le groupe Mosquito, Relief Sonore, Supravox.

Evidemment, un grand merci à tous les auditeurs qui ont passé mes tests.

Je remercie le C.N.R.S. et la Région Provence-Alpes-Côtes d'Azur qui ont financé ma bourse de thèse, ainsi que la S.F.A. et l'Ecole Doctorale qui par leur soutien financier m'ont permis de participer à des conférences à l'étranger.

Je remercie le L.M.A. pour son accueil, et tout particulièrement les équipes SACADS et APIM au sein desquelles il a été très agréable de travailler. Un grand merci à Marie-Madeleine pour avoir fait que les choses soient si simples (même le week-end) et à Michèle qui a su me trouver des références introuvables. Merci à Georges pour avoir toujours patiemment répondu à mes nombreuses questions, à Guy pour les bricolages en tout genre, à Isa pour avoir été un sonimaître infallible, à Flo et Nico pour l'aventure JJCAAS 2005, et à Lionel, seule victime de cette même aventure.

Un grand merci à ma famille pour son soutien. L'enthousiasme l'a toujours emporté face la pointe d'incompréhension suscitée par mes explications. Merci aux tontons qui m'ont accueilli dans leur ville. Mille bisex à Mira qui a rempli ces années de bonheur, et qui n'a presque pas paniqué devant la multiplication des pensées-bête.

Pour finir, merci à mes chaussures de footing qui m'ont permis de bien dormir sans sortir du placard, et à la cantine du C.N.R.S. qui a fini par interdire de prendre deux desserts le midi, limitant ainsi les dégâts.

SOMMAIRE

Glossaire

1	Termes relatifs aux modèles auditifs	xiii
2	Notations mathématiques	xiv

Introduction

Chapitre I

Etat de l'art

I.1	Limitations des mesures physiques normalisées	5
I.2	Nécessité du contrôle des tests d'écoute	6
I.3	Études sur la perception de la reproduction sonore	9
I.3.1	Dimensions perceptives mises en évidence	10
I.3.2	Evaluations globales de la qualité	14
I.4	Tout protocole expérimental constitue un compromis	15
I.5	Recherches de mesures objectives pertinentes	20
I.5.1	Comparaison par corrélation "visuelle"	21
I.5.2	Comparaison par corrélation grandeur perceptive/grandeur objective	23
I.5.3	Bilan des recherches	24
I.6	Le recours aux signaux de synthèse	26
I.7	Bilan des différentes approches	29

Chapitre II

Approche et protocole expérimental choisis

II.1	Evaluation des différences relatives entre enceintes acoustiques	33
II.2	Protocole expérimental	34
II.3	Limitations et avantages du protocole	37
II.4	Dissemblances perceptives	39
II.4.1	Comparaison par paires	40

II.4.2	Classification libre	40
II.4.3	Validité de l'expérience psychoacoustique	41
II.5	Dissemblances objectives "classiques"	44
II.5.1	Synchronisation des signaux enregistrés	45
II.5.2	Techniques d'analyse "classiques" considérées	45
II.6	Utilisation de modèles auditifs: dissemblances "perceptives objectives"	48
II.6.1	Densité de sonie totale: estimation du spectre perçu	51
II.6.2	Densités de sonie: estimations du plan temps-fréquence perçu	52
II.7	Bilan de l'approche proposée	54

Chapitre III

Comparaison directe des dissemblances

III.1	Première série d'expériences: influence de l'extrait musical et de la prise de son	59
III.1.1	Enregistrements	59
III.1.2	Tests d'écoute	60
III.1.3	Corrélation entre dissemblances objectives et perceptives	64
III.2	Deuxième série d'expériences: influence du mode de reproduction des enceintes	66
III.2.1	Enregistrements	67
III.2.2	Tests d'écoute	69
III.2.3	Corrélation entre dissemblances objectives et perceptives	69
III.2.4	Analyse de variance	71
III.3	Troisième expérience: influence de la pièce et des stimuli utilisés	72
III.3.1	Enregistrements	72
III.3.2	Tests d'écoute	74
III.3.3	Corrélation entre dissemblances objectives et perceptives	74
III.4	Bilan de la comparaison directe des dissemblances	75

Chapitre IV

Comparaison multidimensionnelle des dissemblances

IV.1	Influence de l'extrait musical et de la prise de son	81
IV.2	Influence du mode de reproduction des enceintes	86
IV.3	Influence de la pièce	89
IV.4	Conclusions intermédiaires	91
IV.4.1	Un seul espace, deux dimensions	91
IV.4.2	Bilan de la comparaison objectif/perceptif pour ces deux dimensions	92
IV.5	Influence des stimuli utilisés	94
IV.5.1	Influence de la tâche perceptive?	96

IV.5.2 Limitations de l'approche objective pour décrire les quatre dimensions . . .	98
IV.6 Attributs objectifs préliminaires	99
IV.6.1 Balance grave-aigu	100
IV.6.2 Emergence du bas-médium	103
IV.6.3 Bilan sur les attributs objectifs préliminaires	106
IV.7 Bilan de la comparaison multidimensionnelle des dissemblances	107

Chapitre V Perspectives
--

V.1 Analyse du signal audio	111
V.1.1 Autres modèles et autres métriques	111
V.1.2 Méthodes de comparaison objectif/perceptif	112
V.1.3 Intérêt de la recherche d'attributs objectifs	113
V.2 Psychoacoustique	113
V.2.1 Influence de la tâche perceptive	114
V.2.2 Perception de la reproduction sonore	115
V.3 Mesures de réponses impulsionnelles	116
V.4 Enceintes acoustiques	117
V.4.1 Dynamique	117
V.4.2 Interaction avec la salle	118
V.4.3 Lien avec les paramètres physiques	121
V.5 Bilan	121

Conclusion

BIBLIOGRAPHIE

Annexes

Annexe A Matériel utilisé lors des expériences

A.1 Prises de son testées	iii
A.2 Panel d'enceintes considéré	iii
A.3 Matériel d'enregistrement	v
A.4 Photographies des séances d'enregistrements	vi
A.5 Evaluation de la durée de réverbération des salles d'écoute	vii
A.6 Dispositif expérimental lors des tests psychoacoustiques	viii

Annexe B Méthodes d'analyse statistique utilisées

B.1	Corrélation	xiii
B.2	Comparaison de partitions: indice de Rand	xiii
B.3	Analyse en arbre hiérarchique	xiv
B.4	Analyse de variance (ANOVA)	xv
B.5	Analyse multidimensionnelle (MDS)	xvii
B.5.1	Modèle utilisé	xvii
B.5.2	Détermination du nombre de dimensions	xviii
B.5.3	Orientation des axes	xx
B.5.4	Stabilité de l'espace perceptif	xx

Annexe C Données complémentaires

C.1	Analyses en arbre hiérarchique sur les auditeurs	xxv
C.2	Corrélations entre dissemblances objectives et perceptives	xxviii
C.3	Analyses de variance de la deuxième série d'expériences	xxx
C.4	Analyses en arbre hiérarchique sur les enceintes	xxxii
C.5	Espaces objectifs obtenus à partir des techniques d'analyse "classiques"	xxxv
C.6	Compléments sur les attributs objectifs préliminaires	xxxviii

Annexe D Synchronisation de signaux audio

D.1	Méthode basée sur le cepstre complexe	xliii
D.2	Méthode basée sur l'intercorrélacion	xliii
D.3	Minimisation de la distance quadratique dans le domaine temporel	xliv
D.4	Solution choisie	xliv

Annexe E Méthodes d'analyse objective utilisées

E.1	Déterminations du spectre et des pondérations fréquentielles	xlvii
E.2	Détermination de la représentation temps-fréquence	xlviii
E.3	Détermination des densités de sonie	xlviii

GLOSSAIRE

1 Termes relatifs aux modèles auditifs

sonie	intensité sonore perçue
tonie ou hauteur tonale	échelle de fréquence perçue
densité de sonie ou sonie spécifique	densité de sonie suivant l'échelle de tonie ou hauteur tonale
échelle des Barks	modélisation de l'échelle de tonie dans les modèles auditifs de Zwicker <i>et coll.</i>
bandes de Bark ou bandes critiques	subdivisions de l'échelle des Barks (annexe E.3)
densité de sonie totale	densité de sonie selon l'échelle des Barks, évaluée de façon globale à partir du spectre du signal considéré comme stationnaire
densité de sonie 1	densité de sonie selon l'échelle des Barks en fonction du temps, calculée en appliquant le modèle permettant d'obtenir la densité de sonie totale à des portions successives de 100 ms du signal, ces portions étant alors considérées comme stationnaires
densité de sonie 1 moyenne	moyenne arithmétique temporelle de la densité de sonie 1
densité de sonie 2	densité de sonie selon l'échelle des Barks en fonction du temps, calculée toutes les 10 ms à l'aide d'un modèle auditif destiné aux sons non-stationnaires, prenant en compte l'étalement temporel du signal dû au fonctionnement du système auditif
densité de sonie 2 moyenne	moyenne arithmétique temporelle de la densité de sonie 2

2 Notations mathématiques

$\langle x \rangle$	moyenne arithmétique de la variable x
$\langle x \rangle_t$	moyenne arithmétique de x sur l'échelle temporelle
$\langle x \rangle_f$	moyenne arithmétique de x sur l'échelle fréquentielle
$\langle x \rangle_b$	moyenne arithmétique de x sur l'échelle des Barks
$\ c\ $	module de la variable complexe c
$\text{Min}\{a,b\}$	plus petite des valeurs a et b
$\text{Max}\{a,b\}$	plus grande des valeurs a et b

Introduction

Les mesures physiques normalisées utilisées de nos jours pour comparer les enceintes acoustiques ne permettent pas de faire le lien avec ce que les auditeurs perçoivent lorsqu'ils utilisent ces enceintes. Elles fournissent de l'information, mais quelle partie de cette information est véritablement pertinente vis-à-vis de notre perception de la reproduction sonore ? Ces mesures normalisées sont exclusivement basées sur une approche physique, mettant en jeu certains paramètres du système de reproduction tels que sa réponse en fréquence et sa directivité. L'aspect perceptif n'est absolument pas pris en compte au cours de ce processus, alors qu'il caractérise pourtant de manière essentielle l'écoute qui peut être faite du champ acoustique rayonné par une source sonore. Les évaluations perceptives sont laissées aux seuls jugements des journalistes de la presse spécialisée, l'évaluation étant alors davantage subjective que perceptive. Ainsi dépourvus de critères objectifs permettant d'évaluer pleinement leur travail, les fabricants sont freinés quant aux améliorations qu'ils pourraient apporter à leurs systèmes.

En vue d'une discrimination objective plus pertinente des enceintes acoustiques, nous avons cherché à évaluer un panel d'enceintes selon deux approches, physique et perceptive, menées en parallèle. L'objectif du travail présenté dans ce document était d'établir des correspondances entre ces deux types d'approche. Le lancement de cette thématique de recherche au sein du laboratoire a donc débuté par la mise en place d'un protocole expérimental approprié. Ce projet de recherche devra permettre à terme de relier la variation d'un paramètre physique de l'enceinte à son influence au niveau de l'écoute, par l'intermédiaire des mesures physiques effectuées. Notre travail se concentrant dans un premier temps sur le rapprochement des approches physique et perceptive, nous n'avons pas tenté d'apporter de nouvelles connaissances concernant la perception de la reproduction sonore ou les mesures physiques d'enceintes acoustiques considérées individuellement. Notre étude a porté sur le lien entre ces deux domaines de recherche au départ disjoints.

Ce projet s'inscrit dans le cadre général de la qualité sonore [1][2]. Il faut comprendre ici le terme "qualité" au sens "caractéristique" plutôt qu'au sens "bon ou mauvais". Nous nous sommes en effet intéressés à l'évaluation des différences relatives entre enceintes acoustiques et non pas à celle de leur qualité absolue. La qualité sonore et la qualité de la reproduction sonore

différent en ce sens que la qualité sonore étudie les caractéristiques du son tandis que la qualité de la reproduction sonore étudie les caractéristiques du système qui restitue ce son. L'évaluation de ces caractéristiques pose le problème de sa dépendance vis-à-vis du signal qui doit nécessairement être utilisé, un système de reproduction ne pouvant pas être évalué sans signal à reproduire.

Le terme "mesure" sera souvent employé dans ce document. Il ne désignera pourtant pas toujours une mesure au sens strict de la comparaison à un étalon. Nous avons parfois décidé de conserver ce terme par souci de clarté du propos, mais le terme plus approprié d'"évaluation" lui sera souvent préféré.

Le chapitre I contient un état de l'art des recherches concernant la perception de la reproduction sonore et les mesures physiques proposées pour l'évaluer. Notre approche pour aborder le problème de la confrontation des mesures physiques et perceptives sur les enceintes acoustiques est présentée dans le chapitre II. Le protocole expérimental qui a été mis au point pour réaliser cette confrontation est décrit. L'utilisation d'un modèle auditif en vue d'une analyse objective plus pertinente de la reproduction sonore est notamment proposée. Deux axes sont ensuite développés en parallèle au cours des deux chapitres suivants : la comparaison objectif/perceptif et l'influence des différents paramètres expérimentaux. Le chapitre III décrit nos expériences, ainsi que la comparaison directe des résultats provenant des évaluations objectives et perceptives. Le chapitre IV présente une comparaison tenant compte du caractère multidimensionnel de la perception de la reproduction sonore. Les avantages et les limitations de l'analyse objective proposée sont détaillés. Nos résultats sont comparés à ceux de la littérature au fur et à mesure de leur présentation. Nos travaux ont également soulevé un certain nombre de questions. Le chapitre V contient des propositions d'expériences visant à y répondre. Il évoque également certaines études que nous envisageons de réaliser grâce à notre protocole expérimental. La conclusion de ce document présente de manière synthétique les apports de notre travail de recherche, tandis que certaines données complémentaires sont fournies en annexe.

CHAPITRE I

Etat de l'art

----- oOo -----

Un état de l'art des recherches concernant la perception de la reproduction sonore et les mesures physiques proposées pour l'évaluer a été réalisé. Après avoir montré les limitations des mesures normalisées et la nécessité du contrôle rigoureux à apporter aux tests d'écoute, ce chapitre présente les différentes approches envisagées pour relier les mesures physiques et perceptives, les étapes de ce rapprochement et les résultats finalement obtenus.

----- oOo -----

SOMMAIRE

I.1	Limitations des mesures physiques normalisées	5
I.2	Nécessité du contrôle des tests d'écoute	6
I.3	Études sur la perception de la reproduction sonore	9
I.3.1	Dimensions perceptives mises en évidence	10
I.3.2	Evaluations globales de la qualité	14
I.4	Tout protocole expérimental constitue un compromis	15
I.5	Recherches de mesures objectives pertinentes	20
I.5.1	Comparaison par corrélation "visuelle"	21
I.5.2	Comparaison par corrélation grandeur perceptive/grandeur objective	23
I.5.3	Bilan des recherches	24
I.6	Le recours aux signaux de synthèse	26
I.7	Bilan des différentes approches	29

I.1 Limitations des mesures physiques normalisées

Les normes [3][4] présentent l'ensemble des mesures qui doivent être effectuées sur des enceintes acoustiques pour les évaluer objectivement. Elles impliquent notamment des mesures de réponse en fréquence, de directivité et de distorsion harmonique. Ces mesures sont réalisées en champ libre dans une salle anéchoïque, ou en champ diffus dans une salle réverbérante. La résolution fréquentielle recommandée est le tiers d'octave. Les normes [5][6] indiquent les performances minimales que doivent satisfaire les enceintes comme gage de leur qualité. Malheureusement, les limitations de ces mesures normalisées sont précisées dès l'introduction des normes : "Les caractéristiques choisies, qui sont représentatives des qualités physiques des enceintes acoustiques, ne tiennent compte ni de certains facteurs extérieurs à l'enceinte acoustique tel que le couplage de l'appareil au local d'écoute, ni des facteurs physiologiques et psychoacoustiques de l'audition" [6]. "Les techniques actuellement utilisées pour les mesures objectives ne permettent pas une définition complète de la qualité de restitution sonore d'un haut-parleur ; néanmoins, les mesures objectives peuvent donner une information essentielle sur la qualité des haut-parleurs" [5]. Les mesures normalisées fournissent en effet de l'information, mais quelle partie de cette information est véritablement pertinente vis-à-vis de notre perception de la reproduction sonore ? Comment analyser cette information pour retrouver les différences perçues par les auditeurs ?

Les mesures physiques normalisées utilisées pour comparer les enceintes acoustiques ne permettent pas actuellement de faire le lien avec ce que les auditeurs perçoivent lorsqu'ils utilisent ces enceintes, "(...) such data are not enough to realize how the reproduced sound will actually be perceived. The reason for this is that we know too little about the psychological relations between the physical parameters and the perceived sound quality" Gabrielsson et Sjögren [7]. Toole [8] fait un bilan des différentes méthodes utilisées pour mesurer la réponse en fréquence des enceintes. Il montre qu'il n'existe pas de consensus sur la manière dont cette mesure doit être effectuée pour rendre compte de la qualité du système. Cette question est toujours d'actualité comme l'écrit Bramsløw [9] en 2004 : "There are many standardized, and relevant measurements that are used for assessment of the "quality" of a device, e.g. frequency response, distortion, signal/noise ratio. They give some indication of the performance of the device, but often little knowledge about the sound quality perceived by the listener".

Les mesures normalisées sont effectuées en salle anéchoïque ou en salle réverbérante. Elles présentent l'avantage d'être reproductibles et comparables d'un laboratoire à l'autre. Elles sont également caractéristiques des enceintes seules, puisque l'influence de la pièce a été minimisée. Ces avantages sont reconnus par Klippel [10], mais il souligne également que de ce fait les conditions d'écoute réelles et les mécanismes psychoacoustiques intervenant dans la perception des auditeurs ne sont pas pris en compte. "Considered from the standpoint of perception psychology, it would be preferable to make measurements in the actual listening room, since this would give a better

description of the stimulus reaching the listener's ears. On the other hand, such measurements are not standardized and present many problems..." Gabrielsson *et coll.* [11].

Afin de rechercher des méthodes de mesures objectives qui soient pertinentes vis-à-vis de la perception de la reproduction sonore, il faut être capable d'évaluer correctement cette perception. Les mesures objectives doivent être reliées à des mesures perceptives qui soient également pertinentes. Pour cela, de nombreux paramètres expérimentaux sont à contrôler pour que les tests perceptifs soient exploitables.

I.2 Nécessité du contrôle des tests d'écoute

Si les tests d'écoute mis en place pour évaluer la perception de la reproduction sonore par des enceintes acoustiques ne sont pas rigoureusement contrôlés, les résultats qu'ils fournissent ne sont pas exploitables, car rien ne permet de dire si ces résultats sont caractéristiques de la reproduction sonore des enceintes ou de tout autre facteur expérimental ayant eu une influence lors des tests. Tous les paramètres de l'expérience doivent être contrôlés et la seule variable restante doit être celle qui fait l'objet de l'étude. Le contrôle des tests perceptifs a donc pour but de s'assurer de ce qui est évalué lors de l'expérience.

Lipshitz et Vanderkooy dans [12] et Toole dans [13][14] détaillent les facteurs expérimentaux à considérer et les précautions à prendre pour réaliser des tests d'écoute sur des enceintes acoustiques. La salle d'écoute, la position des enceintes et de l'auditeur, le mode de reproduction des enceintes, le niveau sonore, l'expérience de l'auditeur, les extraits musicaux utilisés et la procédure expérimentale employée pour réaliser le test ont tous une influence sur les résultats obtenus. La suite de ce paragraphe présente les effets recensés dans la littérature concernant ces différents paramètres expérimentaux.

Salle d'écoute et positions des enceintes Bech [15] a montré que la perception du son rayonné par une enceinte acoustique est influencée par la salle d'écoute. La position de l'enceinte apparaît également cruciale. Il a réalisé ses tests perceptifs en considérant trois positions pour chaque enceinte utilisée en reproduction monophonique. Les tests ont été répétés dans trois pièces différentes. Quatre modèles d'enceintes étaient impliquées. Ses résultats suggèrent que les auditeurs peuvent maintenir leurs impressions d'une pièce à l'autre pourvu qu'ils soient suffisamment entraînés. La salle d'écoute et la position de l'enceinte ont un effet sur l'évaluation absolue des enceintes, mais aussi sur leurs différences relatives.

Une étude parallèle menée par Olive *et coll.* [16] a confirmé l'influence de la position de l'enceinte. Deux modèles d'enceintes ont été comparés en reproduction monophonique, pour quatre positions différentes, dans une même salle d'écoute. Ce test d'écoute en direct était complété par un test d'écoute au casque utilisant des enregistrements binauraux impliquant les mêmes

enceintes et les mêmes positions. Ce test au casque a permis une comparaison immédiate de deux enceintes à la même position ou d'une même enceinte à deux positions différentes. Les deux expériences ont fourni des résultats concordants, montrant que la différence de position des enceintes pouvait avoir plus d'effet que leur différence de modèle. Des tests d'écoute en direct où différentes enceintes sont comparées à différentes positions peuvent donc être biaisés par cette différence de position [16].

L'influence de la pièce et de la position des enceintes semble encore plus critique dans le cas d'une reproduction stéréophonique. Ainsi Toole [14] a dû abandonner des essais visant à comparer des paires d'enceintes l'une à côté de l'autre, car les différences créées par le déplacement de l'image sonore semblaient parfois plus grandes que les différences réelles entre les enceintes.

Mode de reproduction des enceintes Dans la plupart des tests d'écoute rencontrés dans la littérature, les enceintes sont utilisées soit en reproduction monophonique, soit en reproduction stéréophonique. Les deux modes de reproduction ont rarement été impliqués dans la même expérience. Ceci a limité l'étude directe de l'influence de ce paramètre expérimental.

Toole [14][17] a globalement obtenu les mêmes évaluations de qualité pour des enceintes comparées en reproduction monophonique ou stéréophonique, la reproduction stéréophonique apparaissant cependant moins discriminante et plus flatteuse. Les enceintes de mauvaise qualité étaient notamment jugées plus proches des autres modèles lors de l'écoute stéréophonique. Il indique également une plus grande variabilité des résultats individuels des auditeurs pour le test utilisant la reproduction stéréophonique. Dans un premier temps, il considère que la stéréophonie masque les défauts et complique la tâche des auditeurs, mais il revient plus tard sur ses propos : "(...) loudspeakers that rated poorly in monophonic listening tests frequently received higher "fidelity" ratings in stereo listening. It now appears that this change is indicative of the high perceptual priority of spatial presentation rather than a reduced ability to detect resonances in a spatially enriched sound field" [18].

Niveau sonore Gabrielsson *et coll.* [19] ou Eisler [20] ont démontré que la sonie, le niveau sonore perçu, apparaissait comme un des critères importants utilisés par les auditeurs pour différencier des enceintes acoustiques dont les niveaux de reproduction n'avaient pas été égalisés. Illényi et Korpássy [21] ont trouvé une forte corrélation entre les évaluations de qualité et le niveau de reproduction de leurs enceintes lors de leurs tests perceptifs.

L'influence d'une différence de sonie doit être prise en compte pour réaliser des tests d'écoute sur des enceintes acoustiques [22][23], comme pour toute autre expérience de qualité sonore [24]. Ces différences de sonie présentent peu d'intérêt concernant l'évaluation de la reproduction sonore. Afin d'éviter qu'elles ne masquent des différences plus fines, le niveau sonore des reproductions doit être égalisé avant les tests d'écoute. Étant donnée la difficulté d'évaluer précisément

par la mesure la sonie de sons non-stationnaires [25] aux contenus fréquentiels pouvant être différents [14], l'égalisation précise finale est souvent réalisée à l'oreille par les expérimentateurs [14][22].

Concernant le niveau sonore absolu de la reproduction, Staffeldt [26] a réalisé des tests à deux niveaux différents et n'a pas trouvé d'effet lié à ce paramètre. Pourtant, Gabriellson *et coll.* [11] rappellent que le niveau sonore absolu a une influence sur les différentes dimensions perceptives mises en jeu lors de leurs tests. Une enceinte n'est pas un système linéaire, et ne se comporte pas de la même façon suivant le niveau de sollicitation auquel elle est soumise. Des reproductions à des niveaux sonores différents nécessiteront des niveaux de sollicitation différents, qui entraîneront des changements du comportement de l'enceinte. Ces changements seront plus ou moins perceptibles suivant l'amplitude des variations de niveau considérées. Lors d'un test d'écoute, il faut garder à l'esprit que c'est l'enceinte à un niveau de sollicitation donné qui est évaluée.

Expérience de l'auditeur Gabriellson et Lindstrom [22], Bech [27] puis Olive [28] ont montré que les auditeurs entraînés étaient plus stables et plus discriminants dans leurs jugements que les auditeurs n'ayant pas subi d'entraînement, tout en conduisant à des résultats moyens équivalents. Le recours à de tels auditeurs permet donc de réduire le nombre de tests en impliquant moins d'auditeurs, pour une même variance des résultats. Il est important de noter qu'il est ici question d'auditeurs spécialement entraînés pour l'évaluation de la reproduction sonore [28] ou pour le test particulier de l'expérience [27].

Extrait musical L'extrait musical utilisé lors des tests perceptifs affecte bien évidemment l'évaluation des enceintes acoustiques. Cette influence a été révélée par toutes les études faisant intervenir plusieurs extraits musicaux que nous avons rencontrés dans la littérature [7][12][13][14][15][16][19][20][22][26][28]. Il n'est pas étonnant qu'une enceinte se comporte différemment lorsqu'elle est sollicitée par des signaux différents. Ces différents signaux peuvent notamment révéler des caractéristiques différentes de la reproduction sonore.

Procédure expérimentale "The psychophysical procedures used in the experiments can influence the results of sound evaluations considerably" Fastl [29]. La procédure expérimentale employée pour les tests d'écoute doit donc minimiser tout biais pouvant corrompre les résultats du test. L'expérience doit être menée en double-aveugle [12], l'auditeur ne doit bien sûr pas voir les enceintes ni savoir quelles enceintes il compare, mais l'expérimentateur ne doit pas non plus contrôler l'ordre de présentation des différents modèles. Pour éviter tout effet d'ordre, les enceintes doivent être présentées aléatoirement [23].

Lipshitz et Vanderkooy [12] indiquent également qu'il est important d'utiliser des tests les plus

simples possible. Par conséquent, en l'absence de stimulus de référence, ils conseillent une tâche basée sur des évaluations de différence relatives plutôt qu'une tâche impliquant des jugements absolus.

Bilan des recommandations à suivre Lors d'un test d'écoute sur des enceintes acoustiques, c'est l'ensemble enceinte-salle qui est évalué, pour une position déterminée de l'enceinte et de l'auditeur, avec un extrait musical et un niveau sonore donnés. Tous ces facteurs influencent la perception du champ sonore rayonné par les enceintes. Ils doivent donc être rigoureusement contrôlés pour que les résultats des tests d'écoute soient exploitables. Les études portant sur la perception de la reproduction sonore ont mené à l'élaboration de normes ou recommandations [30][31] récapitulant l'ensemble des précautions à prendre pour réaliser des tests perceptifs valides.

La comparaison de différentes enceintes doit se faire dans une même pièce, pour des positions de l'enceinte et de l'auditeur identiques, avec un même extrait musical et au même niveau sonore. Une évaluation globale des enceintes devra faire intervenir plusieurs extraits musicaux de style différents, plusieurs positions d'enceinte et les reproductions monophonique et stéréophonique devront être considérées.

Il convient de remarquer qu'il est nécessaire d'effectuer les tests perceptifs dans une pièce correspondant à un salon domestique "moyen" [30][31], alors que les mesures objectives normalisées doivent être réalisées en chambre anéchoïque ou en salle réverbérante (paragraphe I.1, page 5). L'environnement d'un salon domestique tombe malheureusement entre ces deux cas extrêmes. Cette différence d'environnement ne favorise pas le rapprochement des deux types de mesure.

Avant de s'intéresser aux mesures objectives qui ont été développées pour tenter ce rapprochement, il convient de considérer les résultats qui ont pu être obtenus à partir de tests perceptifs sur des enceintes acoustiques.

I.3 Études sur la perception de la reproduction sonore

Les normes ou recommandations évoquées précédemment concernant les tests d'écoute sur les enceintes acoustiques [30][31] conseillent d'évaluer trois grandes catégories de caractéristiques de la reproduction sonore : la restitution du timbre, la composante spatiale de la reproduction et la dynamique des enceintes, c'est-à-dire l'évolution de leur comportement en fonction du niveau auquel elles sont sollicitées. A notre connaissance, aucune étude approfondie n'a été menée sur cette dernière caractéristique, si ce n'est les quelques expériences évoquées précédemment ayant testé plusieurs niveaux absolus de la reproduction lors des tests d'écoute. La plupart des études réalisées dans les années 70 et 80 ont porté sur la "sound quality" [32] des enceintes, considérant essentiellement la restitution du timbre. Ces recherches ont par contre laissé place ces dernières

années aux recherches portant sur la "spatial quality" [32] de la reproduction sonore, en parallèle du développement des systèmes de restitution multicanaux. Comme il sera expliqué dans la suite du document, nous avons rapidement décidé de limiter notre recherche, au moins dans un premier temps, à la restitution du timbre par les enceintes acoustiques. Nous ne détaillerons donc pas les recherches concernant la composante spatiale de la reproduction sonore. Le lecteur pourra pour cela se référer par exemple aux travaux de Rumsey [33][34], Bech [35][36], Zielinski *et coll.* [37] ou Griesinger [38].

Les études sur la perception de la reproduction sonore ont tout d'abord cherché à déterminer les dimensions perceptives utilisées par les auditeurs pour évaluer les enceintes acoustiques lors des tests d'écoute. Ces études se sont par la suite davantage intéressées à des évaluations plus globales de la qualité des systèmes de reproduction. La finalité de toutes ces recherches reste bien sûr de relier les dimensions perceptives, la qualité globale et les mesures objectives.

I.3.1 Dimensions perceptives mises en évidence

"It may be assumed that perceived sound quality is constituted by a (limited) number of separate perceptual dimensions, and that it would be possible to give a perceptual description of sound-reproducing systems by stating their positions in such dimensions" Gabrielsson et Sjögren [7]. Toutes les recherches ayant pour but de révéler les dimensions ou critères perceptifs utilisés lors de l'évaluation des enceintes acoustiques partent de cette hypothèse. Ces recherches ont principalement impliqué deux techniques d'analyse statistique pour identifier les dimensions à partir des jugements des auditeurs. L'analyse multidimensionnelle de proximité vise à représenter des dissemblances estimées entre stimuli par des distances entre ces mêmes stimuli symbolisés par des points dans un espace multidimensionnel (annexe B.5). Cet espace permet d'explorer le nombre et la nature des dimensions perceptives sous-jacentes aux estimations de dissemblances provenant des tests d'écoute [39]. Des évaluations selon un différentiel sémantique [40], associées à une analyse factorielle [41], peuvent également être utilisées pour explorer ces dimensions. L'évaluation est alors absolue et est effectuée sur un grand nombre d'attributs potentiels de la reproduction sonore. L'analyse factorielle est basée sur l'étude des corrélations entre ces attributs et permet de regrouper les attributs redondants, d'éliminer les attributs non pertinents, afin d'obtenir finalement un nombre plus ou moins restreint de facteurs correspondant aux critères utilisés par les auditeurs pour effectuer leurs jugements. L'analyse factorielle peut aussi être associée à des évaluations globales telles que les estimations de qualité ou de fidélité. La matrice de résultats peut par exemple contenir les évaluations de chaque enceinte en fonction des différents extraits musicaux utilisés. L'analyse factorielle appliquée à cette matrice a alors pour but de révéler les facteurs étant intervenus dans les jugements des auditeurs, les poids apportés à ces facteurs en fonction des différents extraits musicaux, et les "scores" des enceintes selon ces

mêmes facteurs.

Eisler [20] a montré l'intérêt d'utiliser une telle méthode pour révéler les dimensions sous-jacentes aux jugements de qualité effectués par ses auditeurs sur dix enceintes reproduisant vingt-quatre extraits musicaux différents. Il a ainsi mis en évidence neuf facteurs et en a interprété sept : "sound level, purity of transients, environmental information, bass boost, full-treble reproduction, high-treble relative midrange, disturbing directional effects". Eisler précise que ces attributs sont à considérer avec beaucoup de précaution, ses tests perceptifs n'ayant pas été rigoureusement contrôlés.

Lors de tests de comparaison par paires portant sur différentes configurations d'un système de reproduction multicanal utilisant de une à huit enceintes acoustiques, Nakayama *et coll.* [42] ont mis en évidence trois dimensions perceptives : "fullness", "clearness" et "depth of the image sources". Les tests ayant été menés en chambre anéchoïque plutôt qu'en salle d'écoute, ces résultats apparaissent cependant difficilement généralisables.

Staffeldt [26] a réalisé des tests de préférence sur cinq paires d'enceintes acoustiques évaluées deux par deux, donc à des positions différentes. Les auditeurs devaient indiquer la paire qu'ils préféreraient mais également quelle reproduction était par exemple la plus "nuanced, brilliant, full-bodied, strong midrange and shrill", avec au total trente-cinq attributs à évaluer. La différence entre deux paires d'enceintes a ensuite été estimée en calculant la fréquence des préférences de chaque paire. L'analyse de ses résultats a révélé deux dimensions principales : "emphasized treble" et "emphasized bass".

Dans [19], Gabriellsson *et coll.* présentent deux expériences distinctes. La première faisait intervenir cinq enceintes situées à des positions différentes, tandis que la deuxième n'impliquait qu'une seule enceinte dont la réponse en fréquence et le niveau de reproduction étaient variables, selon six combinaisons possibles. Les auditeurs devaient fournir des jugements absolus de fidélité des enceintes, ainsi que des jugements relatifs basés sur une évaluation de la similarité des systèmes considérés. Des analyses multidimensionnelle et factorielle ont ensuite été appliquées sur les résultats de ces tests. En plus de la sonie ("loudness"), le niveau sonore perçu, ces analyses ont révélées deux dimensions perceptives : "distinctness" ("clarity", "clearness") qui serait reliée à la bande passante et au niveau de distorsion de l'enceinte, et "brightness-darkness" semblant dépendante de la forme de la réponse en fréquence et de l'équilibre grave-aigu. Ces dimensions avaient les mêmes interprétations pour les trois extraits musicaux employés dans les tests d'écoute.

Gabriellsson et Sjögren [7] ont complété cette étude en réalisant des évaluations absolues sur cinquante-cinq échelles différentes correspondant à des attributs potentiels de la reproduction sonore. Seulement trois enceintes étaient impliquées, mais l'une d'elle pouvait subir des modifications de réponse en fréquence au niveau des aigus ou des graves ainsi qu'un ajout de distorsion. Au total neuf systèmes étaient évalués à trois positions différentes, à l'aide de plusieurs extraits musicaux. Le résultat de ces évaluations a été soumis à une analyse fac-

torielle dans le but de faire apparaître les attributs pertinents. Quatre facteurs ressortent de cette analyse : 1) "clearness/distinctness", "feeling of space", "nearness" ; 2) "sharpness/hardness-softness" ; 3) "brightness-darkness" ; 4) "disturbing sounds". Le même type d'expérience a été mené en parallèle sur des casques et des prothèses auditives. Les tests sur les prothèses impliquaient en plus un test de comparaison par paires associés à une analyse multidimensionnelle. Dans la plupart des tests, le son produit par les prothèses était enregistré pour être reproduit au casque pendant l'expérience. Les tests sur les casques et les prothèses auditives ont conduit à des attributs perceptifs du même type que ceux mis en évidence avec les enceintes acoustiques, les facteurs résultant des analyses factorielles pouvant combiner ces attributs plus ou moins différemment selon les expériences. Gabrielsson et Sjögren [7] récapitulent ainsi les huit dimensions perceptives révélées par l'ensemble de leurs tests : "clearness/distinctness", "sharpness/hardness-softness", "brightness-darkness", "fullness-thinness", "feeling of space", "nearness", "disturbing sounds" et "loudness". Ils font remarquer que ces dimensions peuvent être redondantes et qu'il peut également en exister d'autres. Ils précisent que leur nom est provisoire et qu'il est susceptible d'être modifié par de nouvelles recherches.

Gabrielsson et Lindstrom [22] ont ensuite testé la validité de ces dimensions perceptives au cours de trois expériences réalisées en reproduction stéréophonique. Les auditeurs devaient évaluer les différentes reproductions sonores selon ces dimensions et selon deux dimensions globales de fidélité et d'agrément. Les deux premières expériences impliquaient de cinq à six paires d'enceintes par test. La troisième expérience a fait intervenir vingt paires d'enceintes, au cours de seulement quatre séances, les paires d'enceintes ne pouvant donc pas être évaluées toutes à la même position en même temps. Ces expériences ont confirmé la validité des dimensions perceptives révélées précédemment et leur pertinence vis-à-vis des évaluations globales. Elles ont également soulevé la question importante de l'influence du contexte lors des tests : l'évaluation d'une enceinte pouvant être dépendante des autres enceintes impliquées dans le même test.

Dans [10], Klippel décrit succinctement les tests perceptifs sur lesquels il a basé ses recherches. Ils ont été effectués dans trois environnements différents : une salle d'écoute, un studio et une chambre anéchoïque. Aucune précision n'est malheureusement donnée concernant l'influence de ces environnements sur les résultats. Les tests impliquaient de quatre à huit enceintes acoustiques, en reproduction monophonique ou stéréophonique, avec de nombreux extraits musicaux. Ils ont tous été basés sur des comparaisons par paires, avec des enceintes devant donc être placées à des positions différentes. Les auditeurs ont réalisé soit des évaluations de similarité qui étaient ensuite associées à une analyse multidimensionnelle, soit des évaluations de rapports selon de multiples attributs, en indiquant par exemple qu'une enceinte était deux fois plus "clear" ou "sharp" qu'une autre. Ces évaluations de rapports ont ensuite été traitées par une analyse factorielle. Deux à trois facteurs ont été mis en évidence par test, les dimensions finalement obtenues sur l'ensemble des expériences étant : "clearness", "treble stressing" ("sharpness"), "general bass emphasis"

("volume"), "low bass emphasis", "feeling of space", "clearness of basses", "brightness".

Dans le cadre de l'évaluation de la qualité des prothèses auditives, Bramsløw [9] a réalisé des tests d'écoute au casque, en utilisant deux extraits ayant subi différentes transformations dans les basses, moyennes et hautes fréquences. Les auditeurs devaient évaluer les stimuli selon six échelles choisies à partir des travaux de Gabrielsson *et coll.* [7][19][22]: "loudness", "clearness", "sharpness", "fullness", "spaciousness", "overall impression". Les quatre facteurs révélés par l'analyse factorielle de ses résultats ont été : 1) "clearness", "overall impression" ; 2) "sharpness" vs. "fullness" ; 3) "spaciousness", "loudness" ; 4) "loudness", "spaciousness". La sonie apparaît comme un critère du fait d'une égalisation seulement approximative des niveaux de reproduction. Bramsløw conclue que les échelles "clearness" et "sharpness" sont les plus pertinentes pour évaluer la qualité sonore de ses signaux. Il a également constaté que ces échelles n'étaient pas absolue mais relatives. Elles permettent de classer les reproductions sonores les unes par rapport aux autres.

Finalement, en plus de la sonie qui intervient lorsque les niveaux de reproduction sonore des enceintes n'ont pas été égalisés avant les tests, les dimensions perceptives mises en évidence par toutes ces études peuvent être présentées de manière synthétique dans la liste suivante :

- "clearness/distinctness",
- "sharpness/hardness-softness" et "fullness-thinness",
- "brightness-darkness",
- "feeling of space" et "spaciousness",
- "nearness",
- "bass emphasis" ou "treble emphasis",
- "disturbing sounds".

Connaître les critères utilisés par les auditeurs pour différencier les enceintes lors des tests d'écoute peut aider l'expérimentateur à relier les évaluations perceptives aux mesures objectives. Avoir une idée du type d'information qui doit être recherché dans le signal de mesure peut en effet constituer un atout précieux.

Après avoir étudié les dimensions perceptives associées à la reproduction sonore, les études ont tenté d'évaluer de façon globale la qualité des enceintes. Ces évaluations ont parfois été associées à des évaluations selon différents attributs perceptifs, principalement ceux révélés par les études de Gabrielsson *et coll.* [7][19][22], afin de trouver le lien entre dimensions perceptives et qualité globale.

I.3.2 Evaluations globales de la qualité

Les évaluations globales de la qualité sonore liée à la restitution du timbre font intervenir différents termes selon les études : qualité sonore globale [43], impression globale [9], fidélité [14][15][19][22][27][44], fidélité de reproduction du timbre [15][27], précision [45][46], caractère naturel de la reproduction ("true to nature", "naturalness") [7][10][19][47], préférence [16][26][28][46] ou agrément [7][10][14][22]. Olive *et coll.* [16][45] ont trouvé les jugements de préférence plus discriminants que les jugements de fidélité lorsque le test d'écoute implique des enceintes qui se ressemblent beaucoup. Tous ces termes font référence à une évaluation globale qui est censée prendre en compte, consciemment ou inconsciemment de la part de l'auditeur, l'ensemble des dimensions perceptives évoquées au paragraphe précédent.

Gabrielsson et Lindstrom [22] ont recherché le lien entre les échelles globales de fidélité et d'agrément et les dimensions perceptives provenant de leurs précédentes expériences [7][19]. Ils ont calculé les coefficients de corrélation entre dimensions perceptives et échelles globales, afin d'évaluer l'importance respective des différentes dimensions perceptives vis-à-vis de la qualité. Les dimensions "clarity" et "spaciousness" ont fourni les corrélations les plus élevées, suivies par la dimension "fullness". Les échelles de fidélité et d'agrément se sont avérées très corrélées.

Toole [14] a utilisé ces mêmes dimensions perceptives et échelles globales lors de plusieurs tests d'écoute. Jusqu'à seize enceintes ont été testées en reproduction monophonique, utilisant quatre positions différentes. Des tests en reproduction stéréophonique ont impliqué quatre paires d'enceintes évaluées à la même position grâce à un système de plateaux tournants. Lors de l'écoute stéréophonique, l'auditeur devait en plus juger différents attributs portant sur la composante spatiale de la reproduction [14]. Ces tests de Toole semblent révéler une bonne corrélation entre les jugements de "sound quality", censés être basés sur les caractéristiques plutôt timbrales de la reproduction sonore, et les jugements de "spatial quality" devant plutôt évaluer la dimension spatiale de la reproduction. La différence entre ces deux échelles n'a pas été significative et il conclut en ces termes: "Accurate sound reproduction and good spatial representations appear to go hand-in-hand; and good loudspeaker, used in pairs, becomes a good stereo loudspeaker. For critical loudspeaker evaluations it is probably important to examine the performance in both stereophonic and monophonic test" [14]. Il est important de signaler que ces tests d'écoute impliquaient au maximum quatre paires d'enceintes, du fait de la difficulté expérimentale pour les comparer rapidement à la même position. Ce nombre limité de stimuli peut aussi expliquer le fait que la composante spatiale de la reproduction sonore ne soit pas apparue de façon indépendante.

De la même manière, Staffeldt [26][43] considère que les différences de timbre interviennent de manière prépondérante dans les jugements de qualité globale des systèmes, que l'information stéréophonique est mineure et qu'elle n'influence pas la préférence des auditeurs. Malheureusement ses tests d'écoute n'ont peut être pas donné suffisamment accès à cette information stéréopho-

nique, avec seulement cinq paires d'enceintes comparées deux-à-deux à des positions différentes.

Olive *et coll.* [16][28][46] réalisent des tests de préférence, les auditeurs évaluant les différentes reproductions sonores sur une échelle comprise entre "n'aime vraiment pas" et "aime vraiment". Dans [46], les auditeurs qui étaient entraînés devaient en plus rendre compte de la balance spectrale telle qu'ils la percevaient, en évaluant le niveau perçu dans différentes bandes de fréquences adjacentes, et estimer la quantité de distorsion non-linéaire présente dans la reproduction. Il leur était également demandé de commenter librement les caractéristiques qu'ils considéraient importantes pour juger de la qualité sonore des reproductions qu'ils avaient écouté. Ce test perceptif faisait intervenir treize enceintes en reproduction monophonique. L'utilisation d'un plateau tournant a permis à Olive de placer toutes les enceintes à la même position, en testant quatre enceintes par séance d'écoute. Les résultats de ce test ont montré l'importance de la balance spectrale perçue vis-à-vis de la préférence des auditeurs. La perception des basses fréquences constitue un facteur important, et un excès de medium ou d'aigu est perçu de façon négative par les auditeurs. Finalement : "perceived spectral balance across all band is potentially a good predictor of preference" [46]. La quantité de distorsion n'est pas apparue comme un critère d'évaluation pertinent pour les systèmes de reproduction impliqués. Une étude sémantique des commentaires des auditeurs a révélé que 94% des termes employés faisaient référence à la restitution du timbre. Seulement 3% des termes étaient relatifs à la composante spatiale de la reproduction sonore, mais le test était réalisé en reproduction monophonique. Une analyse factorielle basée sur ces mêmes commentaires a ensuite permis de mettre en évidence neuf facteurs. Seuls les deux facteurs principaux ont pour le moment été interprétés. Le premier serait associé à la "bass quality, overall spectral balance and smoothness" et le deuxième à des "specific timbre colorations in the bass, treble and midrange".

Avant de considérer les différentes mesures objectives définies sur les enceintes acoustiques dans le but de retrouver les résultats des tests perceptifs, il convient de faire le bilan des différents types de tests rencontrés, chacun présentant ses avantages et ses inconvénients.

I.4 Tout protocole expérimental constitue un compromis

Jugements de similarité et analyse multidimensionnelle Lipshitz et Vanderkooy [12] conseillent la comparaison par paires comme la tâche la plus simple qui puisse être demandée à un auditeur durant un test d'écoute. Ils recommandent des évaluations de différences relatives en l'absence de référence, ce qui est le cas pour la reproduction sonore des enceintes acoustiques [9][10][45][48] : il n'existe pas de système de reproduction parfait auquel les autres systèmes puissent être comparés. Il faudrait pouvoir comparer des musiciens jouant en direct et le même passage musical reproduit sur des enceintes. Olive [45] a détaillé les nombreuses raisons qui font que ce type de test ne peut pas être réalisé dans des conditions suffisamment contrôlées

pour que les résultats soient exploitables, et notamment la difficulté d'avoir un passage musical à reproduire qui soit une parfaite représentation des musiciens jouant en direct. Une évaluation des différences relatives entre les systèmes de reproduction apparaît donc pertinente de ce point de vue. Les jugements de similarité semblent de plus constituer une tâche "peu contraignante puisqu'elle n'impose aux sujets ni les paramètres responsables de la différence perceptive entre les sons, ni leur signification. Le sujet est libre de répondre sur ce qu'il a réellement perçu" Guyot [49].

Gabrielsson *et coll.* [19] ont utilisé ce type d'évaluation lors de leurs premiers tests. Ils ont alors pu vérifier que l'analyse multidimensionnelle permet de remonter aux dimensions perceptives utilisées par les auditeurs, à partir de leurs seuls jugements de similarité. Par contre, l'interprétation de ces dimensions reste ensuite à la charge de l'expérimentateur [10], et cette interprétation n'est pas toujours évidente. De plus, le nombre de dimensions pouvant être ainsi mises en évidence est malheureusement limité par le nombre d'enceintes impliquées dans les tests de comparaison par paire. La représentation multidimensionnelle obtenue sera en effet d'autant moins bonne que le nombre de stimuli impliqué sera plus faible [39] (annexe B.5). Kruskal et Wish [50] cités par Marozeau [51] proposent une règle empirique selon laquelle N stimuli ne peuvent conduire qu'à un maximum de $(N-1)/4$ dimensions pour que l'espace obtenu puisse être considéré comme robuste. Cette règle est également évoquée dans [32].

Pour pouvoir réaliser des évaluations de similarité, les enceintes doivent être situées à la même position dans la salle d'écoute. Dans le cas contraire il n'est plus possible de savoir si ce sont les différences de position ou d'enceinte qui sont évaluées. Il est également nécessaire que les enceintes soient comparées l'une aussitôt après l'autre, du fait de notre mémoire auditive limitée [16][45]. Une solution est d'utiliser un plateau tournant comme Toole [14] et Olive [46], même si ces auteurs n'ont pas employé ce plateau pour des tests basés sur la similarité. Cependant, ce dispositif ne permet de tester qu'un nombre limité d'enceintes au cours de la même séance, quatre dans le cas de Toole et Olive. A notre connaissance, très peu de tests perceptifs évoqués dans la littérature ont cherché à évaluer directement la similarité de la restitution du timbre par des enceintes acoustiques. Nous avons évoqué ceux de Gabrielsson *et coll.* [19] et de Klippel [10]. Dans les deux cas les enceintes étaient comparées à des positions différentes. Les tests impliquaient cinq [19] ou sept à huit [10] enceintes. Aarts [52] a réalisé une étude cherchant à vérifier qu'une enceinte égalisée pour ressembler à une autre était en effet perçue comme plus similaire après égalisation. Il a pour cela enregistré le son produit par ses différentes enceintes et a réalisé les évaluations de similarité au casque. Dans cette étude, seulement quatre à sept stimuli étaient considérés. C'est la méthode d'égalisation qui faisait l'objet de la recherche et pas directement la perception de la reproduction sonore. Les expériences de Nakayama *et coll.* [42] utilisaient également des jugements de similarité, mais portaient sur différentes configurations d'un système multicanal en chambre anéchoïque.

Différentiel sémantique et analyse factorielle Face à ces difficultés liées à la nécessité de comparaison rapide imposée par les évaluations de similarité, les recherches se sont orientées vers des évaluations absolues selon un différentiel sémantique, associées à une analyse factorielle. C'est la méthode qui fut utilisée par Gabrielsson et Sjögren [7] pour mettre en évidence des dimensions de la reproduction sonore qui ont été par la suite largement utilisées pour réaliser des évaluations de systèmes de reproduction.

Les évaluations absolues selon les différents attributs d'un différentiel sémantique n'impose pas en théorie de faire des comparaisons rapides entre enceintes. Ces enceintes peuvent donc être évaluées une par une, à la même position dans la salle d'écoute. L'analyse factorielle permet ensuite de remonter aux dimensions perceptives de la reproduction sonore, en déterminant les facteurs pertinents utilisés par les auditeurs lors de leurs évaluations. Ce type de test semble donc s'affranchir des contraintes imposées par la comparaison par paires des tests de similarité.

S'agissant des études recensées dans la littérature, nous constatons que dans la pratique, que les évaluations absolues aient été globales ou réalisées selon des attributs particuliers, elles ont utilisé des comparaisons par paires. Celles-ci ont l'avantage d'être plus discriminantes que les évaluations successives d'enceintes considérées de façon isolée [15][16][45]. Le problème du contrôle de la position des enceintes évoqué pour les tests de similarité se pose donc toujours. Les tests ont ainsi souvent été réalisés avec des enceintes situées à des positions différentes [7][10][14][15][19][22][26][27][43]. Il s'agissait alors de tester plusieurs fois les mêmes enceintes, à toutes les positions, la position changeant aléatoirement d'un test à l'autre, afin de réduire au maximum le biais lié à ce facteur expérimental. Comme les contraintes des tests de similarité, ceci limite le nombre d'enceintes pouvant être impliquées dans un même test. Toole [14] a réalisé un test avec seize enceintes en utilisant quatre positions, mais toutes les enceintes n'ont pas été évaluées à toutes les positions et toutes les combinaisons possibles pour les comparaisons n'ont pas été effectuées. Il en fut de même pour le test de Gabrielsson et Lindstrom [22] qui impliquait vingt enceintes, les auteurs ayant d'ailleurs noté une influence du contexte sur l'évaluation : les jugements concernant une enceinte dépendant des autres enceintes testées au cours de la même séance.

L'utilisation du différentiel sémantique pose le problème de l'interprétation des attributs par les auditeurs. Ainsi, Gabrielsson et Sjögren [7] annoncent à propos de leurs résultats obtenus à partir de ce type de test : "The present labels of the dimensions are provisional and may be changed in the future. There are many problems of language here. It is sometimes said that verbal terms are inadequate for describing sounds — there seem to be no words for what you perceive. Or it can be suspected that different people use different words in different ways so the same word may mean different things to different individuals. For this reason investigators of "timbre" often use multidimensional scaling techniques requiring only judgments of similarity or dissimilarity, avoiding (more or less) other descriptive terms like adjectives." Il se peut que

les auditeurs entendent la même chose, mais ne l'expriment ensuite pas de la même façon quand il leur est demandé d'utiliser des mots. L'expérience de l'auditeur semble alors déterminante pour interpréter et évaluer les attributs. La définition de la dimension "softness" donnée par Toole [14] à ses auditeurs avant les tests constitue un bon exemple: "Softness—Refers to the quality of high-frequency sounds. These should be smoothly natural, neither overly subdued and mild nor excessively hard, shrill, strident, or sharp". N'importe quel auditeur ne sera pas forcément capable d'évaluer cet attribut, et rien ne dit que différents auditeurs le feront tous de la même façon.

Ce problème d'interprétation des attributs explique pourquoi nous n'avons pas cherché à traduire les dimensions utilisées ou révélées dans les différentes études présentées. Ceci nous permet d'éviter d'imposer notre propre interprétation des termes, dans une traduction qui peut s'avérer parfois délicate.

Les évaluations absolues selon certains attributs soulèvent également la question de la pertinence des attributs utilisés par l'expérimentateur dans le cadre de son test d'écoute. Les dimensions perceptives mises en jeu dans un test dépendent du contexte [53], et notamment du nombre et de la nature des enceintes qui sont testées. Il peut arriver que certains attributs choisis par l'expérimentateur ne soient pas pertinents dans le cadre de son expérience particulière. Olive [46] a ainsi trouvé que les évaluations de quantité de distorsion non-linéaire perçue étaient très corrélées aux évaluations de préférence des auditeurs, alors que ces mêmes auditeurs ont indiqué dans leurs commentaires que la présence de distorsion n'avait pas été un critère pertinent pour effectuer leurs jugements. Les auditeurs ont donc évalué des différences de distorsions alors qu'ils ne les ont pas entendues. Olive explique ces résultats par un "halo effect", un biais décrivant la tendance des auditeurs à évaluer différentes dimensions selon leurs préférences relatives parmi les stimuli présentés. "This bias reinforces the need to carefully consider the choice and relevancy of scales and cross-referencing the ratings with other scales, post-interviews and objective measurements" Olive [46].

Comparer rapidement à la même position Nous avons pu constater les problèmes expérimentaux posés par la combinaison du contrôle de la position des enceintes et de la nécessité de pouvoir comparer ces enceintes rapidement. L'utilisation d'un plateau tournant semble être une très bonne solution, mais elle reste restreinte aux tests impliquant relativement peu d'enceintes. Une deuxième solution envisageable consiste à enregistrer le son rayonné par différentes enceintes à la même position, et à réaliser ensuite les tests d'écoute au casque. Certains travaux ont comparé les résultats de tests faits en direct à ceux de tests réalisés au casque [16][44][54][55]. Même si les buts de ces différentes études étaient divers, elles ont toutes montré que les deux configurations donnaient des résultats similaires en ce qui concerne l'évaluation relative de la restitution du timbre.

Toole [44] a réalisé deux tests d'écoute à l'aide de quatre enceintes utilisées en reproduction monophonique. Les auditeurs devaient évaluer la fidélité des systèmes. La composante spatiale de la reproduction n'était pas considérée. Le premier test a été fait en écoute directe, avec quatre positions différentes des enceintes. Le deuxième test a été effectué à partir d'enregistrements binauraux réalisés dans la même configuration que celle du premier test. Les résultats ont montré que le classement des enceintes selon l'échelle considérée était le même pour les deux tests, le test au casque pouvant être plus discriminant.

Dans le but d'étudier l'influence de la position des enceintes sur des évaluations de "préférence", Olive *et coll.* [16] ont comparé un test en direct et un test au casque. Quatre positions et deux modèles d'enceinte, avec deux enceintes par modèle, ont été utilisés pour le test en direct et la réalisation d'enregistrements binauraux. Les enceintes étaient comparées par quatre lors du test en direct. En plus des évaluations de préférence, le test au casque a permis des comparaisons par paires, avec notamment les comparaisons des deux modèles différents aux mêmes positions, ce qui n'est pas faisable lors d'une écoute en direct. Les résultats des deux tests ont été similaires, le test au casque permettant de mesurer des effets plus subtils tels que l'interaction des facteurs enceinte, position et extrait musical.

Bech *et coll.* [54][55] ont comparé deux tests perceptifs utilisant des signaux de synthèse. L'étude portait sur la perception des basses fréquences. Le premier test a été effectué à partir d'une enceinte en chambre anéchoïque, tandis que le deuxième utilisait une reproduction au casque. Les auteurs n'ont pas observé de différences significatives entre les données obtenues à partir de ces deux méthodes de reproduction [54]. Cette étude utilisant des reproductions synthétiques, les résultats sont peut être moins généralisables à des tests impliquant de véritables enceintes. Pour les deux tests, toutes les variations dans les signaux étaient contrôlées par l'expérimentateur. L'équivalence des deux types de test a donc été démontrée uniquement pour un nombre limité de paramètres choisis des signaux.

D'autres tests perceptifs évoqués précédemment ont été effectués au casque à partir d'enregistrements impliquant différents systèmes de reproduction, mais ils n'ont pas été comparés à des tests réalisés en direct sur ces mêmes systèmes. Gabrielsson et Sjögren [7] ont ainsi enregistré le son produit par différentes prothèses auditives, afin d'étudier les dimensions perceptives associées à leur reproduction sonore. Aarts [52] a utilisé des enregistrements impliquant des enceintes ayant ou non subi une égalisation, afin d'évaluer leur similarité.

Ces tests réalisés au casque à partir d'enregistrements permettent une comparaison rapide des enceintes placées à la même position dans la salle d'écoute. Par contre, par rapport à une écoute en direct, le signal délivré aux oreilles de l'auditeur est modifié par les phases d'enregistrement et de restitution au casque. Cette démarche n'est donc pas adaptée aux évaluations absolues de la reproduction sonore. Il semblerait également délicat d'étudier la composante spatiale de la reproduction des enceintes par un tel protocole. Celle-ci pourrait en effet être trop déformée par

l'écoute au casque.

Faire un compromis L'expérience idéale consisterait à pouvoir comparer rapidement en écoute directe de nombreuses enceintes à la même position. Ceci n'est malheureusement pas possible. L'ensemble des expériences décrites dans la littérature nous a montré que du fait des nombreuses contraintes inhérentes au test d'écoute sur les enceintes acoustiques, tout protocole expérimental constitue un compromis. Les différents types de test possèdent leurs avantages et leurs inconvénients, aucun d'eux n'étant parfait. Le contrôle de la position des enceintes et leur comparaison rapide représente une des plus grandes difficultés pour l'expérimentateur. Les choix à effectuer dépendent bien sûr du but final de l'expérience.

L'objectif de tous ces tests est bien sûr de trouver une ou des méthodes de mesure objectives qui permettent de les remplacer. Il convient donc d'étudier les différentes recherches qui ont été menées pour rapprocher les mesures objectives du résultat de ces tests perceptifs.

I.5 Recherches de mesures objectives pertinentes

Toutes les recherches de mesures objectives pertinentes pour évaluer la restitution du timbre par les enceintes acoustiques s'accordent sur un point : "the frequency response of the loudspeaker is the most important factor related to perceived sound quality" Olive [56]. Par contre, Toole [8] montre qu'il existe beaucoup de divergences concernant la façon de mesurer cette réponse en fréquence. Il récapitule les différentes méthodes qui peuvent être utilisées. Une mesure anéchoïque dans l'axe de l'enceinte évalue le champ direct, tandis que prendre en compte quelques directions hors axe permet de considérer l'influence que pourront avoir les premières réflexions. La puissance rayonnée dans toutes les directions donne des indications sur le champ réverbéré. Ces mesures sont ensuite effectuées dans différents environnements : champ anéchoïque, salle réverbérante ou salle d'écoute. Toole fait alors remarquer que toutes ces réponses en fréquence ne peuvent pas être plates en même temps. Le critère devant rendre compte de la qualité du système doit donc être déterminé plus précisément. Il ajoute à propos de ce lien entre la forme de la réponse en fréquence des enceintes et leur qualité perçue : "evidence that offends the eye may or may not indicate the presence of a problem that is offensive to the ear" [8]. Bücklein [57] a par exemple montré qu'à "taille équivalente" les trous dans une réponse fréquentielle sont moins audibles et gênants que les pics.

La question de l'interprétation des réponses en fréquence se pose en effet, et avant elle, celle de la façon la plus pertinente d'observer les signaux de mesure. Ainsi Janse et Kaiser [58][59] conseillent d'utiliser la distribution de Wigner-Ville pour comparer les réponses d'enceintes, tandis que Heyser [60] propose un plan temps-fréquence qui serait plus proche de la perception. A notre connaissance, ces propositions ne semblent pas avoir été exploitées dans la pratique et leur

application est restée limitée à des considérations théoriques ou à des simulations numériques.

Olive dans [46][56] récapitule les différents modèles objectifs envisagés pour rendre compte des résultats de tests perceptifs. Il sépare ces modèles en fonction du type d'environnement de mesure auxquels ils font appel : mesures en salle réverbérante, en chambre anéchoïque ou en salle d'écoute. Il présente cette séparation comme révélatrice de points de vue différents concernant ce qui est important dans la reproduction sonore, l'accent étant mis sur le champ direct, les premières réflexions ou le champ réverbéré. Nous avons préféré regrouper les études en fonction du mode de comparaison objectif/perceptif qu'elles ont utilisé.

I.5.1 Comparaison par corrélation "visuelle"

La recherche de corrélations "visuelles" consiste à comparer visuellement les réponses en fréquence des enceintes impliquées dans les tests d'écoute, et à déceler des tendances sur la forme de ces réponses fréquentielles qui pourraient expliquer le résultat des évaluations perceptives.

Staffeldt [26] a comparé la forme de mesures de réponse en fréquence en tiers d'octave réalisées en chambre anéchoïque, en salle réverbérante et dans sa salle d'écoute, avec les deux dimensions "emphasized treble" et "emphasized bass" révélées par ses tests perceptifs. La mesure dans la salle d'écoute, moyennée à trois positions dans une zone correspondant à la tête de l'auditeur et associée à un modèle auditif [61][62], fournit les meilleurs résultats. Ils sont meilleurs en ce sens que la forme de la réponse fréquentielle ainsi déterminée est plus facilement interprétable par l'expérimentateur pour expliquer le résultat de ses tests perceptifs. La mesure en salle réverbérante donne des résultats proches de la mesure en salle d'écoute, meilleurs que ceux de la réponse anéchoïque. Mais Olive [46] signale à ce sujet que la salle utilisée par Staffeldt était justement très réverbérante et s'écartait des recommandations concernant la salle d'écoute à utiliser pour un test d'écoute sur des enceintes acoustiques [30][31]. Suite à ces expériences, Staffeldt propose un modèle permettant de prédire la restitution du timbre par les enceintes [63], à partir de leur réponse fréquentielle en salle d'écoute. Il montre ensuite que lorsque cette réponse est égalisée pour deux enceintes au départ bien différentes, leur reproductions sonores sont perçues comme plus proches après égalisation, démontrant ainsi la pertinence de sa méthode de mesure [43].

Gabrielsson *et coll.* [11] ont aussi cherché à confronter des mesures en chambre anéchoïque, en salle réverbérante et en salle d'écoute, avec les évaluations d'enceintes acoustiques selon leurs différentes échelles perceptives [22]. Pour mener cette confrontation, il se sont appuyés sur les résultats d'une étude [64] au cours de laquelle ils avaient testé l'influence de différentes modifications de la réponse en fréquence sur ces dimensions perceptives, à l'aide de signaux de synthèse et d'une reproduction au casque. Cette étude leur avait permis d'émettre des hypothèses quand au lien entre la forme de la réponse fréquentielle d'un système de reproduction et son évaluation selon les différentes dimensions perceptives. Il serait trop long de reporter le détail de ces hypothèses

ici, l'important est de savoir que c'est l'accord avec ces hypothèses qui a servi de critère pour évaluer la pertinence des différentes méthodes de mesure de réponse en fréquence d'enceintes dans [11]. Les dimensions "clarity" et "spaciousness" ont été les deux dimensions perceptives principalement considérées lors de cette recherche de corrélation visuelle. Les auteurs rapportent un bon accord avec les hypothèses de [64] pour la mesure en salle d'écoute, tandis que cet accord n'est que limité pour les deux autres types de mesure. La mesure en salle d'écoute était réalisée grâce à un microphone omnidirectionnel situé à la position d'écoute centrale. Seule l'enceinte gauche de la paire stéréophonique impliquée dans les tests d'écoute était mesurée. La résolution fréquentielle était d'un octave. "It is also reasonable to assume that the listener makes a similar smoothing of the complex stimulus in perceiving and judging the sound quality. Rather than tracing all the ups and downs in the stimulus spectrum, the listener probably makes an integration (smoothing) into a limited number of frequency, e.g., bass, midrange and treble, or perhaps higher bass, lower and higher midrange, and lower and higher treble. Each of these categories encompasses at least an octave." Gabrielsson *et coll.* [11]. Ils n'ont donc pas conservé une résolution en tiers d'octave, pourtant davantage en accord avec la largeur des filtres auditifs [65][66][67], car ils ont considéré que les auditeurs effectuaient une intégration sur des domaines fréquentiels plus étendus. Afin de prendre en compte l'influence de l'extrait musical, les réponses en fréquence des enceintes ont également été multipliées par le spectre des différents extraits utilisés pendant les tests d'écoute. Le type de réponse fréquentielle obtenu vérifiait également bien les hypothèses devant relier réponse en fréquence et perception.

Toole [68] a lui aussi réalisé des corrélations visuelles entre les trois types de réponse fréquentielle évoqués et les évaluations de fidélité obtenues lors de ses tests d'écoute [14]. Il considère la moyenne des mesures effectuées à plusieurs positions. "The purpose of spatial averaging is to eliminate irregularities in the curves caused by interference effects that change with microphone location, leaving intact the irregularities due to other causes" Toole [68]. Ses mesures ont également une grande résolution fréquentielle, de un dixième à un vingtième d'octave sur l'ensemble de la gamme de fréquence audible, et il considère que les mesures en tiers d'octave "cannot reveal the fine details of the amplitude response that the present tests indicate are important to the identification of the highest rated loudspeakers" [68]. Il trouve finalement que les mesures en salle d'écoute sont imprécises, dures à interpréter et difficile à généraliser du fait de l'influence de la pièce. Il préfère les mesures anéchoïques, même si elles ne correspondent pas à l'environnement d'écoute. Elles permettent selon lui de prédire la réponse en salle d'écoute si les caractéristiques de la pièce sont connues. Il propose de combiner les réponses anéchoïques mesurées à différents angles autour de l'enceinte, avec une pondération spécifique pour des régions angulaires qu'il relie au champ direct en situation d'écoute, aux premières réflexions et au champ réverbéré. C'est l'interprétation de ce type de mesure qui lui permet d'expliquer au mieux les jugements de fidélité. De nombreux éléments sont à considérer simultanément sur ces courbes de réponse pour

expliquer les jugements des auditeurs : réponse dans l'axe et hors axe, courbe plate et absence d'irrégularité, bande passante... "What we do not learn (...) are the priorities of these performances measures. Need they all be equally good?" Toole [68]. C'est un des inconvénients d'une approche perceptive basée sur une évaluation globale de qualité ne mettant pas en évidence les dimensions perceptives mises en jeu, ni leur importance relative.

Le problème de cette approche basée sur une simple observation des formes de réponses en fréquence est soulevé par Moore et Tan [47] lorsqu'ils écrivent à propos de l'étude de Gabrielsson *et coll.* [11] : "no quantitative relationships were established between frequency-response shapes and the perceptual ratings", ou encore à propos de l'étude de Toole [68] : "no attempt was made to quantify the relationship between the degree of smoothness of the frequency response and the preferences of the listeners". L'évaluation du lien entre mesures objectives et perceptives reste alors en effet essentiellement basée sur l'interprétation de l'expérimentateur. Les comparaisons visuelles semblent discriminantes mais elles peuvent être assez complexes à réaliser, comme l'ont montré les nombreux critères à interpréter sur les courbes de réponses proposées par l'étude de Toole [68].

I.5.2 Comparaison par corrélation grandeur perceptive/grandeur objective

Klippel [10][48] a eu une approche assez différente et hybride par rapport à celles présentées au paragraphe précédent. Il a utilisé des mesures anéchoïques pour simuler les conditions des tests perceptifs dans la salle d'écoute. La réponse anéchoïque dans l'axe de l'enceinte lui a donné accès au champ direct, auquel il a ajouté une approximation du champ réverbéré produit par l'enceinte dans la pièce. Cette approximation a été calculée à partir de la puissance rayonnée par l'enceinte en chambre anéchoïque, du temps de réverbération et des dimensions de la pièce. Il a ainsi simulé la réponse fréquentielle de l'enceinte dans la salle d'écoute, à la position de l'auditeur. En la combinant au spectre des extraits musicaux utilisés pendant les tests perceptifs, il a obtenu une approximation du spectre des signaux écoutés par ses auditeurs. Il a ensuite utilisé un modèle auditif [61][62] pour simuler les phénomènes perceptifs et se rapprocher au plus près de ce qui a été perçu lors des tests. L'ensemble de l'évaluation s'est faite en tiers d'octave. C'est à partir de ce spectre final "perçu" qu'il a défini des attributs objectifs qui ont pu être bien corrélés aux jugements des auditeurs.

Dans [46], Olive constate que les réponses fréquentielles anéchoïques moyennées sur la "fenêtre d'écoute" ("listening window"), zone de l'espace définie par des directions horizontales à plus et moins trente degrés par rapport à l'axe de l'enceinte et des directions verticales à plus et moins dix degrés, sont très corrélées avec les évaluations de balance spectrale perçue des auditeurs, avec un coefficient de corrélation de 0,95. Il conclut donc que la prise en compte du champ direct est importante pour évaluer la perception du timbre, et que les seules mesures de puissance rayonnée

ne suffisent pas.

Il construit ensuite un modèle permettant de prédire les jugements de préférence des enceintes à partir de mesures objectives anéchoïques [56]. Les mesures considérées sont la réponse en fréquence dans l'axe, dans la "fenêtre d'écoute", la puissance rayonnée, et une simulation de la réponse en salle d'écoute obtenue à partir des mesures anéchoïques. Après avoir défini plusieurs attributs objectifs sur ces différentes courbes de réponse, il obtient un modèle de préférence à partir d'une régression multiple sur ces attributs. Ce modèle s'avère meilleur que d'autres modèles impliquant seulement les réponses fréquentielles en salle d'écoute, mesurées ou calculées, ou bien uniquement la courbe de puissance rayonnée. Il montre aussi que la résolution fréquentielle en tiers d'octave est insuffisante et préconise une résolution en vingtième d'octave. Finalement, l'utilisation d'un modèle auditif est une perspective envisagée pour améliorer son modèle.

Un modèle du même type a été développé par Bramsløw [9] pour évaluer de façon objective les "clearness" et "sharpness" d'extraits musicaux ayant subi différentes modifications fréquentielles. Ce modèle a été proposé dans le cadre de l'évaluation de la qualité des prothèses auditives. Il est basé sur des tests perceptifs, un modèle auditif utilisé pour analyser les signaux et pouvant prendre en compte des pertes auditives potentielles, et un réseau de neurones permettant de relier la sortie du modèle auditif aux résultats des tests perceptifs.

I.5.3 Bilan des recherches

Les recherches de mesures objectives pertinentes vis-à-vis de la perception de la restitution du timbre ont toutes montré que la mesure de la réponse en fréquence des enceintes était déterminante. Mais il reste encore des interrogations concernant la manière la plus adaptée pour effectuer cette mesure.

Quel environnement de mesure? Gabrielsson *et coll.* [11] et Staffeldt [26][43][63] ont trouvé que la mesure directe en salle d'écoute, dans la configuration utilisée pour les tests perceptifs, permettait le mieux d'expliquer les résultats de leurs tests. Toole [68] et Olive [46][56] préfèrent les mesures anéchoïques plus contrôlées et permettant selon eux de prédire les performances en salle d'écoute. Klippel [10][48] adopte une approche hybride dans laquelle il utilise les mesures anéchoïques pour simuler a posteriori les conditions de ses tests d'écoute.

Une chambre anéchoïque constitue un environnement neutre pour les enceintes qui permet de s'affranchir de l'influence de la pièce et d'effectuer une mesure caractéristique des enceintes. Par contre cette mesure ne considère pas les enceintes dans leur environnement normal d'utilisation. Avec des mesures en salle d'écoute, c'est l'ensemble enceinte-salle qui est évalué, mais c'est également le cas lors des tests d'écoute. La mesure objective semble donc plus facilement fiable à l'évaluation perceptive correspondante. Par contre, tout comme cette évaluation perceptive, elle sera modifiée lorsqu'une autre pièce sera considérée. Il y a donc un compromis à trouver

entre une mesure à la fois indépendante de la pièce et en même temps facilement fiable aux résultats des tests perceptifs.

Quelle résolution fréquentielle? Les études évoquées ont également montré qu'il n'y avait pas de consensus sur la résolution fréquentielle requise pour évaluer la réponse en fréquence des enceintes. Alors que Gabriëlsson *et coll.* [11] se contentent d'une résolution d'un octave, Staffeldt [26][43][63] et Klippel [10][48] utilisent le tiers d'octave, tandis de Toole [68] et Olive [46][56] prônent les mesures haute résolution jusqu'au vingtième d'octave. Ces derniers conseillent de réaliser une moyenne spatiale des réponses, pour faire disparaître les irrégularités dues aux interférences acoustiques tout en conservant une résolution fréquentielle permettant de détecter les phénomènes de résonances caractéristiques de l'enceinte.

Quels signaux de mesure? La question des signaux à considérer pour faire le lien entre mesures objectives et perceptives se pose également. Staffeldt [26][43][63], Toole [68] et Olive [46][56] considèrent directement les réponses en fréquence des enceintes tandis que Gabriëlsson *et coll.* [11], Klippel [10][48] et Bramsløw [9] prennent en compte l'influence de l'extrait musical en multipliant les réponses fréquentielles par le spectre de différents extraits utilisés.

Utilisation de modèles auditifs? Certains ont également cherché à tenir compte des processus perceptifs impliqués dans le jugement des auditeurs, en ayant recours à un modèle auditif. C'est le cas de Staffeldt [26][43][63], Klippel [10][48] et Bramsløw [9]. Olive [56] évoque également cette perspective pour améliorer son modèle. D'autres auteurs ont conseillé l'utilisation de tels modèles pour évaluer la reproduction sonore [69][70][71].

Klippel indique concernant ces modèles: "several effects observed in listening tests can be directly explained by the loudness spectrum. Dips in frequency response show lesser perceptibility than peaks with the same magnitude, since dips are leveled by masking effects, and peaks produce additional flanking excitations" [10]. Cette remarque est à mettre en relation avec celle de Toole citée au début de ce paragraphe à propos de la recherche de mesures objectives pertinentes, indiquant que ce n'est pas forcément ce qui se voit le plus sur une réponse en fréquence qui sera le plus perçu à l'écoute. La remarque de Klippel semble indiquer que l'utilisation d'un modèle auditif peut en ce sens faciliter l'interprétation des courbes de réponse.

Une même tendance : simuler les conditions d'écoute Les mesures objectives normalisées ne permettant pas de faire le lien avec les mesures perceptives, les différentes recherches menées pour trouver ce lien ont toutes tenté de définir des mesures qui se rapprochent davantage des conditions d'écoute. Lorsque la mesure n'est pas effectuée directement en salle d'écoute, cet environnement est simulé plus ou moins complètement à partir de mesures anéchoïques. Après

avoir pris en compte l'influence de la pièce, certaines mesures objectives considèrent également l'influence de l'extrait musical, puis celle du système auditif de l'auditeur. La mesure devient alors moins caractéristique des enceintes et plus dépendante des différents facteurs expérimentaux, mais il est souvent plus facile de la relier aux jugements des auditeurs.

Plusieurs types de comparaison objectif/perceptif Les études décrites ont montré qu'il pouvait être délicat de faire correspondre les évaluations globales de la qualité avec la forme des réponses fréquentielles mesurées. Les auteurs ont parfois recherché simplement des relations visuelles qui posent le problème de l'interprétation complexe des multiples critères à considérer simultanément. Les modèles d'Olive [46][56] ou Bramsløw [9] relient qualité perçue et note objective, mais ne permettent pas forcément d'explicitier davantage ce lien qui est obtenu en optimisant une régression linéaire multiple ou en utilisant un réseau de neurones. L'approche de Klippel [10][48] vise d'abord à rechercher des attributs objectifs décrivant les dimensions perceptives révélées par ses tests d'écoute. Ensuite, il utilise ces attributs pour définir une évaluation objective globale de la qualité de ses enceintes acoustiques.

Afin de vérifier des hypothèses établies en comparant les réponses fréquentielles d'enceintes et les résultats de tests perceptifs, ou encore pour émettre de nouvelles hypothèses concernant le lien entre la forme de ces courbes de réponse et la perception, certaines études ont été menées en utilisant des signaux de synthèse.

I.6 Le recours aux signaux de synthèse

Gabrielsson *et coll.* [64] ont réalisé des tests perceptifs au casque pour déterminer l'influence de la forme de la réponse en fréquence et du niveau sonore sur les dimensions perceptives de la reproduction sonore mises en évidence lors d'études précédentes [7][19][22]. Trois programmes monophoniques ont été utilisés, soit sans avoir été modifiés, soit filtrés avec une amplification dans trois gammes de fréquence différentes : les basses fréquences en dessous de 200 Hz, le médium autour de 1 kHz et les aigus autour de 4 kHz. Le niveau sonore constituait une deuxième variable. Les tests perceptifs consistaient à évaluer les signaux selon les différentes dimensions perceptives étudiées. Les résultats obtenus ont ensuite été étendus à la perception de la reproduction sonore en général, et utilisés par Gabrielsson *et coll.* [11] comme hypothèses de travail pour rechercher la méthode la plus pertinente pour mesurer la réponse en fréquence des enceintes acoustiques — la méthode la plus pertinente étant celle qui donnerait des résultats conformes aux hypothèses établies dans [64].

Bech *et coll.* [54][55] ont étudié l'influence de différents paramètres de la réponse aux basses fréquences des enceintes : le niveau sonore, la fréquence de coupure basse, l'ordre du filtre passe-haut correspondant à cette coupure, et l'amplitude d'"ondulations" dans la réponse en fréquence

simulant l'effet produit par des pièces aux durées de réverbération différentes. Quatre signaux musicaux ont été filtrés selon ces différents paramètres. Les tests d'écoute ont été réalisés à la fois au casque et à l'aide d'une enceinte en chambre anéchoïque, les deux types de tests ayant donné des résultats analogues. Les signaux étaient comparés à une référence fixe et l'auditeur devait indiquer s'il percevait plus ou moins de "lower bass" et "upper bass" dans le signal filtré. Cette étude a montré une influence de la fréquence de coupure basse sur le niveau de "lower bass" lorsque le niveau est au dessus du seuil d'audition dans les bandes de fréquences correspondantes. L'ordre du filtre passe-haut n'a pas eu d'effet dans les conditions testées. Les ondulations dans la réponse en fréquence, simulant l'effet d'une pièce, modifient également le niveau de "lower bass" et "upper bass".

Moore et Tan [47] ont effectué des tests perceptifs au casque pour étudier l'effet de différentes modifications de la réponse en fréquence d'un système. De nombreuses distorsions linéaires ont été testées à l'aide de différents filtres : "spectral tilts and ripples of various types, variations in upper and lower cutoff frequency, and combinations of these". Même si les signaux musicaux considérés étaient large bande, les distorsions ont été appliquées sur une gamme fréquentielle restreinte, de 100 à 7000 Hz, correspondant à la bande passante de téléphones de haute-qualité. L'auditeur avait pour tâche d'évaluer la qualité des signaux sur une échelle variant de "very unnatural-highly colored" à "very natural-uncolored". Les résultats indiquent que les distorsions ont d'autant plus d'effet qu'elles sont appliquées sur une large région fréquentielle. Les filtres à pente linéaire ("tilts") entraînent la même dégradation de qualité pour des pentes positives ou négatives. Lorsque la fréquence de coupure basse est relativement élevée, alors une modification de la fréquence de coupure haute a relativement peu d'influence. Il en est de même pour une modification de la fréquence de coupure basse si la fréquence de coupure haute est déjà basse. Pour les filtres à bande limitée ayant une réponse plate dans la bande passante, les fréquences de coupure ont une influence importante. Les ondulations ("ripples") dégradent d'autant plus la qualité que leur amplitude est plus grande et qu'elles sont plus resserrées.

Plusieurs études se sont également intéressées à la perception des résonances dans une réponse en fréquence, ou à celle des réflexions provenant des différents murs de la pièce lors d'une reproduction par des enceintes en salle d'écoute, ces réflexions étant responsables de résonances dans la réponse en fréquence de l'enceinte mesurée dans la pièce. Ces études ont ainsi cherché à prévoir l'influence de la salle d'écoute sur la reproduction sonore des enceintes acoustiques, dans le but de prédire la qualité de la reproduction dans un environnement donné.

Les résultats de Bücklein [57] ont été évoqués précédemment. Il a étudié l'effet des pics et vallées dans la réponse en fréquence d'un système électroacoustique, en réalisant des tests au casque à l'aide de nombreux extraits sonores. Il a ainsi montré que les pics sont plus audibles et gênants que les vallées de taille équivalente. Ils réduisent davantage l'intelligibilité pour les signaux de parole. Les pics et vallées sont d'autant plus audibles qu'ils sont larges. Toole et

Olive [18] ont mesuré les seuils de perception de résonances dans diverses conditions, en regardant l'influence de ces conditions et de nombreux autres facteurs expérimentaux : la fréquence, le facteur de qualité, l'amplitude relative, et le retard temporel des résonances, la performance auditive de l'auditeur, le programme utilisé, la directivité du haut-parleur et la réverbération ajoutée ou non lors de l'écoute au casque, en chambre anéchoïque ou en salle d'écoute.

Bech [72] a également cherché à évaluer l'influence des réflexions dans une salle d'écoute sur la restitution du timbre par une enceinte acoustique. Il a pour cela utilisé un dispositif constitué de haut-parleurs entourant l'auditeur dans une chambre anéchoïque. Ce dispositif avait pour but de simuler le champ rayonné par une enceinte omnidirectionnelle dans une salle d'écoute : le champ direct, les premières réflexions et le champ réverbéré pouvant être contrôlés indépendamment. Le but de l'étude était d'évaluer l'influence individuelle des premières réflexions. Les seuils de perception de ces réflexions considérées individuellement ont été déterminés pour un signal de bruit et un signal de parole. Ces seuils ont été comparés aux niveaux de ces mêmes réflexions prédits par la simulation du champ sonore dans une pièce d'écoute standard. Si le seuil d'une réflexion est au dessus de son niveau prédit, alors cette réflexion ne devrait pas être perçue et ne devrait pas influencer la restitution du timbre par l'enceinte acoustique. Les réflexions du premier ordre sur le plafond et le sol sont les plus perceptibles, suivies peut être par la réflexion sur le mur opposé à l'enceinte simulée. Les autres réflexions ont des seuils de perception au dessus des niveaux prédits et ne devraient pas avoir d'influence. Le système de simulation a ensuite été amélioré [73]. La directivité de l'enceinte et la dépendance fréquentielle des absorbants dans la pièce ont été prises en compte. Seule la réflexion sur le sol contribue alors individuellement au timbre dans le cas du bruit, tandis qu'aucune des réflexions n'intervient dans le cas de la parole. Bech conclue que seule la réflexion sur le sol risque d'avoir un effet sur la restitution du timbre. La gamme de fréquence comprise entre 500 Hz et 2 kHz semble particulièrement déterminante. Les changements dans cette gamme de fréquence détermineraient les changements de timbre, et donc les seuils de détections des réflexions. Il trouve dans [73] que les seuils ont augmenté par rapport à l'étude utilisant la simulation précédente [72], et conclut donc que les résultats de [72] correspondent à la situation "the most sensitive for the majority of reflections, and that the introduction of acoustically more realistic conditions will reduce the influence of these individual reflections" [73].

Concernant le recours aux signaux de synthèse, certaines recherches ont également impliqué des enceintes dont la réponse en fréquence était contrôlable par l'expérimentateur. Gabrielson *et coll.* [7][19] ont utilisé de telles enceintes pour explorer les dimensions perceptives de la reproduction sonore. Une telle approche doit tout de même être abordée avec précaution, l'expérimentateur pouvant avec cette démarche générer lui-même artificiellement les dimensions qu'il cherche à découvrir. Après avoir mis en évidence certaines dimensions perceptives et des attributs objectifs les décrivant, Klippel [10][48] a employé des enceintes égalisables pour étudier

les liens entre ses attributs objectifs et des évaluations globales d'agrément ou de naturel de la reproduction ("naturalness").

Le recours aux signaux de synthèse offre à l'expérimentateur l'avantage de connaître précisément, de contrôler et de pouvoir faire varier les mesures objectives. Le lien entre grandeurs perceptives et objectives est conservé lors de toute l'expérience. Par contre, le cadre de l'expérience s'éloigne parfois des situations d'écoute réelles et les résultats peuvent apparaître plus difficiles à généraliser [47]. Si l'étude vise à révéler des dimensions de la reproduction sonore, ces signaux "artificiels" peuvent aussi conduire à des dimensions "artificielles" qui sont en fait créées par l'expérimentateur, et qui ne seraient peut être pas apparues en comparant des systèmes réels.

I.7 Bilan des différentes approches

Face aux limitations des mesures objectives normalisées, de nombreuses études ont cherché une mesure objective plus pertinente pour caractériser la perception de la reproduction sonore. Elles ont tout d'abord nécessité la mise en place de tests d'écoute rigoureusement contrôlés. Ces tests ont eu pour but de révéler les dimensions perceptives utilisées par les auditeurs pour effectuer leurs jugements, ou de réaliser des évaluations globales de la qualité des systèmes. La recherche des dimensions perceptives impliquées dans les tests fait intervenir des analyses statistiques telles que l'analyse multidimensionnelle ou l'analyse factorielle. L'utilisation de signaux de synthèse peut également être envisagée pour mieux comprendre le lien objectif/perceptif.

Les recherches ont toutes montré que la restitution du timbre par les enceintes acoustiques était caractéristique de leur réponse en fréquence. Par contre, les avis divergent concernant la meilleure méthode pour mesurer cette réponse fréquentielle. Les différentes approches essaient par contre toutes, d'une façon ou d'une autre, de prendre en compte les conditions d'écoute pendant les tests perceptifs.

Les études rencontrées dans la littérature tentent de rapprocher a posteriori des mesures perceptives et objectives au départ disjointes. Il sera montré au chapitre suivant que notre approche consiste à effectuer le cheminement inverse. Nous avons tenté de garder les mesures objectives et perceptives le plus proche possible en les menant en parallèle, afin de mieux comprendre leur lien. Une fois ce lien bien établi, le but est de séparer progressivement les évaluations objective et perceptive, pour rechercher une mesure objective plus caractéristique des enceintes acoustiques et moins dépendante des conditions de mesure.

Le but premier de notre recherche est donc le rapprochement objectif/perceptif. Nous verrons dans la suite que nous n'avons pas essayé d'apporter de nouvelles connaissances concernant la perception de la reproduction sonore et les mesures objectives d'enceintes acoustiques considérées séparément. C'est le lien entre ces deux domaines qui a fait l'objet de notre recherche.

CHAPITRE II

Approche et protocole expérimental choisis

----- oOo -----

Ce chapitre présente la façon dont nous avons décidé d'aborder le problème de la confrontation des évaluations objectives et perceptives des enceintes acoustiques. Le protocole expérimental qui a été mis au point pour réaliser cette confrontation est décrit. Les outils d'analyse employés sont également détaillés.

----- oOo -----

SOMMAIRE

II.1	Evaluation des différences relatives entre enceintes acoustiques	33
II.2	Protocole expérimental	34
II.3	Limitations et avantages du protocole	37
II.4	Dissemblances perceptives	39
II.4.1	Comparaison par paires	40
II.4.2	Classification libre	40
II.4.3	Validité de l'expérience psychoacoustique	41
II.4.3.a	Validité de l'estimation des dissemblances perceptives	41
II.4.3.b	Validité de l'analyse des dissemblances perceptives	43
II.5	Dissemblances objectives "classiques"	44
II.5.1	Synchronisation des signaux enregistrés	45
II.5.2	Techniques d'analyse "classiques" considérées	45
II.5.2.a	Domaine temporel	46
II.5.2.b	Domaine fréquentiel	47
II.5.2.c	Plan temps-fréquence	47
II.5.2.d	Domaine fréquentiel avec pondérations	48
II.6	Utilisation de modèles auditifs : dissemblances "perceptives objectives"	48
II.6.1	Densité de sonie totale : estimation du spectre perçu	51
II.6.2	Densités de sonie : estimations du plan temps-fréquence perçu	52
II.7	Bilan de l'approche proposée	54

II.1 Evaluation des différences relatives entre enceintes acoustiques

Avant de chercher à savoir pourquoi une enceinte serait "meilleure" qu'une autre, il nous semble intéressant, dans un premier temps, de comprendre pourquoi elles sont perçues différemment. Notre but est donc de retrouver par une mesure objective les différences perçues par des auditeurs comparant différentes enceintes. Nous avons décidé de baser notre approche psychoacoustique sur l'évaluation des différences relatives entre enceintes acoustiques plutôt que sur celle de leur qualité absolue. En évitant les jugements de qualité, de fidélité ou de préférence, nous espérons être moins dépendants des goûts et a priori des auditeurs. Les jugements de différences perçues nous donnerons également accès aux dimensions perceptives particulières impliquées dans nos tests. Connaître ces dimensions pourrait nous donner une idée du type de mesure objective qui serait le plus adapté pour les retrouver.

Nous n'avons pas non plus utilisé d'évaluation absolue sur les échelles perceptives mises en évidence par les recherches antérieures que nous avons évoquées au chapitre précédent. Quand de telles échelles sont utilisées, le problème de leur interprétation par les auditeurs se pose. De plus, en supposant que les auditeurs aient compris à quoi correspondaient ces échelles, et qu'ils l'aient tous compris de la même manière, il faut ensuite qu'ils soient capables d'évaluer les enceintes selon ces échelles, ce qui pourrait être très dépendant de l'expérience personnelle et du niveau d'expertise de chacun. Le fait de demander d'évaluer simplement la différence entre enceintes acoustiques sans avoir à nommer ou à caractériser les différences perçues semble être une tâche plus simple et naturelle pour l'auditeur, moins dépendante de son niveau d'expérience. Par ailleurs, nous ne savons pas à l'avance quelles dimensions perceptives seront impliquées dans nos tests d'écoute. Tout dépend des enceintes qui sont incorporées dans le test. Notre démarche nous permet donc d'éviter également notre propre a priori, celui de l'expérimentateur qui choisit plus ou moins arbitrairement les échelles selon lesquelles seront évaluées les enceintes.

De manière parallèle à l'approche psychoacoustique, l'approche physique vise à mettre en évidence des différences objectives entre les enceintes mesurées. Ces différences objectives sont systématiquement comparées aux différences perçues par les auditeurs. Notre but est ainsi d'identifier une méthode de mesure qui différencie les enceintes de la même manière que les auditeurs lors des tests d'écoute. Une méthode de mesure pertinente vis-à-vis de notre perception de la reproduction sonore constituerait un outil en vue d'une évaluation objective de la qualité des enceintes acoustiques (Figure II.1). Une fois que les enceintes auront été différenciées de manière objective et pertinente par une mesure, il faudra les identifier en fonction de leur qualité absolue. La méthode de mesure précédente permettra alors de différencier celles qui auront été identifiées comme les "bonnes" et les "mauvaises" enceintes.

Pour se donner une chance de trouver un lien entre les approches physique et psychoacous-

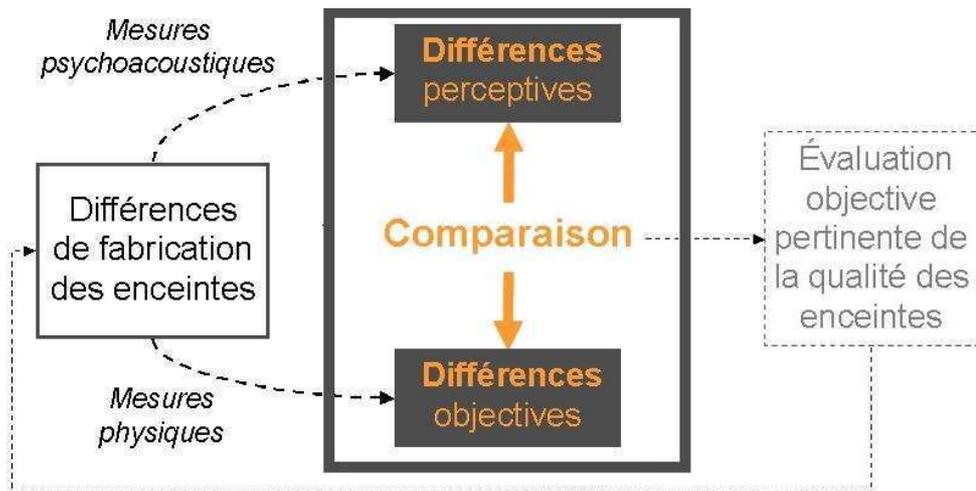


FIGURE II.1 – Approche basée sur l'évaluation des différences relatives entre enceintes acoustiques

tique, nous avons tenté de les garder le plus proche possible l'une de l'autre. En effet, nous avons pu constater dans la littérature la difficulté de relier des mesures physiques aux résultats de tests perceptifs lorsque les deux approches ont été menées séparément. Par exemple, Klippel [10][48] est amené à simuler a posteriori le son produit par les enceintes lors de ses tests d'écoute, en partant de mesures normalisées anéchoïques. À l'inverse, nous avons préféré nous éloigner des mesures normalisées et favoriser le rapprochement des mesures physiques et psychoacoustiques dès le début de l'expérience. Il était donc important de réaliser les mesures physiques dans le même environnement que celui des tests perceptifs, et de préférence en même temps, de façon à s'assurer de mesurer le même champ sonore selon les deux approches. Il n'est pas possible de mesurer le champ sonore à la même position que l'auditeur, pendant que l'auditeur réalise le test d'écoute et par conséquent, comme nos prédécesseurs, nous avons dû chercher un compromis.

II.2 Protocole expérimental

Notre protocole expérimental doit respecter les contraintes inhérentes aux tests d'écoute sur les enceintes acoustiques qui ont été détaillées au chapitre précédent. Les enceintes doivent notamment être comparées à la même position et dans la même pièce. Notre approche étant basée sur l'évaluation des différences relatives entre enceintes, il faut pouvoir les comparer l'une aussitôt après l'autre, du fait de notre mémoire auditive limitée. Notre protocole doit enfin être compatible avec les mesures physiques et psychoacoustiques, qui doivent de préférence pouvoir être réalisées simultanément.

Pour répondre à toutes ces contraintes, notre protocole expérimental consiste à enregistrer le champ sonore rayonné par différentes enceintes acoustiques dans une même pièce. Les enre-

gistrements sont ensuite soumis d'une part à des tests psychoacoustiques réalisés au casque, et d'autre part à des analyses faisant intervenir différentes techniques de traitement de signal. Les phases d'enregistrement et de restitution au casque introduisent des distorsions dans le signal acoustique à transmettre aux oreilles de l'auditeur. Notre protocole ne semble donc pas adapté à l'évaluation des enceintes selon des échelles absolues. Mais les distorsions introduites sont a priori les mêmes pour toutes les enceintes, donc nous pouvons chercher à évaluer les différences relatives entre ces enceintes. Par contre, nous ne prétendons pas pouvoir étudier toutes ces différences, car certaines pourraient être masquées par les distorsions introduites par notre protocole.

Nous ne sommes pas les premiers à utiliser des tests d'écoute au casque pour évaluer la reproduction sonore. Nous avons détaillé au chapitre précédent certains travaux ayant comparé les résultats de tests d'écoute en direct à ceux de tests réalisés au casque [16][44][54][55]. Même si les buts de ces différentes études étaient divers, elles ont toutes montré que les deux configurations donnaient des résultats similaires en ce qui concerne la restitution du timbre.

Nous avons décidé de baser notre approche physique sur l'analyse directe des enregistrements soumis aux tests d'écoute, plutôt que de considérer par exemple des estimations de réponses en fréquence des enceintes. Nous voulions, au moins dans un premier temps, rester aussi proche que possible des signaux entendus et jugés par les auditeurs au cours de l'approche psychoacoustique. Une fois le lien trouvé entre ces deux approches, l'approche physique pourra s'éloigner progressivement, de façon à être moins dépendante du signal et plus caractéristique de l'enceinte mesurée.

L'approche psychoacoustique consiste à recueillir les différences perçues par les auditeurs entre les enregistrements faisant intervenir différentes enceintes. Nous parlerons de dissemblances perceptives pour caractériser ces différences. L'approche physique a pour but de retrouver ces différences par l'analyse des enregistrements. Différentes techniques de traitement de signal sont testées (paragraphe II.5.2, page 45). Il s'agit de déterminer celle qui permet de différencier les enregistrements de la même manière que les auditeurs, c'est-à-dire celle qui permet d'extraire l'information pertinente vis-à-vis de leur perception et qui la "présente" de la manière la plus exploitable. En d'autres termes, nous cherchons la meilleure manière d'observer les signaux pour retrouver les différences perçues par les auditeurs. Par exemple, les domaines temporel et fréquentiel contiennent la même information mais la présentent de deux façons différentes. L'une pourrait s'avérer plus adaptée que l'autre pour mettre en évidence les différences perçues par les auditeurs. Nous parlerons de mesures objectives pour caractériser d'une façon générale nos analyses. En effet, ce ne sont pas toutes des mesures physiques pouvant être reliées à des grandeurs physiques mesurables. Pour chacune des techniques d'analyse testée, nous avons défini une façon d'évaluer les différences entre les enregistrements. Nous parlerons de "métrique" associée à la technique d'analyse. Ces évaluations de différences peuvent vérifier la définition d'une distance, mais ce n'est pas toujours le cas. Nous parlerons donc de dissemblances objectives entre les en-

registrements, par analogie avec les dissemblances perceptives estimées par les auditeurs lors des tests psychoacoustiques.

Afin de déterminer la technique d'analyse et la métrique associée les plus appropriées pour retrouver les différences perçues par les auditeurs, les dissemblances objectives et perceptives sont comparées (Figure II.2). Nous pouvons tout d'abord évaluer la corrélation entre ces dissemblances, c'est-à-dire leur degré de ressemblance "linéaire" (annexe B.1). Les dissemblances ne contiennent pas uniquement de l'information, elles sont également "bruitées". Notre but n'est donc pas d'obtenir une corrélation parfaite mais bien de mettre en évidence la technique d'analyse qui donne le meilleur résultat.

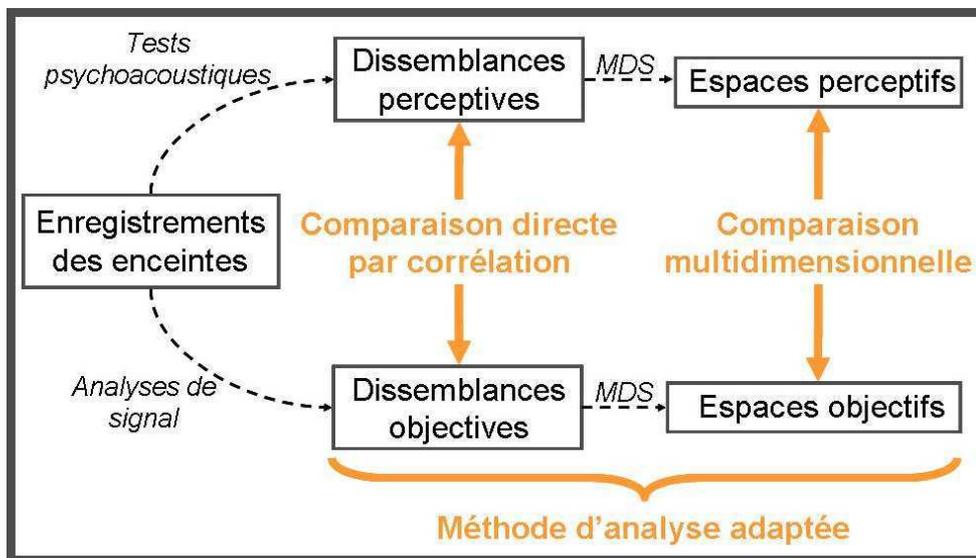


FIGURE II.2 – Le protocole expérimental

Pour compléter cette étude des corrélations entre dissemblances objectives et perceptives, nous pouvons également comparer les espaces obtenus par analyse multidimensionnelle de ces dissemblances (Figure II.2). L'espace perceptif résultant de l'analyse multidimensionnelle des dissemblances perceptives nous donnera accès aux dimensions perceptives principales utilisées par les auditeurs pour différencier les enregistrements lors des tests d'écoute [74]. Les espaces objectifs correspondants sont obtenus en appliquant la même analyse aux dissemblances objectives. La comparaison entre espaces objectifs et perceptifs permet de mieux comprendre les différences d'information contenue dans les dissemblances objectives et perceptives. Ce type de comparaison entre espaces perceptifs et objectifs a notamment permis à Pols *et coll.* [75] de proposer une technique d'analyse appropriée pour étudier la perception des voyelles.

Même si l'espace objectif est différent de l'espace perceptif, et peut ne pas posséder le même nombre de dimensions que celui-ci, certaines dimensions objectives pourraient correspondre à des

dimensions perceptives. Le fait qu'une technique d'analyse permette d'extraire au moins une partie de l'information pertinente vis-à-vis de la perception est déjà une donnée importante. Même si la correspondance entre dissemblances objectives et perceptives n'est que partielle, cela permet d'identifier les apports et limitations de la technique d'analyse considérée. Il convient d'ajouter que la corrélation évalue uniquement le degré de ressemblance linéaire entre les dissemblances. Une relation non-linéaire entre les dissemblances objectives et perceptives pourrait être mieux révélée par la comparaison des espaces résultant de l'analyse multidimensionnelle de ces dissemblances plutôt que par le calcul de leur corrélation. L'espace objectif pourrait également mettre en évidence des dimensions plus fines qui n'ont pas été perçues lors du test perceptif parce que masquées par d'autres dimensions prépondérantes, mais qui seraient révélées par la mesure, et qu'il conviendrait d'explorer par des tests d'écoute complémentaires. L'espace objectif est donc utile pour caractériser l'information extraite par la technique d'analyse et sa métrique associée et apporte un plus par rapport à l'évaluation de la corrélation dans le cas où cette information n'est pas identique à celle contenue dans les dissemblances perceptives : information en moins, en plus, ou information déformée.

II.3 Limitations et avantages du protocole

Des essais préliminaires nous ont confirmé que l'image sonore produite par deux enceintes acoustiques fonctionnant en reproduction stéréophonique est très fortement perturbée si elle est reproduite au casque après avoir été enregistrée. Elle est ressentie comme très "localisée" à l'intérieur de la tête, à tel point qu'il peut devenir risqué de parler encore d'image sonore. Nos essais portaient sur plusieurs techniques de prise de son : prises stéréophoniques AB ORTF, AB NOS et XY, tête artificielle et microphones binauraux. Aucune de ces techniques ne nous a donné de résultats satisfaisants. Ceci provient notamment du fait que la scène sonore à restituer au casque se situe face à l'auditeur, zone où même les prises binaurales ont des performances réduites. L'application de notre protocole est donc limitée à l'étude des capacités de restitution du timbre par les enceintes acoustiques. La composante spatiale de la reproduction sonore ne sera pas considérée. Même si nous ne pouvons pas affirmer que des différences d'image sonore ne peuvent pas influencer les résultats de nos tests, nous ne prétendons pas pouvoir les étudier au moyen de notre protocole. Une étude limitée à la restitution du timbre reste tout de même pertinente étant donnée l'importance de cette composante de la reproduction sonore. Cette importance a été révélée lors des évaluations globales de qualité évoquées au chapitre précédent.

La comparaison au casque limite également notre étude de la perception des très basses fréquences potentiellement reproduites par les enceintes. La perception de ces fréquences peut en effet faire intervenir les vibrations d'autres parties du corps que les tympons, et une écoute au casque n'est pas capable de reproduire ces effets susceptibles d'être perçus lors de l'écoute en

direct. Cette caractéristique de la reproduction sonore par les enceintes ne pourra donc pas être étudié avec le dispositif expérimental proposé et ne fait pas l'objet de cette étude.

Nous n'avons pas utilisé d'enceintes pour reproduire nos enregistrements lors des tests d'écoute. Reproduire les enregistrements au casque fournit un meilleur contrôle de la position des oreilles de l'auditeur [44]. La reproduction d'enregistrements stéréophoniques sur deux enceintes en salle anéchoïque aurait nécessité l'utilisation de traitement du signal pour annuler la transmission de l'enceinte droite vers l'oreille gauche et de l'enceinte gauche vers l'oreille droite ("crosstalk"). "It seemed to be imprudent to investigate loudspeaker performance through other loudspeakers which have been subject to elaborate signal processing" Toole [44]. Ceci aurait peut être pu permettre de considérer l'influence des très basses fréquences qui ne sont pas reproduites au casque, mais en même temps une chambre anéchoïque devient moins anéchoïque dans cette gamme de fréquence. Son influence s'ajouterait donc à celle de la salle d'écoute dans laquelle les enregistrements ont été réalisés. Nous avons préféré limiter au maximum les traitements effectués sur les signaux entre les enregistrements et la reproduction lors des tests d'écoute.

Nous avons vu au chapitre précédent que des différences de niveau de reproduction sonore sont bien sûr perçues par les auditeurs, et ceci de façon prépondérante. Cette dimension de la reproduction sonore présente pourtant peu d'intérêt. Afin d'éviter que des différences de niveau sonore masquent des différences plus fines, la sonie des enregistrements doit être égalisée avant les tests d'écoute. Nous devons également éviter au maximum que des différences de niveau de bruit de fond liées à l'enregistrement n'interviennent dans les jugements des auditeurs. Par conséquent, le niveau sonore de la reproduction lors des sessions d'enregistrements doit être le même pour toutes les enceintes, afin de s'assurer que le rapport signal/bruit soit le même pour tous les enregistrements, ce qui évite ensuite les différences de niveau de bruit de fond lors de leur égalisation de sonie. Le niveau sonore commun utilisé lors des enregistrements ne correspond pas forcément pour toutes les enceintes à leur niveau de sollicitation nominal, niveau de fonctionnement optimal recommandé par le fabricant. Cette limitation n'est cependant pas propre à notre protocole car tous les tests d'écoute sur des enceintes acoustiques nécessitent une égalisation du niveau sonore de la reproduction. Lors d'un test d'écoute, c'est l'enceinte à un niveau de sollicitation donné qui est évaluée. La dynamique d'une enceinte, c'est-à-dire sa capacité à répondre de la même façon pour les forts et faibles niveaux, constitue une autre composante de la reproduction sonore qui ne sera pas étudiée ici, même si nous verrons au chapitre V qu'une telle étude peut être envisagée à l'aide de notre protocole.

Les chaînes d'enregistrement et de restitution au casque modifient le signal délivré par les enceintes acoustiques dans la pièce. De l'information est perdue au cours du processus et notamment certaines différences entre les enceintes peuvent disparaître. Néanmoins, conscients de ces limitations, nous avons décidé d'appliquer ce protocole expérimental qui présente des avantages importants pour étudier l'information restante. D'une part, les enregistrements se prêtent

aussi bien aux tests psychoacoustiques qu'aux analyses objectives. Ces deux approches que nous cherchons à relier sont ainsi gardées très proches. D'autre part, lors de la comparaison des enregistrements, toute différence évaluée provient forcément des enceintes ou de leur interaction avec la salle, car toutes les enceintes sont évaluées dans les mêmes conditions. Ce protocole permet également de ne pas limiter le nombre d'enceintes incluses dans la même expérience, tout en autorisant leur comparaison très rapide pour une même position dans la salle d'écoute.

Les limitations et les avantages de notre protocole expérimental constituent le compromis que nous avons choisi de réaliser afin d'étudier le lien entre nos évaluations objectives et perceptives d'enceintes acoustiques.

II.4 Dissemblances perceptives

L'approche psychoacoustique a pour but d'évaluer les dissemblances perçues entre les enregistrements faisant intervenir les différentes enceintes, en réalisant au casque des tests d'écoute impliquant un nombre suffisant d'auditeurs. L'évaluation perceptive ou psychoacoustique est parfois qualifiée d'évaluation subjective. Le terme est correct en ce sens qu'elle nécessite d'avoir recours à des "sujets", mais il peut porter à confusion en laissant penser que l'évaluation est dépendante de ces sujets, ce qui n'est pas le cas si le test est contrôlé. Comme nous l'avons évoqué au début de ce chapitre, nous avons évité les jugements de qualité ou de préférence de façon à s'affranchir le plus possible du goût des auditeurs et à se concentrer sur leur perception auditive. En choisissant une tâche la plus simple possible pour nos tests, nous avons limité la dépendance vis-à-vis du niveau d'expérience des auditeurs. Enfin, pour chacun des tests, après avoir montré qu'il n'y avait pas de groupe d'auditeurs ayant des stratégies de réponse différentes, nous avons pris la moyenne arithmétique de tous les jugements de dissemblances. Le nombre d'auditeurs est considéré suffisant lorsque cette moyenne se stabilise. Le résultat de l'évaluation correspond alors aux jugements d'un auditeur "moyen".

Les tests d'écoute ont été menés dans une cabine audiométrique, en utilisant un casque électrostatique Stax SR Lambda Professional. Davantage de détails sur le matériel utilisé peuvent être trouvés dans l'annexe A.6. Les tests étaient suivis d'entretiens informels avec les auditeurs, afin de recueillir leurs impressions. Il leur était notamment demandé de décrire autant que possible les critères sur lesquels étaient basés leurs jugements de dissemblances.

Afin de vérifier la présence ou non de classes aux stratégies de réponses différentes parmi les auditeurs, une analyse en arbre hiérarchique a été utilisée [51][76][77]. Cette analyse permet de tracer un dendrogramme à partir d'une matrice de dissemblances, afin d'en étudier la structure (annexe B.3). La dissemblance entre deux auditeurs s'évalue différemment selon le type de test psychoacoustique considéré.

Deux principes de test ont été mis en place pour évaluer les dissemblances perceptives entre

les enregistrements, le principe utilisé dépendant du nombre d'enregistrements impliqués dans le test.

II.4.1 Comparaison par paires

La méthode la plus directe pour obtenir les jugements de dissemblances est de réaliser un test de comparaison par paires [24][40][78]. Les enregistrements sont présentés par paires à l'auditeur qui doit évaluer directement la dissemblance entre les deux enregistrements de chaque paire. L'ordre de présentation des paires est aléatoire et toutes les paires sont présentées à l'auditeur. Lors de nos tests, l'auditeur quantifiait la différence qu'il percevait entre les enregistrements en déplaçant un curseur sur une ligne dont les extrémités s'intitulaient "très semblables" et "très dissemblables". Une valeur de dissemblance est assignée de façon linéaire à chacun des points de cette ligne, si bien qu'une matrice de dissemblances contenant les jugements de toutes les paires est obtenue en fin de test. Il est à noter qu'un stimulus n'est jamais comparé à lui-même et que l'auditeur ne juge qu'une seule fois chaque paire de stimuli, l'ordre des stimuli dans la paire étant tiré aléatoirement pour chaque auditeur. La matrice de dissemblances obtenue est donc triangulaire. Elle est par la suite symétrisée pour permettre l'analyse.

La matrice de dissemblances finale est obtenue en prenant la moyenne arithmétique de l'ensemble des matrices de dissemblances fournies par les auditeurs. Aucune normalisation n'est appliquée sur ces matrices individuelles avant d'effectuer la moyenne. Afin de s'assurer que les auditeurs utilisent tous à peu près la même échelle, il leur est demandé d'utiliser toute l'échelle qui leur est présentée. Une écoute préalable des stimuli impliqués dans le test est réalisée avant de démarrer l'expérience, et quelques paires sont ajoutées à titre d'entraînement au début du test, pour que l'auditeur se familiarise avec l'échelle d'évaluation des dissemblances.

La dissemblance entre deux auditeurs a été évaluée à partir de la corrélation de leurs jugements. Cette dissemblance vaut $1-C$, où C est le coefficient de corrélation entre les évaluations de dissemblances fournies par les deux auditeurs.

L'utilisation de ce type de test est cependant restreinte aux expériences ne devant inclure qu'un nombre limité d'enregistrements. En effet, N enregistrements conduisent à $N(N-1)/2$ paires à juger. Quand le nombre d'enregistrements augmente, le nombre de paires devient vite trop important et le test trop long pour que les auditeurs puissent réaliser la tâche dans des conditions convenables [30].

II.4.2 Classification libre

Une méthode indirecte, basée sur une tâche de classification libre [49][76][77] a été mise en place lorsque nous avons souhaité utiliser un nombre important d'enregistrements. Les enregistrements ont été présentés à l'auditeur sous forme de croix réparties aléatoirement sur un écran.

L'auditeur pouvait bouger chaque croix librement et écouter l'enregistrement correspondant autant de fois qu'il le souhaitait, simplement en cliquant sur la croix. La tâche consistait à grouper les enregistrements en fonction de leur ressemblance : des enregistrements jugés similaires devaient être placés dans le même groupe alors que des enregistrements jugés différents devaient se trouver dans des groupes différents. Le nombre de groupes à utiliser était libre, et donc le degré de similarité entre deux enregistrements pour qu'ils soient placés dans le même groupe était laissé au jugement de l'auditeur. Le groupement ainsi obtenu est ensuite transformé en une matrice de dissemblances entre tous les enregistrements. La dissemblance entre deux enregistrements provenant du même groupe est fixée à 0. Elle est fixée à 1 si les enregistrements sont placés dans des groupes différents. La matrice de dissemblances finale est obtenue en prenant la moyenne arithmétique de l'ensemble des matrices de dissemblances fournies par les auditeurs. Plus deux enregistrements ont été groupés souvent ensemble, plus ils sont supposés être similaires, et plus la valeur de leur dissemblance obtenue est faible.

Les différences de jugement entre auditeurs peuvent provenir de différences de classement des enregistrements, mais aussi du nombre de classes qu'ils ont utilisées. La dissemblance entre deux auditeurs ne doit pas être évaluée directement par la corrélation de leurs réponses. Il s'agit d'évaluer la correspondance entre différentes partitions d'un même ensemble d'éléments. L'indice de Rand est alors employé à la place du coefficient de corrélation (annexe B.2).

II.4.3 Validité de l'expérience psychoacoustique

"Une tâche perceptive ne se réduit pas simplement à une méthode expérimentale. C'est tout d'abord une problématique, puis une méthode expérimentale suivie d'une méthode d'analyse statistique des résultats et enfin une interprétation objective des résultats. Chacune de ces composantes joue un rôle essentiel dans la validité du travail et chaque étape doit référer de manière logique à la problématique" Guyot [49]. Il est en effet primordial de s'interroger sur la validité de notre expérience vis-à-vis de ce que nous cherchons à étudier. Certains parlent alors d'étudier la validité écologique du protocole [76][77]. Est-ce que notre protocole nous permet bien d'étudier la perception de la restitution du timbre par les enceintes acoustiques ou bien est ce que nous manquons notre but ? Quels types de résultats sont recherchés et quels outils statistiques sont utilisés pour cela ? Est ce que ces outils sont valides dans le cadre de notre étude ? Les questions relatives à notre protocole expérimental seront abordées dans un premier temps, celles relatives aux outils d'analyse utilisés seront traitées ensuite.

II.4.3.a Validité de l'estimation des dissemblances perceptives

Dans le cadre de l'étude de la perception sonore, l'étude de la perception de la reproduction sonore est particulière, et ses particularités facilitent l'utilisation de notre protocole expérimental.

tal. Les stimuli sont des extraits musicaux, qui font intervenir une écoute musicale telle que l'a définie Gaver [79]. Ce type d'écoute, basé sur les sons, peut être différent de l'écoute dite quotidienne [79] reposant sur la reconnaissance des événements sonores de notre environnement [74]. Même s'ils ont pour but de susciter des émotions, les sons musicaux ne portent pas de signification particulière comme cela peut être le cas par exemple du son d'une alarme auquel pourront être associés les informations de danger ou d'avertissement. Il faut également ajouter que par rapport aux études sur le timbre des instruments de musique [51][74][80], la comparaison de nos stimuli ne fait pas intervenir de reconnaissance des instruments car, pour nos enregistrements, le même extrait musical est utilisé pour toutes les enceintes. Les études portant sur des sons de l'environnement entraînant une tâche de reconnaissance de la source par l'auditeur sont beaucoup plus sensibles au contexte de l'expérience, car la perception de ces sons peut dépendre de l'activité à laquelle ils sont associés et donc du contexte de l'écoute [74]. Il apparaît parfois délicat de généraliser des résultats obtenus sur de tels sons dans des conditions de laboratoire si celles-ci sont trop éloignées des conditions d'écoute réelle de la vie de tous les jours. La musique n'est pas associée à une autre activité que celle de l'écouter pour le plaisir de l'entendre. L'écoute musicale en laboratoire n'apparaît donc pas très éloignée de son contexte naturel.

Nos tests s'inscrivent dans ce contexte de l'écoute de la musique. Ils font appel à une activité identique à celle que nous cherchons à évaluer qui est l'écoute du son produit par des enceintes acoustiques dans un salon. La reproduction sonore est dès le départ un artifice, c'est la reproduction artificielle de musiciens jouant en direct. Notre but est bien d'étudier cette reproduction artificielle et pas de savoir si cette reproduction "imite" bien la réalité d'un concert par exemple. Le contexte dont nous prétendons ne pas nous éloigner en laboratoire est celui d'une écoute musicale artificielle et pas celui d'un concert en direct. Nous émettons tout de même l'hypothèse que l'écoute musicale au casque est une activité très proche de l'écoute musicale sur des enceintes acoustiques. Nous supposons notamment que l'écoute au casque et l'écoute sur des enceintes font intervenir les mêmes dimensions perceptives associées à la restitution du timbre. Cette hypothèse est supportée par les travaux de Gabrielsson et Sjögren [7] évoqués au chapitre précédent. Si nous mettons en évidence des dimensions perceptives utilisées par les auditeurs pour différencier nos enregistrements, elles ne seront pas décelées seulement du fait du recours à l'écoute au casque. Nous faisons l'hypothèse qu'elles seraient également intervenues lors d'une écoute en direct.

Les tâches perceptives employées lors de nos tests psychoacoustiques ont pour but d'évaluer les dissemblances perçues par les auditeurs, en se basant sur des comparaisons par paires ou une classification libre. Ces tâches nous semblent être simples, naturelles et libres pour l'auditeur, et par conséquent peu susceptibles de modifier son écoute [49][77]. L'influence de la tâche perceptive dans le cadre de notre étude sera considérée par la suite, au même titre que l'influence des autres paramètres expérimentaux.

Notre protocole semble donc a priori valide pour étudier la restitution du timbre par les

enceintes acoustiques. Nous prendrons soin, tout au long de notre étude, d'évaluer l'influence des différents paramètres de ce protocole expérimental.

II.4.3.b Validité de l'analyse des dissemblances perceptives

Les dissemblances perceptives obtenues à l'aide de nos tests psychoacoustiques seront utilisées de deux façons. Elles seront d'une part comparées directement aux dissemblances objectives résultants des différentes techniques de traitement de signal appliquées aux enregistrements. D'autre part, elles seront soumises à une analyse multidimensionnelle (annexe B.5). Ceci nous permettra notamment d'explorer les dimensions perceptives sous-jacentes aux estimations des auditeurs.

L'analyse multidimensionnelle peut être appliquée à des dissemblances résultants aussi bien d'un test de comparaison par paires que d'une tâche de classification libre [39][76][81]. C'est au moment de l'interprétation de l'espace correspondant qu'il convient d'être particulièrement prudent. L'analyse multidimensionnelle représente les dissemblances perceptives par des distances entre des points dans un espace aux dimensions continues (annexe B.5). Malgré tout, rien ne prouve que ces dimensions correspondent à des dimensions perceptives qui permettent de passer de façon continue d'un stimuli à l'autre. L'analyse multidimensionnelle n'est qu'un mode de représentation des données de dissemblance, c'est en interprétant cet espace que l'expérimentateur fait des hypothèses sur ses données. L'hypothèse de l'existence de dimensions perceptives continues n'est en rien justifiée par l'analyse multidimensionnelle elle-même. Si nous prenons l'exemple d'une évaluation de dissemblances entre différentes sortes de pommes de terre et de carottes, il se peut que l'analyse multidimensionnelle des dissemblances représente deux groupes de points, chacun correspondant à un des deux légumes. Rien ne prouvera pour autant qu'il existe une dimension perceptive continue qui permette de passer d'une pomme de terre à une carotte. Ce sera à l'expérimentateur de justifier l'hypothèse de continuité.

La possibilité d'interpréter les dimensions résultant de l'analyse multidimensionnelle comme des dimensions perceptives continues semble dépendre fortement de la nature des stimuli impliqués dans le test. Susini [24] relate une étude portant sur des sons de l'environnement très hétérogènes qui a ainsi conduit à un espace perceptif marqué par des catégories très nettes, associées notamment à l'identification de la source sonore par les auditeurs. Les jugements de dissemblances apparaissaient alors liés à la catégorie d'appartenance des stimuli plutôt qu'à leurs propriétés individuelles. L'hypothèse de dimensions perceptives continues partagées par les stimuli fut alors rejetée. Dans notre cas impliquant un corpus de stimuli suffisamment homogènes, qui ne font a priori pas intervenir de tâche d'identification, il semble légitime de chercher s'il existe des dimensions perceptives continues permettant de différencier les stimuli [24].

Nos analyses multidimensionnelles sont précédées d'analyses en arbre hiérarchique (annexe B.3),

qui permettent de tester l'homogénéité de nos stimuli en mettant ou non en évidence la présence de catégories parmi les enregistrements, à partir de leur matrice de dissemblances résultant des tests psychoacoustiques [24]. En l'absence de groupements nets entre les enregistrements, les dimensions continues révélées par l'analyse multidimensionnelle seront jugées pertinentes pour décrire la perception des auditeurs.

La nature de la tâche perceptive utilisée pour obtenir les dissemblances ne semble pas être un critère approprié pour définir si l'hypothèse de continuité des dimensions révélées par l'analyse multidimensionnelle est adaptée. L'évaluation directe des dissemblances utilisant la comparaison par paires sur des stimuli très hétérogènes conduira à la formation de catégories. Dans une tâche de classification libre sur des stimuli homogènes ne comportant pas de réelles catégories, la nature des groupements réalisés par les auditeurs dépendra du seuil de similarité qu'ils ont utilisé pour placer les stimuli dans des groupes différents. Ce seuil pourra varier d'un auditeur à l'autre. Ainsi, les dissemblances moyennes ne feront pas forcément apparaître de catégories. C'est donc bien la nature des stimuli plutôt que celle de la tâche perceptive qui semble déterminante vis-à-vis de la structure finale des données de dissemblances.

La tâche de classification libre présente néanmoins une résolution beaucoup plus faible que la tâche de comparaison par paires. La dissemblance évaluée par un auditeur vaut 1 ou 0 dans le premier cas, tandis qu'elle peut prendre un nombre considérable de valeurs dans le second cas. Il faudra donc s'assurer d'impliquer un nombre beaucoup plus important d'auditeurs pour la tâche de classification, afin que les matrices binaires de chaque auditeur soient suffisamment moyennées. Chaque auditeur fournit moins d'information — mais dans des tests portant sur un nombre plus important d'enceintes — il faudra donc davantage d'auditeurs pour obtenir suffisamment d'information.

II.5 Dissemblances objectives "classiques"

L'approche objective a pour but de retrouver les dissemblances perçues par les auditeurs lors des tests psychoacoustiques, par l'analyse des enregistrements ayant été utilisés lors de ces tests. Différentes techniques de traitement de signal ont été testées et pour chacune de ces techniques, nous avons défini une "métrique" associée permettant d'évaluer les dissemblances objectives entre les enregistrements. Nos analyses s'appliquent à des signaux monophoniques et dans le cas d'enregistrements stéréophoniques, les voies droite et gauche sont analysées indépendamment.

Avant de procéder aux analyses des signaux à comparer, ils ont dû être synchronisés les uns par rapport aux autres. En effet, de la même façon que la sonie des enregistrements doit être égalisée pour éviter d'introduire des dissemblances sans intérêt, nous avons voulu éviter qu'un décalage de quelques échantillons entre deux signaux à comparer ne se traduise par une différence artificiellement grande au niveau de leurs analyses objectives. Ces faibles décalages temporels,

bien que cruciaux lors de l'analyse des signaux, restent imperceptibles lors des tests psychoacoustiques. Les analyses objectives doivent imiter ce comportement et le décalage temporel ne doit pas créer de dissemblances objectives non pertinentes. Les enregistrements doivent donc être synchronisés afin d'éliminer cette information non perçue.

II.5.1 Synchronisation des signaux enregistrés

Différentes méthodes permettant d'évaluer le décalage temporel entre deux signaux audio ont été testées (annexe D). Notre but était de déterminer les décalages temporels entre les enregistrements devant être comparés, afin de les synchroniser ensuite à l'échantillon près. Cette détermination n'est pas une tâche triviale et est sensible à la gamme de fréquence considérée. Elle semble donc dépendante des réponses en fréquence des enceintes impliquées dans les enregistrements à synchroniser. La réverbération et les premières réflexions dans la pièce lors des enregistrements compliquent encore davantage la tâche.

Aucune des méthodes testées n'a donné de résultats parfaits, et nous avons rencontrés des incohérences du type : le décalage entre les signaux A et C n'est pas égal au décalage entre A et B plus le décalage entre B et C. Deux possibilités s'offraient alors à nous. La première consistait à utiliser un signal comme référence pour synchroniser tous les signaux par rapport à cette unique référence. Mais, nous n'étions alors en aucun cas certains que ces signaux seraient tous synchronisés les uns par rapport aux autres. La deuxième possibilité, que nous avons mise en oeuvre, ne cherche pas à réaliser cette synchronisation en une fois sur l'ensemble des signaux avant leur analyse. Nous avons en effet considéré les signaux par paires. Le décalage entre les deux signaux de la paire a été calculé en minimisant la distance quadratique entre leurs deux formes temporelles. La synchronisation a ainsi été effectuée au sein de chaque paire, indépendamment des synchronisations faites pour les autres paires. Les dissemblances objectives sont ensuite évaluées au sein de chaque paire, sur les signaux dont le décalage temporel a été minimisé. Cette minimisation est réalisée à un échantillon près, la résolution étant limitée par le caractère discret de nos signaux numériques échantillonnés.

II.5.2 Techniques d'analyse "classiques" considérées

Des analyses ont été menées dans les domaines temporel, fréquentiel et temps-fréquence. Deux pondérations fréquentielles ont également été testées.

Dans un premier temps, les dissemblances objectives sont évaluées en tout point du domaine considéré, par exemple à chaque instant pour le domaine temporel ou à chaque fréquence pour le domaine fréquentiel. Ces dissemblances sont ensuite intégrées sur tout le domaine pour ne fournir qu'une seule valeur de dissemblance globale entre les deux signaux comparés. L'intégration temporelle est réalisée en moyennant la valeur instantanée de la dissemblance sur l'ensemble

de l'extrait, ce qui rend cette dissemblance indépendante de la longueur de signal sur laquelle elle est déterminée. Prendre la moyenne arithmétique plutôt que d'effectuer une réelle intégration n'entraîne qu'une différence de normalisation. L'intégration fréquentielle est effectuée en considérant la moyenne arithmétique des dissemblances sur l'ensemble du domaine fréquentiel. La dissemblance globale ainsi définie est alors indépendante de la gamme de fréquences sur laquelle elle a été calculée. Nous avons choisi d'effectuer ces intégrations pour obtenir une dissemblance objective globale pour chaque paire de signaux, par analogie à l'évaluation de la dissemblance perceptive globale réalisée pour chacune de ces paires par les auditeurs durant les tests psychoacoustiques. Comme nous ne pouvons pas supposer qu'un instant soit plus important qu'un autre lors de cette évaluation, ou qu'une gamme de fréquence soit privilégiée, nous avons choisi une moyenne arithmétique donnant le même poids à tous les points.

Les dissemblances étant évaluées en chaque point du domaine d'analyse considéré, il se peut qu'elles soit calculées en des points où il n'y aurait pas de signal mais seulement du bruit de fond. Les dissemblances présentées ici ne font intervenir que des différences de signaux, et pas de rapport. Le bruit de fond est de faible amplitude par rapport au signal, donc une différence de bruit de fond sera également de faible amplitude. Les dissemblances qui seraient déterminées sur du bruit de fond resteront donc de faible amplitude par rapport aux dissemblances calculées sur le signal utile, et la dissemblance globale entre les deux signaux ne devrait être que faiblement influencée. Le problème serait plus critique dans le cas de dissemblances définies à partir de rapports. Des rapports de bruit de fond pourrait en effet conduire à des valeurs élevées de dissemblances qui ne seraient pas pertinentes.

Les différentes techniques d'analyse utilisées sont décrites dans ce qui suit. Des renseignements complémentaires concernant le détail des paramètres d'analyse employés peuvent être trouvés dans l'annexe E.

Les dissemblances objectives "classiques" que nous avons définies et utilisées lors de notre étude sont toutes homogènes au carré des grandeurs comparées. Des recherches sont en cours afin de tester d'autres types de dissemblances, elles sont évoquées au chapitre V.

II.5.2.a Domaine temporel

Dans le **domaine temporel**, la dissemblance entre deux signaux est évaluée en calculant la distance quadratique entre les deux formes temporelles, c'est-à-dire en évaluant le carré de la différence entre les deux signaux. Après l'intégration temporelle, nous avons en fait conservé la valeur minimum des résultats obtenus à partir du carré de la différence et du carré de la somme des signaux, dans l'éventualité de deux signaux en opposition de phase.

$$Diss_{temps}\{x(t),y(t)\} = Min\{\langle [x(t) - y(t)]^2 \rangle_t, \langle [x(t) + y(t)]^2 \rangle_t\}$$

II.5.2.b Domaine fréquentiel

Plusieurs dissemblances ont été définies dans le domaine fréquentiel. La première est calculée sur les **spectres** des deux signaux en déterminant, comme dans le domaine temporel, le minimum entre l'intégration fréquentielle du carré de la différence des deux spectres et celle du carré de leur somme. La deuxième dissemblance ne prend en compte que le **module des spectres**, elle ne fait pas intervenir leur phase. Nous la calculons en moyennant le carré de la différence des modules sur l'ensemble du spectre. Dans les deux cas, le spectre des signaux est évalué par l'intermédiaire de leur transformée de Fourier discrète (TFD).

$$Diss_{spectre}\{x(t),y(t)\} = Min\{\langle\|TFD(x) - TFD(y)\|^2\rangle_f, \langle\|TFD(x) + TFD(y)\|^2\rangle_f\}$$

et

$$Diss_{module_spectre}\{x(t),y(t)\} = \langle\|TFD(x)\| - \|TFD(y)\|\rangle_f^2$$

Nous avons également défini une dissemblance en considérant la **densité spectrale de puissance** (DSP) dont le calcul est précisé dans l'annexe E.1. La dissemblance est calculée en prenant la moyenne fréquentielle du carré de la différence des racines carrées des densités spectrales de puissance des signaux. Le calcul de cette dissemblance ne s'est pas avéré sensible à la résolution fréquentielle utilisée pour évaluer la densité spectrale de puissance.

Cette méthode d'évaluation de la dissemblance entre deux densités spectrales de puissance n'apparaît pas être la solution la plus simple. Nous aurions pu intégrer plus directement la valeur absolue de la différence des densités spectrales. Nous voulions ici définir la dissemblance de façon parallèle à celle que nous avons calculée sur le module du spectre, afin de pouvoir les comparer.

$$Diss_{PSD}\{x(t),y(t)\} = \langle[\sqrt{DSP(x)} - \sqrt{DSP(y)}]^2\rangle_f$$

II.5.2.c Plan temps-fréquence

Des dissemblances objectives sont également évaluées dans le plan temps-fréquence, en considérant la transformée de Fourier à court terme (TFCT) de nos signaux. Le calcul de cette transformée et les différents paramètres utilisés sont décrits dans l'annexe E.2. La première dissemblance définie sur ces transformées **temps-fréquence** est analogue à celle définie sur le spectre dans le domaine fréquentiel. Les dissemblances instantanées sont évaluées sur les spectres obtenus aux différents instants, puis ces dissemblances sont moyennées sur la durée totale de l'enregistrement. La dissemblance globale correspond au minimum entre le carré de la différence des deux transformées de Fourier à court terme et le carré de leur somme, après intégrations fréquentielle et temporelle de ces carrés. Une deuxième dissemblance ne prenant en compte que

le **module des transformées temps-fréquence** a également été testée. Nous la calculons en prenant les moyennes fréquentielles et temporelles du carré de la différence des modules.

$$Diss_{TFCT}\{x(t),y(t)\} = Min\{\langle\langle\|TFCT(x) - TFCT(y)\|^2\rangle_f\rangle_t, \langle\langle\|TFCT(x) + TFCT(y)\|^2\rangle_f\rangle_t\}$$

et

$$Diss_{module_TFCT}\{x(t),y(t)\} = \langle\langle\|\|TFCT(x)\| - \|TFCT(y)\|\|^2\rangle_f\rangle_t$$

II.5.2.d Domaine fréquentiel avec pondérations

A partir du domaine fréquentiel, nous avons considéré des domaines d'analyse dérivés, reposants sur l'application de pondérations spectrales devant rendre compte de la sensibilité de l'oreille. Deux pondérations ont été testées. La première est la **pondération "A"** couramment utilisée en acoustique et obtenue à partir de la ligne isotonique à 40 phones. La deuxième est une **pondération basée sur la courbe isotonique à 70 phones** qui correspond approximativement au niveau perçu de la reproduction des enregistrements lors de nos tests psychoacoustiques. Les détails concernant les courbes isotoniques peuvent être trouvés dans [82]. Les pondérations ont été appliquées sur les densités spectrales de puissance des signaux. Les dissemblances ont ensuite été évaluées de la même façon que celle définie sur la densité spectrale de puissance non pondérée évoquée précédemment.

$$Diss_{Pond}\{x(t),y(t)\} = \langle[\sqrt{DSP_{Pond}(x)} - \sqrt{DSP_{Pond}(y)}]^2\rangle_f$$

II.6 Utilisation de modèles auditifs : dissemblances "perceptives objectives"

Les méthodes d'analyse classiques décrites jusqu'ici ont l'avantage d'être objectives, mais à notre connaissance, elles ne permettent pas de faire facilement le lien avec les mesures perceptives sur les enceintes, qui sont elles pertinentes, mais difficiles à mettre en place.

L'utilisation de pondérations spectrales cherche à prendre en compte la sensibilité de l'oreille afin de se rapprocher de ce qui est perçu par les auditeurs, mais nous restons encore très loin de ce qui est réellement perçu. Le fonctionnement complexe du système auditif modifie beaucoup le spectre des signaux écoutés, et le "spectre" perçu est bien différent de celui qui peut être mesuré à l'entrée du système auditif [29].

La Figure II.3 illustre cette différence en présentant les spectres en tiers d'octave de trois signaux musicaux et deux évaluations des spectres perçus correspondants. Les signaux musicaux considérés sont ceux qui seront utilisés par la suite dans nos expériences. Ils sont décrits plus

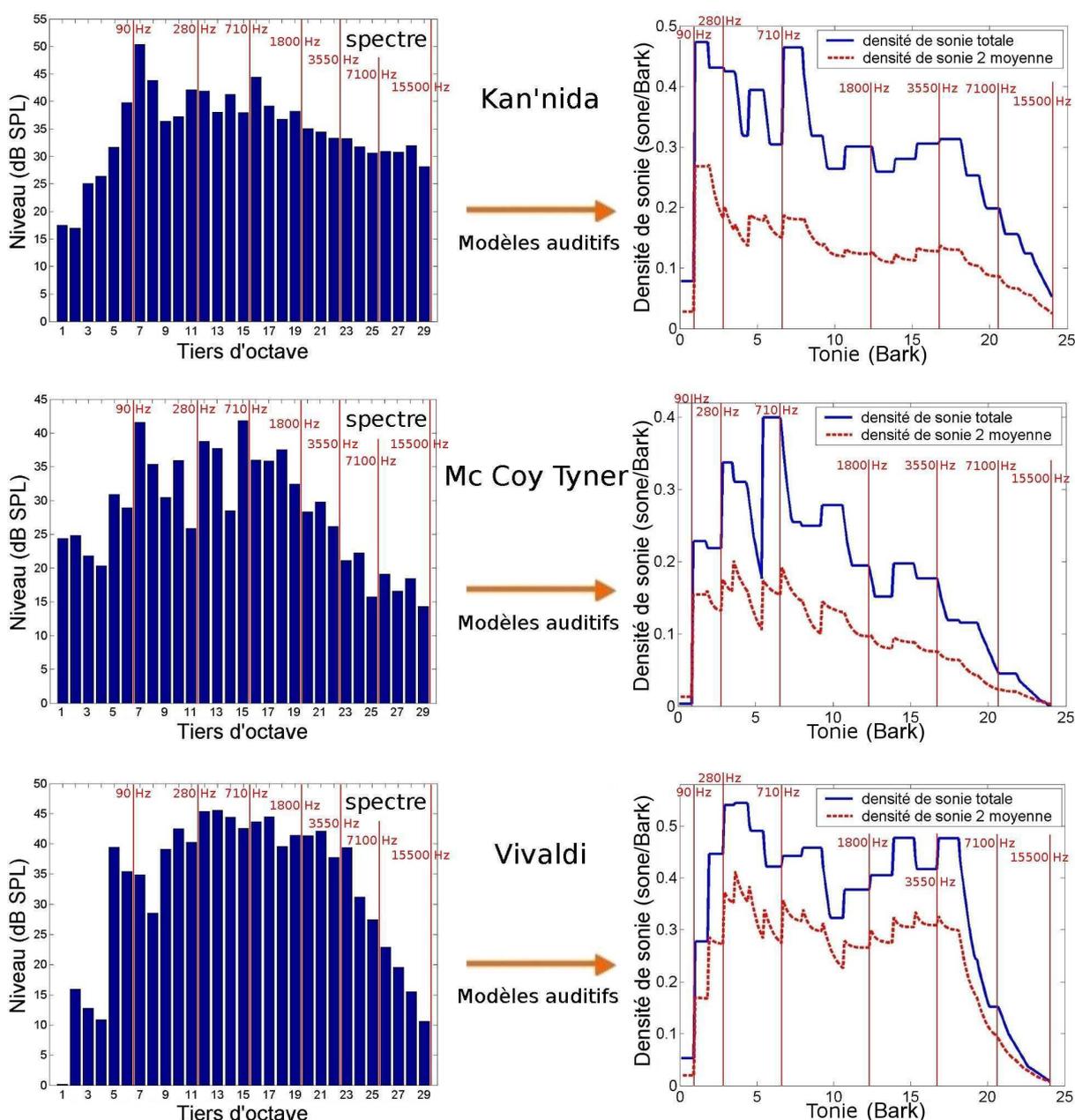


FIGURE II.3 – Estimations des modifications du spectre par le système auditif pour trois extraits musicaux

précisément dans le prochain chapitre. Les modèles auditifs utilisés pour tracer ces évaluations de spectre perçu ont été développés par Zwicker *et coll.* [61][62][65][66][83]. Ils sont décrits plus en détails dans la suite de ce chapitre, ainsi que dans l'annexe E.3. Ils permettent d'évaluer la sonie spécifique ou densité de sonie de nos signaux. La sonie correspond à l'intensité sonore perçue et la sonie spécifique à la densité de sonie suivant l'échelle de tonie ou hauteur tonale, c'est-à-dire l'échelle de fréquence perçue. L'échelle de tonie est modélisée par l'échelle des Barks dans les

programmes que nous avons utilisés. Les limites fréquentielles des tiers d'octave et des bandes de Bark sont précisées en annexe E.3. Des repères ont été placés sur la Figure II.3 pour faire la correspondance entre ces deux échelles fréquentielles, et faciliter la comparaison des graphiques présentés. Les différences entre densité de sonie totale et densité de sonie moyenne sont détaillées plus tard dans ce chapitre. Ce sont deux évaluations différentes du spectre perçu. Le fonctionnement du système auditif, et notamment celui de la cochlée, entraîne un étalement spectral et temporel du signal, qui dépend des caractéristiques de ce même signal. Cet étalement est responsable des différents phénomènes de masquage auditif [65][66][67], et entraîne les déformations du spectre mises en évidence sur la Figure II.3. Par exemple, pour l'extrait Mc Coy Tyner, le pic d'énergie dans le septième tiers d'octave, après 90 Hz, n'apparaît plus sur les évaluations du spectre perçu. Le niveau important des aigus pour l'extrait Kan'nida, entre 7100 et 15500 Hz, est lui aussi très atténué lorsque les modèles auditifs sont considérés.

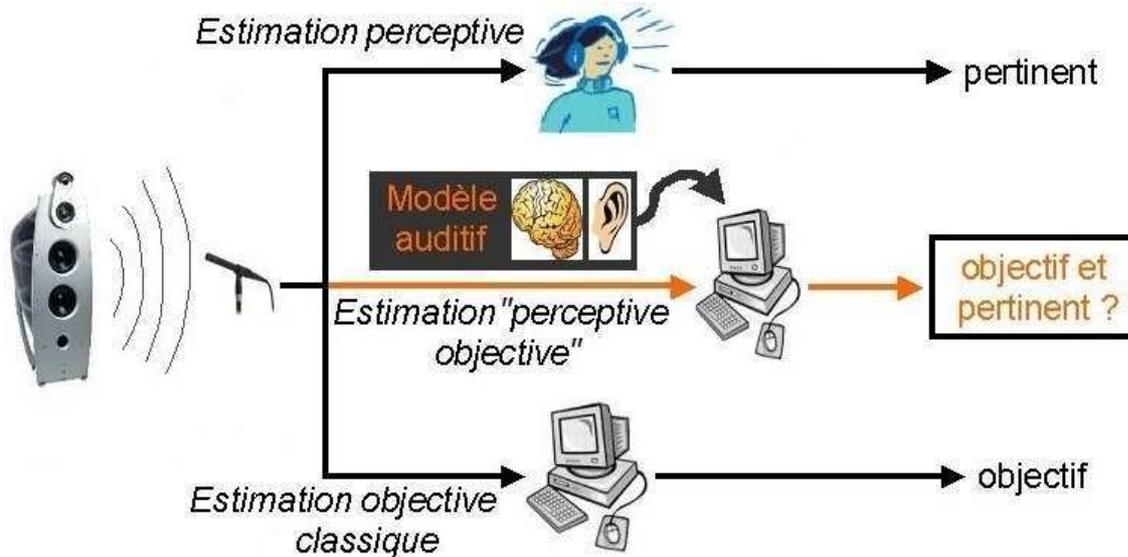


FIGURE II.4 – *Évaluation de dissemblances perceptives objectives par l'utilisation de modèles auditifs*

En parallèle des mesures perceptives et des mesures objectives classiques, nous avons décidé de tester l'utilisation de modèles auditifs (Figure II.4). En faisant appel à une connaissance a priori pour analyser les signaux, cette troisième voie nous a permis d'estimer des dissemblances "perceptives objectives" entre nos enregistrements — perceptives du fait du recours au modèle auditif, et objectives car les auditeurs et la mise en place de test d'écoute ne sont plus nécessaires. Cette idée d'utiliser des modèles auditifs dans le cadre de l'évaluation de la reproduction sonore n'est pas nouvelle [9][10][26][43][48][63][69][70][71]. Staffeldt [26][43][63] a notamment appliqué un modèle auditif à ses mesures de réponses impulsionnelles d'enceintes pour relier plus facilement

la forme des courbes de réponses obtenues aux résultats de ses tests d'écoute. Klippel [10][48] a considéré la densité de sonie de ses signaux pour définir des attributs objectifs lui permettant d'expliquer les attributs perceptifs révélés par ses tests psychoacoustiques. Le modèle développé par Bramsløw [9] pour évaluer la qualité des prothèses auditives est lui aussi basé sur la densité de sonie des signaux analysés.

De la même façon, plusieurs modèles permettant d'évaluer la densité de sonie de nos signaux ont été utilisés. Ce sont ceux qui ont permis de tracer les graphiques de la Figure II.3. Ils sont tous basés sur les modèles de Zwicker *et coll.* [61][62][65][66][83]. Seuls ces modèles nous étaient disponibles. Nous avons donc décidé de les tester pour voir ce qu'ils pouvaient ou non apporter, en gardant à l'esprit que d'autres modèles pourraient être essayés par la suite. Les signaux sur lesquels ont été appliquées ces analyses sont identiques à ceux utilisés lors des autres analyses objectives. Nous avons défini des métriques permettant d'évaluer les dissemblances perceptives objectives entre les densités de sonie de nos enregistrements. Cette troisième voie sera comparée aux mesures objectives classiques, en déterminant si les dissemblances perceptives objectives calculées sur les densités de sonie permettent de se rapprocher davantage des dissemblances perceptives que ne le font les autres dissemblances objectives.

II.6.1 Densité de sonie totale : estimation du spectre perçu

Un premier modèle nous a permis de déterminer la densité de sonie totale de nos signaux [61][62]. Il est normalement destiné à l'analyse des sons stationnaires. La sensibilité de l'oreille et l'étalement spectral lié au fonctionnement du système auditif sont modélisés. Cette modélisation est appliquée de façon globale sur le spectre des signaux. La densité de sonie totale estime un spectre perçu des signaux, selon l'échelle des fréquences perçues ici modélisée par l'échelle des Barks. C'est ce type de modèle qu'ont utilisé Staffeldt [26][43][63] et Klippel [10][48].

Deux valeurs de sonie, et donc deux valeurs de densité de sonie, doivent être comparées en considérant leur rapport et non leur différence [65][66][67]. Les rapports entre les densités de sonie des enregistrements pris deux-à-deux ont ainsi été évalués en tout point de l'échelle des Barks. Avant de prendre les rapports, nous avons seuillé les densités de sonie afin de ne pas obtenir des valeurs de dissemblances artificiellement grandes là où il n'y a en fait pas de signal mais du bruit. Nous avons ainsi évalué les dissemblances seulement là où il y avait du signal utile. Le seuil a été arbitrairement choisi à 0,03 sone/Bark, toutes les valeurs inférieures à ce seuil ont été fixées à 0,03. En prenant les rapports, nous avons fait attention à toujours mettre au numérateur la plus grande des deux valeurs de densités et au dénominateur la plus petite. Cette opération est analogue à l'utilisation de la valeur absolue dans le cas d'une différence. Nous avons également retranché la valeur 1 à chaque rapport afin de s'assurer que des enregistrements identiques, dont les rapports de densité de sonie valent 1 en tout point, conduisent à une dissemblance nulle. Nous

obtenons finalement la dissemblance globale entre les deux densités par intégration sur la tonie, en prenant la moyenne arithmétique des dissemblances sur l'ensemble de l'échelle des Barks. Cette intégration est réalisée par analogie à la sommation de sonie qui permet d'obtenir la sonie d'un son en intégrant sa densité de sonie [65][66][67]. Ce calcul de dissemblance est illustré par un exemple présenté sur la Figure II.5.

La dissemblance calculée à partir de la **densité de sonie totale** peut finalement être exprimée par la formule :

$$Diss_{Dens_sonie_t}\{x(t),y(t)\} = \langle Max\{Dens_t(x),Dens_t(y)\}/Min\{Dens_t(x),Dens_t(y)\} - 1 \rangle_b$$

où $Dens_t$ est la densité de sonie totale des signaux, seuillée à 0,03 sone/Bark.

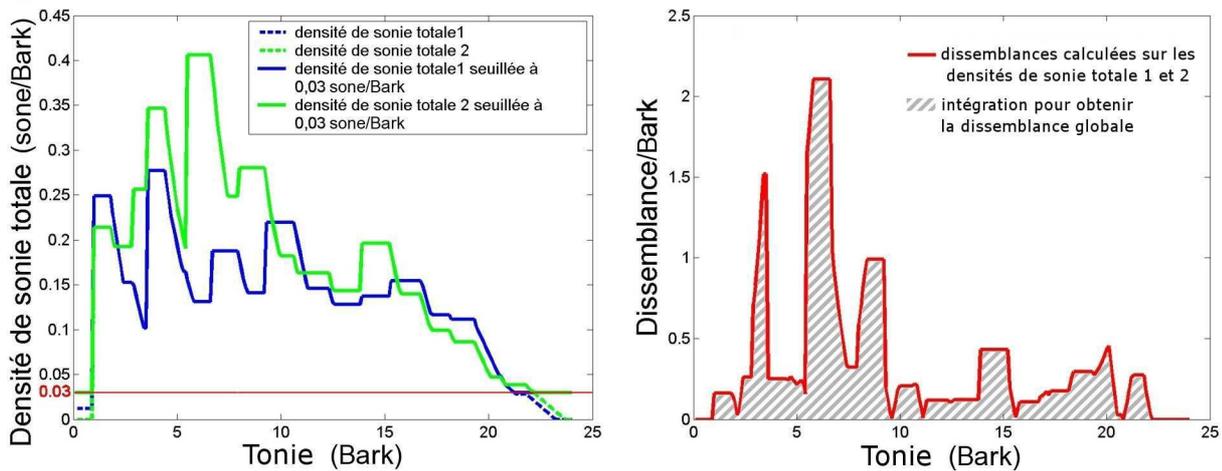


FIGURE II.5 – Illustration du calcul de la dissemblance entre deux densités de sonie totale

II.6.2 Densités de sonie : estimations du plan temps-fréquence perçu

Nous avons considéré la densité de sonie de nos signaux en fonction du temps en utilisant deux méthodes. La première consiste simplement à appliquer le modèle évoqué au paragraphe précédent à des portions successives de 100 ms du signal, ces portions étant alors considérées comme stationnaires. La densité de sonie ainsi calculée sera appelée densité de sonie 1 dans la suite du document. La deuxième méthode fait appel à un modèle dédié à l'étude des signaux non-stationnaires et ajoute la prise en compte de l'étalement temporel du signal dû au fonctionnement du système auditif [66][83]. La densité de sonie est alors calculée toutes les 10 ms. Elle sera nommée densité de sonie 2.

Les phénomènes d'étalement spectral liés au système auditif et modélisés lors du calcul des différentes densités de sonie dépendent du contenu spectral des signaux considérés, ce contenu pouvant évoluer au cours du temps. Pour la densité de sonie totale, le modèle auditif est appliqué de façon globale sur l'ensemble du signal, alors que le calcul des densités de sonie 1 et 2 est effectué en fonction du temps. Par rapport à la densité de sonie totale, ces deux dernières analyses tiennent compte du fait que les phénomènes d'étalement spectral liés au système auditif dépendent de l'évolution temporelle des signaux non-stationnaires considérés.

Ces deux analyses nous permettent d'estimer les dissemblances objectives entre nos signaux dans un plan temps-fréquence perçu [70], ou temps-fréquence sensoriel tel que nommé par Canévet *et coll.* [84]. A chaque instant, les dissemblances sont évaluées de manière analogue aux dissemblances définies sur la densité de sonie totale évoquées au paragraphe précédent. Comme pour les dissemblances objectives "classiques", l'intégration temporelle des dissemblances est réalisée en prenant leur moyenne arithmétique sur la durée totale du signal. La dissemblance ainsi calculée à partir des **densités de sonie 1 ou 2** peut être exprimée par la formule :

$$Diss_{Dens_sonie}\{x(t),y(t)\} = \langle \langle Max\{Dens(x),Dens(y)\}/Min\{Dens(x),Dens(y)\} - 1 \rangle_b \rangle_t$$

où Dens est la densité de sonie 1 ou 2 des signaux, seuillée à 0,03 sone/Bark.

Nous avons également calculé les **densités de sonie 1 et 2 moyennes** de nos signaux en prenant les moyennes temporelles de leurs densités de sonie 1 et 2. Nous avons une nouvelle fois choisi la moyenne arithmétique qui permet de donner une importance identique à tous les instants. Ces moyennes ont été déterminées pour les densités de sonie 1 et 2. Comme la densité de sonie totale du paragraphe précédent, elles estiment un spectre perçu des signaux, mais celui-ci est calculé différemment, en tenant notamment compte du caractère non-stationnaire des signaux étudiés. Les dissemblances objectives sont ensuite déterminées de la même façon que pour la densité de sonie totale. La différence par rapport aux dissemblances définies à partir des densités de sonie variant dans le temps est que pour les densités de sonie moyennes, l'intégration temporelle est réalisée sur les densités de sonie, avant d'évaluer les dissemblances.

$$Diss_{Dens_sonie_m}\{x(t),y(t)\} = \langle Max\{Dens_m(x),Dens_m(y)\}/Min\{Dens_m(x),Dens_m(y)\} - 1 \rangle_b$$

où $Dens_m$ est la densité de sonie 1 ou 2 moyenne des signaux, obtenue à partir de la densité de sonie correspondante seuillée à 0,03 sone/Bark.

Le modèle de Bramsløw [9] est lui aussi basé sur la densité de sonie en fonction de temps des signaux. Il utilise notamment la moyenne et l'écart-type temporels de cette densité de sonie.

II.7 Bilan de l'approche proposée

Notre approche pour aborder le problème de la confrontation des mesures physiques et perceptives sur les enceintes acoustiques consiste à nous concentrer dans un premier temps sur l'évaluation des différences entre enceintes. Comme nous l'avons vu au chapitre précédent ces évaluations relatives sont peu répandues dans la littérature, les évaluations absolues étant souvent préférées. Une approche basée sur les différences nous semble pourtant plus neutre. Nous n'avons pas cherché à évaluer directement la qualité globale des enceintes. Si nous sommes capables de développer un outil permettant de différencier les enceintes de façon pertinente, cet outil permettra de faire la différence entre "bonnes" et "mauvaises" enceintes, à partir du moment où elles auront été identifiées comme telles. Nous avons mis au point un protocole expérimental nous permettant de répondre aux exigences de contrôle des tests d'écoute, mais également de conserver les approches physiques et perceptives très proches l'une de l'autre tout au long de l'expérience. Nous espérons ainsi nous donner plus de chance de trouver le lien entre ces deux approches.

Le protocole expérimental consiste à enregistrer le champ sonore rayonné par les enceintes acoustiques. Les dissemblances perceptives entre ces enregistrements sont évaluées lors de tests psychoacoustiques réalisés au casque. Des dissemblances objectives définies par des métriques associées à différentes techniques de traitement de signal sont calculées entre les enregistrements. Certaines de ces analyses objectives font intervenir des modèles auditifs, introduisant ainsi des connaissances a priori dans l'analyse afin de tenir compte du fonctionnement complexe du système auditif. Les dissemblances objectives et perceptives sont ensuite comparées. Notre but est ainsi de déterminer les techniques d'analyse et les métriques associées qui permettent de retrouver les différences perçues par les auditeurs, c'est-à-dire celles qui permettent d'extraire des signaux l'information pertinente d'un point de vue perceptif. Notre approche peut être résumée par ces phrases de Houix [76]: "La production de jugements de proximité par des auditeurs sur des événements sonores, des expériences de classification de sons, peuvent permettre de refléter les processus perceptifs. Mais ces analyses peuvent être effectuées sur la proximité entre des événements sonores sur une base physique ou acoustique, pour les mettre en correspondance avec les résultats perceptifs". Le recours à l'utilisation du casque lors des tests psychoacoustiques nous empêche cependant d'étudier la composante spatiale de la reproduction sonore. Notre étude est donc dédiée à l'évaluation de la restitution du timbre par les enceintes acoustiques.

Les deux prochains chapitres décrivent l'ensemble des expériences que nous avons menées. Le chapitre III décrit les expériences proprement dites et la comparaison directe des dissemblances

perceptives et objectives obtenues. Le chapitre IV présente la comparaison des espaces perceptifs et objectifs résultant de l'analyse multidimensionnelle de ces dissemblances. Ces chapitres sont articulés autour de deux axes développés en parallèle. D'une part, la comparaison entre les dissemblances objectives et perceptives, effectuée sur différentes expériences impliquant notamment des stimuli différents en nombre et en nature. D'autre part, l'évaluation de l'influence des différents paramètres expérimentaux du protocole proposé. Nous avons étudié l'influence des paramètres d'enregistrements : la prise de son et l'extrait musical utilisés, ainsi que le mode de reproduction des enceintes. Les influences de la salle d'écoute et de la tâche perceptive ont également été considérées. Il s'agissait tout d'abord de savoir si la comparaison objectif/perceptif dépendait des paramètres expérimentaux. Ensuite, nous souhaitions vérifier si notre interprétation de la nature des dissemblances perceptives changeaient avec ces mêmes paramètres, c'est-à-dire si nous mesurons la même chose dans les différentes conditions.

CHAPITRE III

Comparaison directe des dissemblances : corrélation objectif/perceptif

————— oOo —————

Les trois séries d'expériences que nous avons menées sont décrites dans ce chapitre. Les sessions d'enregistrements et les tests d'écoute réalisés sont détaillés. La comparaison directe des dissemblances perceptives et objectives est dans un premier temps effectuée en évaluant leur corrélation. Ce critère de corrélation est sévère puisqu'il évalue l'existence d'une relation linéaire entre les dissemblances objectives et perceptives. Rien ne nous permet de pré-supposer la nécessité de cette linéarité lorsque nous recherchons le rapprochement entre les dissemblances objectives et perceptives. Nous verrons pourtant que les valeurs de corrélations obtenues indiquent qu'un rapprochement a été réalisé.

Ces premières applications de notre protocole expérimental avaient pour but de répondre à plusieurs questions. Bien que l'écoute se fasse au casque, les auditeurs perçoivent-ils des différences entre les reproductions sonores des différentes enceintes? Si oui, pouvons-nous nous rapprocher des différences perçues à l'aide d'une méthode d'analyse objective? Et dans ce cas, le rapprochement objectif/perceptif est-il conditionné par les différents paramètres expérimentaux?

————— oOo —————

SOMMAIRE

III.1 Première série d'expériences : influence de l'extrait musical et de la prise de son	59
III.1.1 Enregistrements	59
III.1.2 Tests d'écoute	60
III.1.3 Corrélation entre dissemblances objectives et perceptives	64
III.2 Deuxième série d'expériences : influence du mode de reproduction des enceintes	66
III.2.1 Enregistrements	67
III.2.2 Tests d'écoute	69
III.2.3 Corrélation entre dissemblances objectives et perceptives	69
III.2.4 Analyse de variance	71
III.3 Troisième expérience : influence de la pièce et des stimuli utilisés	72
III.3.1 Enregistrements	72
III.3.2 Tests d'écoute	74
III.3.3 Corrélation entre dissemblances objectives et perceptives	74
III.4 Bilan de la comparaison directe des dissemblances	75

III.1 Première série d'expériences : influence de l'extrait musical et de la prise de son

Les normes ou recommandations concernant les tests d'écoute sur les enceintes acoustiques [30][31] conseillent de considérer les reproductions monophoniques et stéréophoniques pour que l'évaluation soit complète. Notre étude portant sur la restitution du timbre, nous avons enregistré le champ sonore produit par les enceintes utilisées en reproduction monophonique. "Timbre-related accuracy is much more easily heard in single-loudspeaker listening" [31]. Éliminer l'information spatiale fournie par la reproduction stéréophonique peut en effet permettre aux auditeurs de se concentrer davantage sur les aspects de restitution du timbre [27]. L'influence du mode de reproduction des enceintes a tout de même été considéré lors de notre deuxième série d'expériences (paragraphe III.2, page 66). Pour cette première série d'expériences, nous avons utilisé une reproduction monophonique. Différents extraits musicaux et différentes prises de son ont été testés.

III.1.1 Enregistrements

La tête artificielle a permis de réaliser une prise de son binaurale. Trois principes de prise de son stéréophonique ont été mis en oeuvre : le couple AB ORTF, le couple XY et le couple MS. Une prise de son monophonique omnidirectionnelle a également été effectuée et le couple MS a aussi fourni une prise monophonique cardioïde, par sommation de ses voies de droite et de gauche. Le principe de ces différentes techniques est détaillé dans l'annexe A.1.

Les enregistrements ont été réalisés dans une salle du laboratoire qui n'est pas habituellement utilisée pour la reproduction sonore. La Figure III.1 présente le plan de cette pièce qui ne répond pas exactement aux standards de salle d'écoute [30][31], du fait de sa longueur et de la présence d'une grande vitrine. Le sol est entièrement recouvert de moquette. La durée de réverbération a été évalué à 0,5 s en moyenne entre 500 et 5000 Hz (annexe A.5), pour un volume voisin de 115 m^3 . Comme les enceintes sont toutes utilisées dans la même salle et que nous nous intéressons à leur différences relatives, nous avons décidé d'utiliser cette pièce sans traitement acoustique supplémentaire, considérant qu'elle offrait des conditions d'enregistrement raisonnables.

Douze enceintes ont été impliquées dans ces séances d'enregistrement. Nous avons essayé d'utiliser un panel diversifié, avec des modèles employant des technologies variées et provenant de fabricants différents (annexe A.2). Malgré tout, nous n'avons pas pu disposer d'enceinte très haut de gamme pour cette expérience. Deux des enceintes étaient du même modèle, notre but étant de vérifier si les enregistrements correspondants seraient perçus comme proches lors des tests psychoacoustiques. Il est à noter qu'à notre connaissance douze enceintes incluses dans le même test de comparaison par paire est de l'ordre du maximum de ce qui a été fait dans la

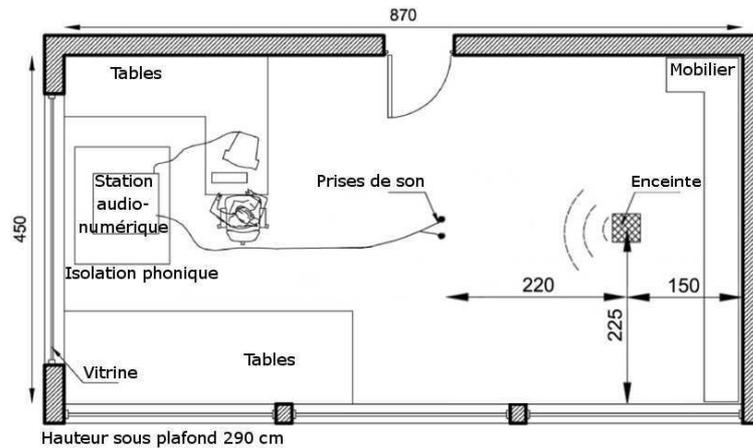


FIGURE III.1 – Salle d'enregistrement et dispositif expérimental utilisés lors de la première série d'expériences

littérature.

Les enceintes étaient toutes placées à la même position dans la salle (Figure III.1). Malheureusement, nous n'avons pas un bon contrôle du positionnement vertical de ces enceintes de tailles différentes. A moins que l'enceinte ne soit spécialement conçue pour être posée directement sur le sol, elles étaient toutes posées sur le même pied dont la hauteur n'était pas réglable. Ceci a été corrigé dans les expériences suivantes. La position des microphones était également fixée. Les enregistrements ont été effectués en deux sessions distinctes. Au cours de la première, nous avons réalisé les prises de son AB ORTF et XY. Lors de la deuxième session, nous avons utilisé la tête artificielle, le couple MS et la prise monophonique omnidirectionnelle. Les prises de son ont été réalisées à environ 1 m du sol, dans une zone restreinte de l'espace correspondant approximativement à la tête d'un auditeur assis à 2,20 m en face de l'enceinte (Figure III.1).

Plusieurs extraits musicaux ont été enregistrés sur un même disque compact avant d'être reproduits sur les enceintes par un système haute-fidélité. Ces extraits étaient des signaux monophoniques obtenus en ne conservant qu'une seule piste des signaux stéréophoniques originaux [30][31]. Seuls certains d'entre eux ont été utilisés par la suite dans les tests d'écoute.

La description de l'ensemble du matériel utilisé et des photographies de l'expérience se trouvent dans l'annexe A.

III.1.2 Tests d'écoute

Stimuli utilisés Afin de limiter le nombre de tests perceptifs à réaliser, nous avons sélectionné trois prises de son et trois extraits musicaux parmi les enregistrements effectués.

Nous avons décidé d'utiliser les enregistrements provenant de trois prises de son choisies pour

leurs différences parmi les prises de sons utilisables :

- la prise stéréophonique AB ORTF,
- la prise stéréophonique MS,
- et la prise monophonique omnidirectionnelle.

La prise binaurale à la tête artificielle, bien que théoriquement la plus adaptée car conçue pour une écoute au casque plus réaliste, c'est-à-dire plus proche d'une écoute directe de l'événement sonore, a dû être abandonnée. En effet, nous n'avons pas réussi à compenser des dégradations trop importantes en terme de réponse fréquentielle qui la rendent inutilisable pour notre étude portant sur la restitution du timbre. Cependant, il est à noter que pour notre étude basée sur une reproduction frontale monophonique dans laquelle la composante spatiale de la reproduction sonore n'est pas étudiée, l'utilisation de techniques binaurales n'apparaît pas cruciale. La prise AB ORTF a été la prise préférée des auditeurs lors d'une écoute informelle où les prises de son étaient comparées. Nous avons indiqué aux auditeurs que nous apportions une importance particulière au réalisme de l'écoute et à la capacité d'évaluer des différences de timbre entre les différents enregistrements provenant d'une même prise de son. La prise MS a été choisie plutôt que la prise XY qui ressemblait davantage à la prise AB ORTF, également basée sur l'utilisation de deux capsules cardioïdes. Enfin la prise monophonique omnidirectionnelle a été choisie plutôt que la prise monophonique cardioïde qui était plus proche de la prise stéréophonique MS dont elle a été extraite.

Nous avons ensuite choisi trois extraits musicaux de styles bien différents dont nous avons sélectionné une très courte séquence :

- Kan'nida ("Konsyans", percussions, présence de transitoires et de basses fréquences ; 1.7 s.),
- Mc Coy Tyner ("Miss Bea", jazz, combinaison d'un petit nombre d'instruments : saxophone, piano, basse, batterie ; 3.3 s.),
- Vivaldi ("L'Europa Galante", orchestre symphonique complet ; 4.7 s.).

Nous avons besoin d'extraits courts pour limiter la durée des tests psychoacoustiques, afin que les auditeurs soient capables de les effectuer dans des conditions convenables. Des durées d'extraits si courtes se sont en fait révélées particulièrement adaptée à l'évaluation de différences entre les enregistrements.

"It seems clear that relatively rapid comparisons are advantageous for maximum discrimination and minimum variability in the judgments" Toole [14]. Ainsi Olive *et coll.* [16] ont d'abord utilisé des extraits de 20 à 60 secondes, puis Olive [28] a considéré des durées de 15 à 30 secondes, ensuite raccourcies entre 10 et 15 secondes pour faciliter la comparaison. De même alors que Gabriellson *et coll.* [7][19] ont commencé leurs études avec des extraits de 30 secondes, ils ont ensuite réalisé des comparaisons par paires en se limitant à 15 secondes de signal [22]. Vogel

et coll. [85] ont constaté que des extraits de 30 secondes étaient encore trop longs pour réaliser des comparaisons par paires. Les études évoquées au chapitre I qui ont eu recours à des signaux de synthèse ont également impliqué de courts extraits. Bücklein [57] a considéré 6 secondes de bruit blanc, de courts extraits musicaux et des signaux de parole contenant deux courtes phrases. Bech [72][73] a utilisé 3,8 secondes de parole et 1 seconde de bruit rose pour effectuer des comparaisons par paires, tandis que Moore et Tan [47] ont considéré 3,1 secondes de parole et 7,3 secondes de jazz ("piano, bass and drums") pour leurs tests. Pour étudier l'émotion suscitée par l'écoute musicale, Bigand *et coll.* [81] ont réalisé des tests d'écoute avec des extraits de 30 secondes, puis les ont réduit à 1 seconde. Ils constatent que cette grosse réduction n'a qu'un faible impact sur leurs résultats.

Nous avons également constaté au cours d'expériences informelles que la tâche de comparaison était plus facile à réaliser sur des extraits courts. Des extraits musicaux plus longs, même s'ils sont homogènes, entraînent une tâche de mémorisation supplémentaire pour être comparés. C'est le test faisant intervenir l'extrait Vivaldi, le plus long de nos signaux qui reste pourtant très court, qui a été jugé le plus difficile par presque tous les auditeurs, notamment du fait de cette tâche de mémorisation supplémentaire.

Tests réalisés Les tests psychoacoustiques menés lors de cette expérience étaient des tests de comparaison par paires dont la tâche consistait à évaluer les ressemblances perceptives entre les enregistrements faisant intervenir les différentes enceintes acoustiques. Le principe de ces tests a été décrit au paragraphe II.4.1, page 40. Lors des enregistrements, les niveaux sonores produits par les différentes enceintes ont été approximativement égalisés à un niveau d'écoute naturel. Une égalisation plus fine de la sonie des enregistrements a été réalisée à l'oreille par les expérimentateurs avant les tests psychoacoustiques. Le niveau perçu de la reproduction au casque était de l'ordre de 70 phons, selon l'appréciation des expérimentateurs. Le niveau correspondant en dB SPL n'a pas été mesuré, notamment du fait de la difficulté de prendre en compte lors de la mesure des phénomènes tels que le couplage entre la tête et le casque ou la sommation binaurale intervenant lors de l'écoute stéréophonique. Néanmoins, nous pouvons affirmer que la reproduction au casque correspondait à un niveau d'écoute musicale confortable, et aucun auditeur n'a fait de remarque concernant ce niveau. Le détail du dispositif expérimental employé et la consigne fournie aux auditeurs sont présentés dans l'annexe A.6.

Nous avons réalisé cinq tests d'écoute répartis en deux séries. Une première série de trois tests utilisait les enregistrements réalisés au couple AB ORTF, avec un des trois extraits musicaux par test. Ensuite, dans une deuxième série de deux tests, nous avons conservé uniquement l'extrait Kan'nida et utilisé les enregistrements provenant du couple MS pour le premier test et de la prise monophonique omnidirectionnelle pour le second. Lors de la première série, l'ordre de passage des tests correspondants aux différents extraits a été réparti uniformément entre les auditeurs.

Pour la deuxième série, la moitié des auditeurs a commencé par le test utilisant le couple MS et l'autre moitié a commencé par celui basé sur la prise monophonique omnidirectionnelle. Nous avons choisi l'extrait Kan'nida pour la deuxième série de tests car c'est celui qui a été jugé le plus facile par les auditeurs pour évaluer les différences entre les enceintes lors de la première série de tests. C'est également l'extrait pour lequel les réponses des auditeurs étaient les plus corrélées entre elles. Il sera montré par la suite que les trois extraits musicaux testés ont conduit à des résultats très similaires.

Auditeurs La première série de tests utilisant les trois extraits musicaux et la prise de son AB ORTF a été réalisée par vingt-sept auditeurs (douze femmes, quinze hommes). Ils appartenaient tous au personnel du laboratoire, du C.N.R.S., ou bien étaient des étudiants de passage dans notre équipe. Aucun n'avait d'expérience dans le domaine de l'écoute comparative d'enceintes acoustiques. Aucun ne présentait de déficit auditif, d'après leurs audiogrammes réalisés selon la méthode de Békésy. Ils avaient de quatorze à cinquante-cinq ans, avec une moyenne d'âge de trente ans et un écart type de neuf ans.

Parmi ces vingt-sept auditeurs, quinze personnes (six femmes, neuf hommes) ont participé à la deuxième série de tests basée sur les deux autres prises de son. Ils avaient de vingt-trois à quarante et un ans, avec une moyenne d'âge de vingt-sept ans et un écart type de cinq ans. Le nombre de quinze auditeurs a été jugé suffisant car, lors de la première série de tests, les dissemblances moyennées sur les auditeurs ne variaient plus beaucoup à partir d'environ douze auditeurs. Or, ce sont ces dissemblances moyennes qui nous intéressent pour analyser les résultats des tests perceptifs.

Avant de considérer la moyenne des dissemblances sur l'ensemble des auditeurs, nous avons vérifié qu'il n'y avait pas de groupe d'auditeurs ayant des stratégies de réponse différentes. Une analyse en arbre hiérarchique [51][76][77] a pour cela été appliquée sur la matrice des dissemblances entre auditeurs (paragraphe II.4.1, page 40). La Figure III.2 présente l'exemple du dendrogramme obtenu après analyse des résultats du test faisant intervenir l'extrait musical Kan'nida et la prise de son AB ORTF. Même si les évaluations des auditeurs 3 et 16 sont moins corrélées à celles des autres auditeurs pour ce test, nous avons conservé leurs résultats. Nous avons préféré avoir recours à un nombre plus important d'auditeurs pour réduire le bruit de mesure, plutôt que d'éliminer les réponses atypiques. Les dendrogrammes correspondants aux autres tests sont présentés en annexe C.1. Aucune de ces analyses n'a révélé de classes parmi les auditeurs, et notamment aucune distinction entre les musiciens et les non-musiciens. Nous avons également constaté que les quelques auditeurs "atypiques" n'étaient pas forcément les mêmes d'un test à l'autre.

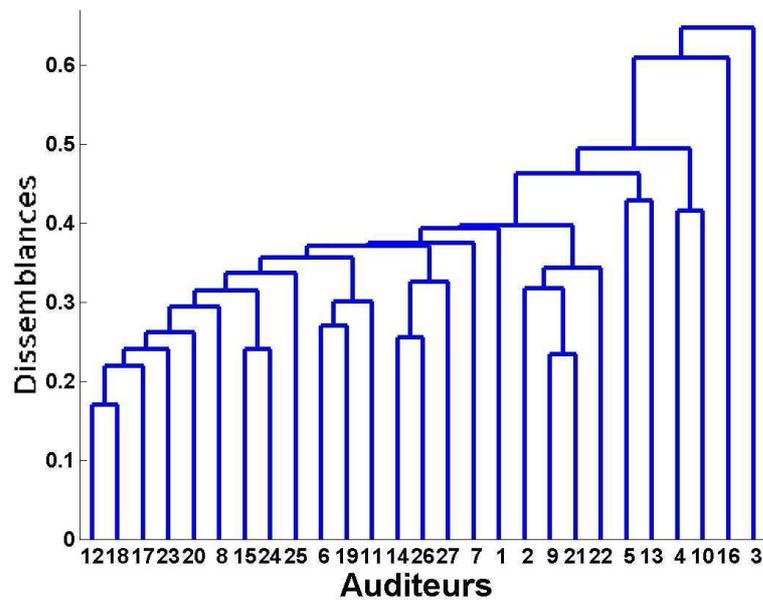


FIGURE III.2 – Exemple de dendrogramme résultant de l'analyse en arbre hiérarchique des dissemblances entre auditeurs (Kan'nida, couple AB ORTF)

III.1.3 Corrélation entre dissemblances objectives et perceptives

En parallèle des tests psychoacoustiques, nos enregistrements ont été soumis aux différentes analyses objectives évoquées au chapitre précédent. Afin de déterminer la technique d'analyse et la métrique associée les plus appropriées pour retrouver les différences perçues par les auditeurs, les dissemblances objectives et perceptives ont été comparées, dans un premier temps directement en évaluant leur corrélation. Lorsque nous cherchons à rapprocher les dissemblances objectives et perceptives, rien ne nous permet de supposer que leur relation doit être linéaire. Nous verrons pourtant que les fortes corrélations observées indiquent que ce rapprochement a bien été effectué.

Nous avons calculé l'ensemble des corrélations entre les dissemblances perceptives provenant de chacun des cinq tests de l'expérience et les dissemblances objectives calculées sur chacune des voies des signaux d'enregistrements correspondants (annexe C.2). Les analyses objectives ne donnent pas exactement les mêmes résultats sur les voies droite et gauche des signaux stéréophoniques, mais les tendances observées restent les mêmes. Nous avons alors considéré les valeurs moyennes des résultats obtenus sur ces deux voies. Ces corrélations moyennes sont représentées sur la Figure III.3.

En comparant les résultats obtenus pour les différents tests, nous retrouvons bien toujours les mêmes tendances quant à la comparaison objectif/perceptif, malgré les modifications de prise de son et d'extraits musical. L'information de phase ne semble pas pertinente vis-à-vis de la perception dans le cadre de nos expériences portant sur la restitution du timbre. En effet, à

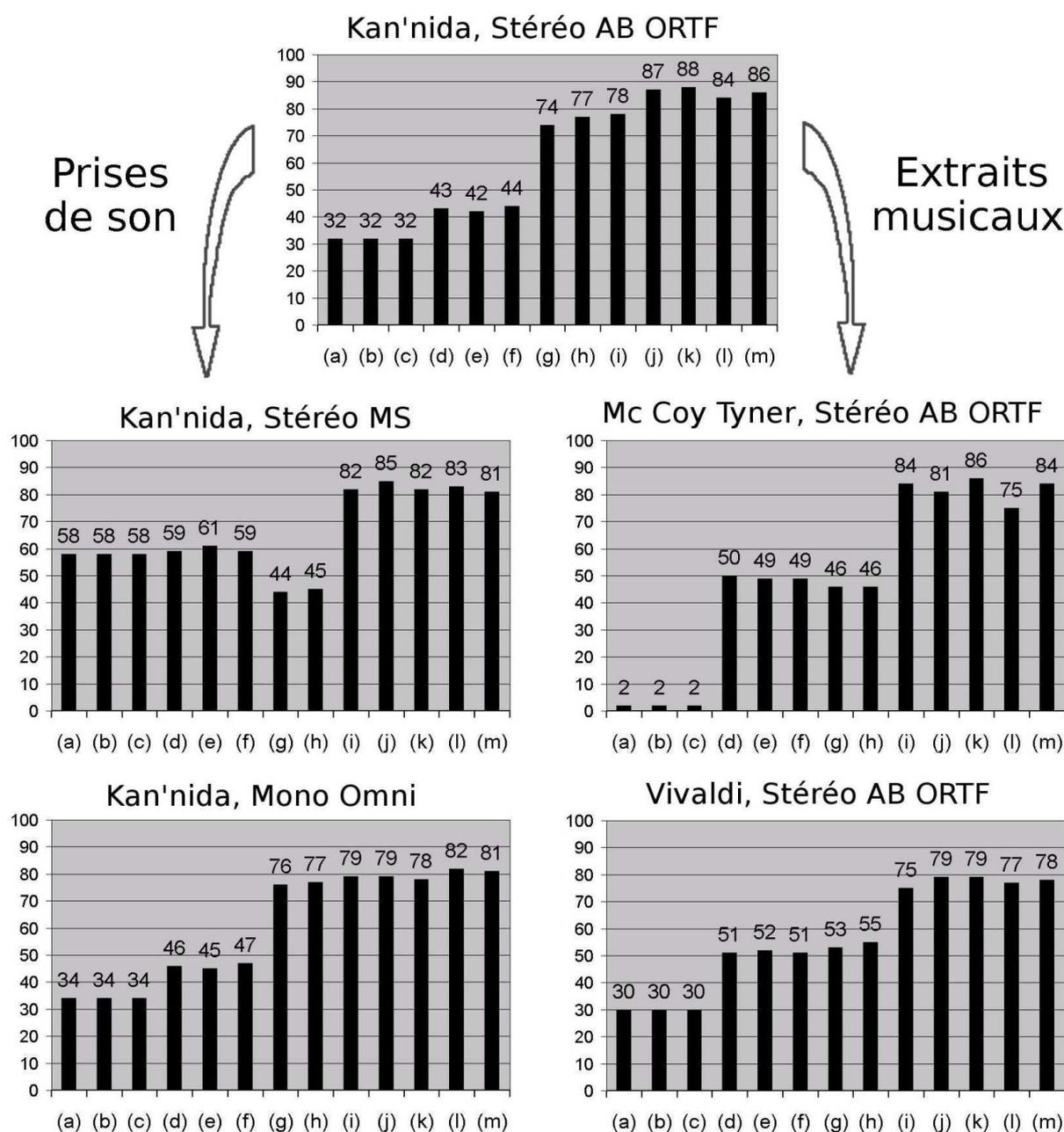


FIGURE III.3 – Première série d'expériences. Corrélations moyennes entre les dissemblances perceptives et objectives (en %) : (a) domaine temporel (b) spectre (c) temps-fréquence (d) module du spectre (e) module du temps-fréquence (f) densité spectrale de puissance (g) pondération A (h) pondération Iso70 (i) densité de sonie totale (j) densité de sonie 1 (k) densité de sonie 1 moyenne (l) densité de sonie 2 (m) densité de sonie 2 moyenne.

partir des domaines fréquentiel (b et d,f) et temps-fréquence (c et e), considérer uniquement le module en éliminant la phase améliore la corrélation avec les dissemblances perceptives. Les

pondérations fréquentielles testées apportent une amélioration pour l'extrait musical Kan'nida et les prises de son AB ORTF et monophonique omnidirectionnelle. Mais cette amélioration n'a pas été confirmée pour les autres signaux, elle s'est même transformée en dégradation de la corrélation dans le cas de l'extrait musical Kan'nida et la prise de son MS. Par contre, la prise en compte du fonctionnement complexe de l'oreille par l'utilisation d'un modèle auditif permet d'augmenter de façon significative la corrélation entre dissemblances perceptives et objectives, et ceci pour tous les signaux testés. A ce stade de l'analyse, nous ne pouvons pas dire quel modèle utilisé pour déterminer la densité de sonie de nos signaux est le plus pertinent. L'évolution temporelle de l'information ne semble pas apporter un plus pour expliquer les dissemblances perceptives qui semblent davantage liées à un jugement moyen sur l'ensemble de l'extrait. En effet, le spectre (b et d,f) et les densités de sonie moyenne (k,m) ou totale (i) donnent respectivement les mêmes résultats que le temps-fréquence (c et e) et la densité de sonie (j,l).

Nous constatons également que plusieurs domaines d'analyse munis de leur métrique associée sont équivalents. Ainsi les domaines temporel, fréquentiel et temps-fréquence, quand les informations de phase sont conservées, fournissent les mêmes résultats. De même, nous pouvons considérer indifféremment le module du spectre, celui du temps-fréquence ou la racine carrée de la densité spectrale de puissance. Par contre, la densité de sonie et la densité de sonie moyenne ne sont pas complètement équivalentes.

Nous n'avons pas su expliquer les valeurs de corrélation particulièrement faibles obtenues avec l'extrait Mc Coy Tyner, à partir du domaine temporel et des domaines spectral et temps-fréquence lorsque les informations de phase sont prises en compte (Figure III.3). Elles pourraient être dues à une synchronisation imparfaite des signaux temporels avant les analyses objectives. Même si le décalage temporel entre les signaux à comparer a été minimisé, la méthode de synchronisation utilisée n'est pas parfaite (annexe D). Ce phénomène n'a pas été observé pour les autres signaux.

Il est à noter que lorsque nous avons observé les valeurs de dissemblances perceptives résultant des différents tests psychoacoustiques, les enregistrements faisant intervenir les enceintes d'un même modèle ont bien été perçus comme très semblables, mais pas forcément toujours comme les plus semblables. Dans le cadre de notre protocole expérimental, certaines enceintes de modèles différents peuvent conduire à des reproductions sonores jugées aussi semblables que celles produites par deux enceintes du même modèle. Il est également important de garder à l'esprit que, comme toute mesure, l'évaluation psychoacoustique possède une résolution limitée.

III.2 Deuxième série d'expériences : influence du mode de reproduction des enceintes

Même si la composante spatiale de la reproduction sonore n'est pas considérée dans notre étude, la reproduction stéréophonique pourrait avoir une influence sur la restitution du timbre

par les enceintes acoustiques. Notre deuxième série d'expériences a donc eu pour but d'évaluer l'influence du mode de reproduction des enceintes dans le cadre de notre protocole. Nous avons décomposé le problème : le fait d'utiliser deux enceintes plutôt qu'une seule et l'emploi du signal stéréophonique ont été étudiés séparément. Nous avons cherché à savoir si le mode de reproduction utilisé lors des séances d'enregistrements modifiait la comparaison objectif/perceptif et la perception des différences entre les enregistrements.

Les enregistrements et les tests psychoacoustiques de cette série d'expériences ont été effectués avec Benjamin Guyot, qui a travaillé avec nous dans le cadre de son stage de Master "Mécanique, Physique et Modélisation, parcours Acoustique".

III.2.1 Enregistrements

Les séances d'enregistrements ont fait intervenir onze paires d'enceintes. Huit des enceintes déjà impliquées dans notre première série d'expériences ont été utilisées. Trois modes de reproduction différents ont été employés. Nous avons d'abord considéré la reproduction monophonique, puis la reproduction stéréophonique, et enfin ce que nous avons appelé la reproduction "double monophonique". La reproduction double monophonique constitue un intermédiaire entre les reproductions monophoniques et stéréophoniques, dans lequel les paires d'enceintes sont placées comme dans la configuration stéréophonique, mais sont toutes les deux alimentées par le même signal qu'en reproduction monophonique. Pour la reproduction monophonique, seule une enceinte de chaque paire a été considérée. Cependant, pour trois paires, nous avons également inclus la seconde enceinte dans les séances d'enregistrements. Ceci a été fait dans le but de pouvoir comparer une différence potentielle entre les reproductions monophonique et double monophonique avec une différence potentielle entre deux enceintes d'une même paire. Si nous trouvions une différence entre les reproductions monophonique et double monophonique, nous avons besoin de savoir si cette différence était due au fait d'utiliser deux enceintes plutôt qu'une seule, ou si elle était due à une différence entre les deux enceintes de la paire.

Les enregistrements ont été réalisés dans la même pièce que celle utilisée lors de la première série d'expériences. Le plan du dispositif expérimental est présenté sur la Figure III.4. Les enceintes et la prise de son étaient placées de telle sorte qu'elles forment un triangle équilatéral dans la configuration stéréophonique, les enceintes étant orientées vers la position d'enregistrement [30][31]. Un compromis a été fait afin de permettre à la fois un espacement suffisant des enceintes de la paire stéréophonique, et de conserver en même temps les enceintes suffisamment loin des murs. En reproduction monophonique, les enceintes sont situées en face de la prise de son, dans le même plan que les paires d'enceintes en reproduction stéréophonique. Seule la prise de son AB ORTF a été employée. Les microphones étaient situés à 1 m au dessus du sol. La position verticale des enceintes était repérée par le point situé entre le haut-parleur du médium

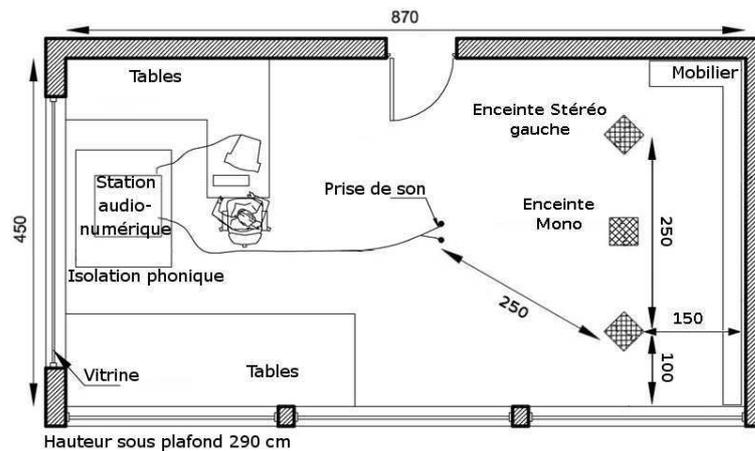


FIGURE III.4 – *Dispositif expérimental utilisés lors de la deuxième série d'expériences*

et le tweeter. Comme les microphones, ce point était placé à une hauteur de 1m, à moins que l'enceinte ne soit spécialement conçue pour être posée directement sur le sol. Nous avons en cela suivi la norme [30] et les exemples de Toole [13][14] et Olive *et coll.* [16][46], qui recommandent de placer le haut-parleur des aigus ou celui des médium/aigus au niveau des oreilles de l'auditeur. Étant donnée la difficulté de déterminer précisément la position du centre acoustique d'une enceinte, nous ne pouvons pas dire que les enceintes sont exactement à la même position dans la salle d'écoute. Malgré tout, nous avons minimisé autant que possible l'influence de ce paramètre expérimental.

Les extraits musicaux utilisés lors de cette série d'expériences étaient les mêmes que ceux de la première série d'expériences, afin de permettre leur comparaison, mais cette fois les deux pistes du signal stéréophonique étaient conservées. Les signaux monophoniques employés pour les reproductions monophonique et double monophonique ont été obtenus par sommation des voies droite et gauche des signaux stéréophoniques. Seul l'extrait Mc Coy Tyner a été utilisé dans les tests psychoacoustiques. Au moment de choisir entre les trois extraits musicaux utilisés lors de notre première série d'expériences, nous avons tenté d'évaluer le caractère stéréophonique de ces extraits. Il nous fallait en effet choisir un extrait "suffisamment stéréophonique" pour se donner une chance de mettre en évidence une éventuelle influence des trois modes de reproduction testés pour les enceintes. La corrélation entre les voies droite et gauche des signaux a d'abord été évaluée. Mc Coy Tyner est l'extrait qui possède les voies droite et gauche les moins décorréélées, avec un coefficient de corrélation entre 0,5 et 0,6 suivant la longueur de signal considérée, contre 0,4 à 0,5 pour les deux autres extraits. Il a pourtant été finalement choisi parce qu'il nous semblait avoir le spectre le plus "équilibré" des trois extraits (Figure II.3, page 49), et que notre expérience porte sur la restitution du timbre et l'influence que pourrait avoir le mode de reproduction des enceintes sur cette restitution. Cet extrait nous a paru posséder une répartition stéréophonique

suffisante, évitant le cas extrême d'instruments présents uniquement sur une des deux voies, et offrant une orchestration assez aérée pour permettre de révéler des effets de salle potentiellement plus marqués pour une reproduction impliquant deux enceintes.

III.2.2 Tests d'écoute

Comme lors de la première série d'expériences, les enregistrements réalisés ont été soumis à des tests de comparaison par paires. Le niveau perçu de la reproduction au casque était toujours d'environ 70 phones. Le dispositif expérimental employé et la consigne fournie aux auditeurs étaient les mêmes que pour les expériences précédentes.

Deux tests d'écoute ont été effectués. Le test 1 faisait intervenir les enregistrements des reproductions monophonique et double monophonique et le test 2 utilisait ceux des reproductions double monophonique et stéréophonique. Au cours de chaque test, les enregistrements de deux modes de reproduction différents n'ont jamais été comparés les uns aux autres. Ils étaient pourtant inclus dans le même test, afin de s'assurer que les auditeurs utilisent la même échelle lorsqu'ils évaluent les dissemblances entre les enregistrements pour les deux modes de reproduction.

Vingt auditeurs ont pris part aux expériences (huit femmes, douze hommes). Ils avaient entre dix-neuf et quarante deux ans, avec une moyenne d'âge de vingt-cinq ans et un écart type de cinq ans. Ils étaient membres du laboratoire ou étudiants, et aucun d'eux n'avait d'expérience particulière dans la comparaison des enceintes acoustiques. Certains avaient tout de même participé aux tests d'écoute de notre première série d'expériences. La moitié des auditeurs a commencé par le test 1, tandis que les autres ont débuté par le test 2. Il leur a fallu en moyenne entre vingt et trente minutes pour réaliser chacun des tests. Aucune classe d'auditeurs ayant des stratégies de réponses différentes n'a été mise en évidence par l'analyse en arbre hiérarchique.

III.2.3 Corrélation entre dissemblances objectives et perceptives

Les dissemblances perceptives ont ensuite été comparées aux dissemblances calculées à l'aide des différentes techniques d'analyse objective. La Figure III.5 présente les valeurs moyennes des corrélations entre les dissemblances perceptives et les dissemblances objectives calculées sur les deux voies de nos signaux, pour chaque analyse objective et pour les deux modes de reproduction de chacun des deux tests psychoacoustiques réalisés. Les dissemblances perceptives des enregistrements de la reproduction double monophonique ont été évaluées deux fois, une fois au cours du test 1 et une fois au cours du test 2. Les corrélations calculées à partir de ces deux évaluations successives sont représentées.

Les résultats obtenus révèlent les mêmes tendances que ceux de notre première série d'expériences (Figure III.3, page 65), et ceci pour les trois modes de reproduction considérés. Les pondérations fréquentielles testées n'apparaissent pas pertinentes pour se rapprocher de la per-

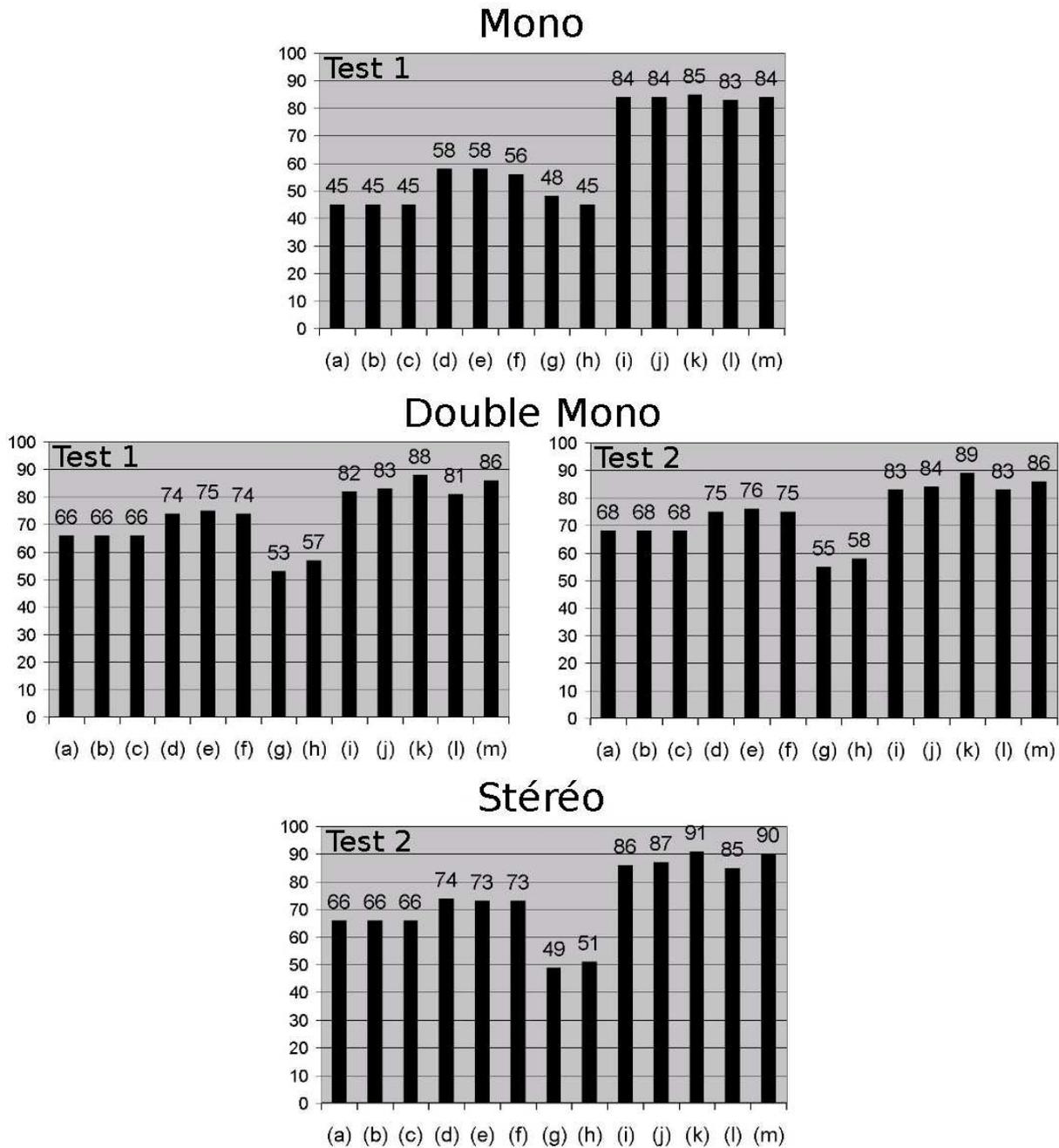


FIGURE III.5 – Deuxième série d'expériences. Corrélations moyennes entre les dissemblances perceptives et objectives (en %) : (a) domaine temporel (b) spectre (c) temps-fréquence (d) module du spectre (e) module du temps-fréquence (f) densité spectrale de puissance (g) pondération A (h) pondération Iso70 (i) densité de sonie totale (j) densité de sonie 1 (k) densité de sonie 1 moyenne (l) densité de sonie 2 (m) densité de sonie 2 moyenne

ception. Alors qu'elles n'avaient pas d'effet lors de la première série d'expériences avec ce même extrait musical (Mc Coy Tyner), elles entraînent ici une baisse de la corrélation avec les dissemblances perceptives, par rapport aux analyses spectrales sans pondération. L'intérêt de considérer les densités de sonie des signaux en ayant recours à un modèle auditif est quant à lui confirmé par cette deuxième série d'expériences.

III.2.4 Analyse de variance

Nous avons vu que les tendances observées en comparant les corrélations entre les différentes dissemblances objectives et les dissemblances perceptives n'étaient pas dépendantes du mode de reproduction des enceintes. Cela ne signifie en rien que les évaluations effectuées, tant perceptives qu'objectives, ne sont pas influencées par ce paramètre. Les matrices de dissemblances obtenues pour les trois modes de reproduction ne sont pas identiques.

Comme chaque test impliquait les enregistrements de deux modes de reproduction différents, l'influence du changement de mode de reproduction a pu être quantifiée grâce à une analyse de variance (ANOVA pour analysis of variance). Les aspects théoriques de cette analyse sont présentés dans l'annexe B.4. L'impact de l'utilisation de deux enceintes plutôt qu'une seule a été évalué par l'analyse des résultats du test 1 faisant intervenir les enregistrements des reproductions monophonique et double monophonique. L'influence de l'utilisation d'un signal stéréophonique a été quantifiée en appliquant la même analyse sur les résultats du test 2 impliquant les enregistrements des reproductions double monophonique et stéréophonique.

La modification du mode de reproduction a eu un effet statistiquement significatif dans les deux cas ($p < 0,01$). Cet effet est cependant très faible car il n'explique qu'une faible part de la variance : moins de 4% pour le premier test et moins de 1% pour le second. Le détail de ces deux analyses de variance est disponible dans l'annexe C.3.

	Moyenne globale des dissemblances
Mono (test 1)	0,54
Double Mono (test 1)	0,51
Double Mono (test 2)	0,54
Stéréo (test 2)	0,50

TABLEAU III.1 – Moyennes globales des dissemblances perceptives pour chaque mode de reproduction lors des deux tests de notre deuxième série d'expériences

Pour chacun des deux tests, nous avons calculé la moyenne globale des dissemblances pour chaque mode de reproduction. Les valeurs obtenues sont présentées dans le Tableau III.1. Ces différences de moyenne, très faibles mais significatives d'après l'analyse de variance, indiquent que la reproduction monophonique serait plus discriminante que la reproduction double mono-

phonique, puisqu'elle conduit à des valeurs de dissemblances globalement plus importantes. De la même façon, la reproduction double monophonique serait plus discriminante que la reproduction stéréophonique. Ces résultats sont à considérer avec précaution étant donnés les faibles écarts observés. Les auditeurs auraient trouvé davantage de différences entre les enregistrements faisant intervenir les enceintes en reproduction monophonique. Ces résultats sont en accord avec ceux de Toole concernant l'influence de la stéréophonie sur l'évaluation de la qualité des enceintes acoustiques [14][17]. Il a trouvé que les classements des enceintes sur une échelle de fidélité étaient les mêmes pour les reproductions monophonique et stéréophonique, mais que l'écoute monophonique était plus discriminante, notamment pour les enceintes de mauvaise qualité. Ces tests étaient alors réalisés en écoute directe.

Ce type d'analyse n'a pas pu être effectué sur les résultats de la première série d'expériences, car les enregistrements correspondant aux différentes prises de son et aux différents extraits musicaux ont été évalués lors de tests psychoacoustiques indépendants. Les résultats de ces tests ne peuvent pas être comparés directement, car les auditeurs n'ont pas forcément utilisé la même échelle pour évaluer les dissemblances lors des différents tests. Pour chacun de ces tests, il était en effet demandé aux auditeurs d'utiliser toute l'échelle fournie pour quantifier les similarités, car aucune normalisation de leurs évaluations n'est appliquée a posteriori. Nous n'avons donc pas accès directement aux informations relatives à des différences de jugement de dissemblance selon que nous utilisons un extrait musical ou un autre, ou bien une prise de son plutôt qu'une autre. Cela aurait été le cas seulement si plusieurs extraits ou prises de son avaient été inclus au sein d'un même test.

III.3 Troisième expérience : influence de la pièce et des stimuli utilisés

Lors de nos séances d'enregistrements, l'enceinte n'est pas mesurée seule, c'est l'ensemble enceinte-salle qui est considéré. Dans le cadre de notre protocole, l'influence de la pièce ne pouvant pas être éliminée, nous avons cherché à l'évaluer. Nous avons donc réalisé une nouvelle série d'enregistrements dans une autre salle d'écoute. Le panel d'enceintes contenait l'ensemble des enceintes étant intervenues dans les expériences précédentes, mais également de nombreux nouveaux modèles. Il s'agissait en effet d'étudier également l'influence du nombre et de la nature des enceintes impliquées sur le résultat de la comparaison objectif/perceptif.

III.3.1 Enregistrements

Le fabricant d'enceintes Mosquito a mis à notre disposition sa salle d'écoute pour la réalisation des enregistrements de cette expérience. Cette pièce est proche des standards de salle

d'écoute [30][31]. Son plan est présenté sur la Figure III.6. Cette salle a fait l'objet de nombreux traitements acoustiques. Les murs ne sont pas parallèles. Des diffuseurs de Schroeder sont placés derrière et sur le côté des enceintes, ainsi que dans les coins de la salle, derrière la zone d'écoute. Le reste des murs est recouvert soit de bois perforé, soit d'une membrane associée à une dalle minérale absorbante située dans les coins de la salle derrière les enceintes. Le sol est entièrement recouvert de moquette. La durée de réverbération a été évaluée à 0,4 s en moyenne entre 500 et 5000 Hz, pour un volume de l'ordre de 110 m^3 . Les enregistrements ont été réalisés avec la prise de son AB ORTF. Les microphones étaient situés à 1 m au dessus du sol, dans une zone restreinte de l'espace correspondant à la tête d'un auditeur assis à 2,40 m en face de l'enceinte.

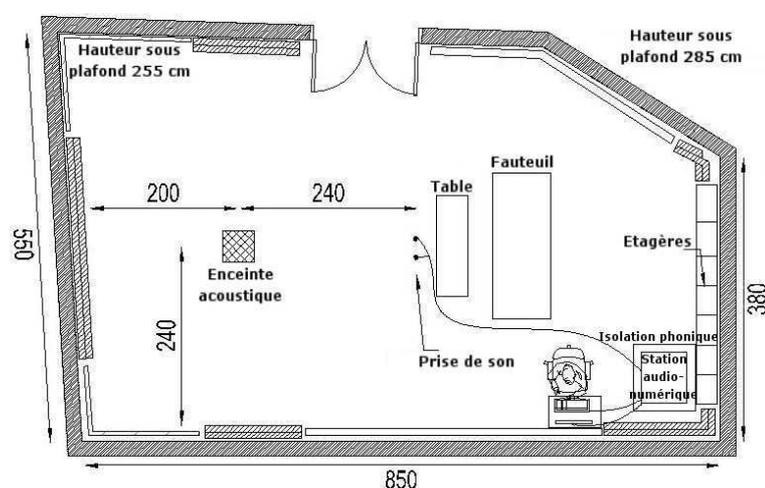


FIGURE III.6 – Salle d'enregistrement et dispositif expérimental utilisés lors de la troisième expérience

Trente-sept enceintes acoustiques ont été utilisées en reproduction monophonique au cours de cinq sessions d'enregistrements. Elles étaient toutes placées au même endroit dans la salle (Figure III.6) et leur position verticale était contrôlée comme lors de la deuxième série d'expériences. Nous avons tenté d'avoir un panel le plus vaste et diversifié possible, couvrant une large gamme de technologies, de prix et de fabricants. Ce panel est présenté en annexe A.2. Nous avons exploité ici un des avantages de notre protocole expérimental basé sur des enregistrements. Il permet d'inclure un nombre très important d'enceintes dans la même expérience, tout en autorisant leur comparaison très rapide.

Les extraits musicaux utilisés lors des enregistrements étaient les mêmes que ceux de la première série d'expériences, mais seul l'extrait Mc Coy Tyner, commun aux expériences précédentes, a été impliqué dans le test psychoacoustique.

III.3.2 Tests d'écoute

Afin de faire évaluer les dissemblances entre un si grand nombre d'enregistrements, un test psychoacoustique basée sur une tâche de classification libre a été mis en place. Le principe de ce test a été décrit au paragraphe II.4.2, page 40. La tâche des auditeurs consistait à grouper les enregistrements en fonction de leur ressemblance. Le niveau perçu de la reproduction au casque était de l'ordre de 65 phons.

Cinquante-six auditeurs ont réalisé la tâche perceptive (trente-quatre hommes et vingt-deux femmes). Ils avaient entre vingt-deux et cinquante-trois ans, avec une moyenne d'âge de trente-deux ans et un écart type de neuf ans. Aucun d'eux n'était entraîné à la comparaison des enceintes acoustiques, même si certains avaient participé aux tests d'écoute de nos expériences précédentes. Les auditeurs ont mis entre dix et quatre-vingt minutes pour effectuer le test, pour une durée moyenne de vingt-huit minutes.

Le nombre de groupes qu'ils ont utilisés lors de la classification libre variait de trois à vingt-cinq, avec une moyenne de huit groupes par auditeurs et un écart type de quatre groupes. Les partitions des auditeurs ont été comparées (paragraphe II.4.2, page 41), mais aucune classe d'auditeurs n'a été mise en évidence par l'analyse en arbre hiérarchique. Comme aucune différence de stratégie de réponse n'est apparue clairement entre les auditeurs lorsqu'ils ont défini leur partitions, ils peuvent avoir simplement utilisé différents degrés de similarité entre les enregistrements pour les placer dans un même groupe. La matrice moyenne des dissemblances sur l'ensemble des auditeurs a été considérée pour la suite des analyses.

III.3.3 Corrélation entre dissemblances objectives et perceptives

La Figure III.7 illustre la comparaison directe des dissemblances perceptives avec les différentes dissemblances objectives calculées sur nos signaux. Elle présente les corrélations moyennes obtenues à partir des deux voies de nos enregistrements stéréophoniques. En plus de l'analyse complète des dissemblances portant sur la totalité des trente-sept enregistrements, nous avons effectué une analyse partielle en ne conservant que les dissemblances impliquant les onze enceintes qui avaient déjà été utilisées lors de notre première série d'expériences. Notre but était de pouvoir ainsi comparer deux expériences faisant intervenir les mêmes enceintes acoustiques utilisées dans deux pièces différentes, les enregistrements étant réalisés avec la même prise de son et le même extrait musical. Les dissemblances perceptives correspondant à ces enregistrements ont simplement été extraites de la matrice de dissemblances complète obtenue lors du test psychoacoustique. Aucun nouveau test n'a été réalisé. Bien sûr, l'influence des autres stimuli impliqués dans le test d'écoute n'a pas pu être éliminée a posteriori, et ceci doit être gardé à l'esprit au moment de l'interprétation des résultats.

Les résultats obtenus à partir de ces deux analyses confirment les tendances observées lors de

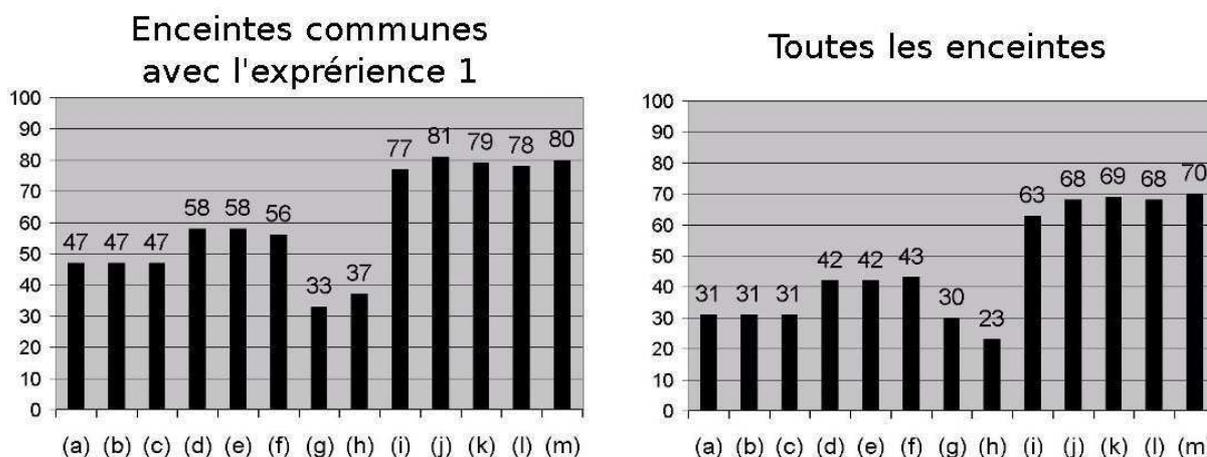


FIGURE III.7 – *Troisième expérience. Corrélations moyennes entre les dissemblances perceptives et objectives (en %) : (a) domaine temporel (b) spectre (c) temps-fréquence (d) module du spectre (e) module du temps-fréquence (f) densité spectrale de puissance (g) pondération A (h) pondération Iso70 (i) densité de sonie totale (j) densité de sonie 1 (k) densité de sonie 1 moyenne (l) densité de sonie 2 (m) densité de sonie 2 moyenne*

nos deux premières séries d'expériences. La comparaison de l'histogramme résultant de l'analyse partielle avec ceux obtenus dans l'autre salle d'écoute (Figure III.3, page 65) indique que la pièce n'influence pas le rapprochement objectif/perceptif mis en évidence par notre protocole expérimental. L'intérêt de considérer les densités de sonie des signaux est confirmé quand un nombre beaucoup plus important d'enceintes est considéré, même si les valeurs de corrélation entre les dissemblances perceptives et les dissemblances objectives basées sur les densités de sonie sont moins élevées.

III.4 Bilan de la comparaison directe des dissemblances

Notre protocole expérimental nous a permis d'évaluer un très grand nombre d'enceintes acoustiques dans deux salles d'écoute différentes, à l'aide de plusieurs prises de son et de plusieurs extraits musicaux. Différents modes de reproduction des enceintes ont également été considérés.

En définissant une métrique associée à chacune des analyses objectives testées, nous pouvons calculer des dissemblances objectives entre nos enregistrements. Calculer la corrélation entre ces dissemblances objectives et les dissemblances perceptives résultant des tests psychoacoustiques nous permet de quantifier la ressemblance "linéaire" entre les deux types d'évaluation. Ce critère de corrélation est sévère puisqu'en cherchant à rapprocher ces évaluations, nous ne pouvons pas supposer qu'il doit exister une relation linéaire entre les dissemblances objectives et perceptives. Les valeurs de corrélation obtenues nous indiquent cependant qu'un rapprochement a au moins

en partie été réalisé et quantifié. Nous avons ainsi pu aller un peu plus loin que la corrélation "visuelle" des résultats de tests d'écoute avec les formes de réponses en fréquence des enceintes, cette forme de comparaison présentée au chapitre I étant très souvent rencontrée dans la littérature. En contrepartie, nous avons dû conserver des approches objectives et perceptives très proches. Nous travaillons ainsi sur les signaux audio écoutés par les auditeurs lors des tests, et non pas sur des mesures physiques des enceintes qui seraient plus caractéristiques du système mesuré et moins dépendantes du signal de mesure.

Nos évaluations portent sur des enceintes utilisées dans une salle selon un certain mode de reproduction. Elles sont effectuées avec une prise de son et un extrait musical donnés. C'est l'ensemble de cette chaîne de reproduction-enregistrement qui est évalué, et bien sûr chacun des paramètres expérimentaux qui la constituent influence cette évaluation. L'effet lié à l'extrait musical et à la salle d'écoute sur la perception du champ sonore produit par des enceintes acoustiques a été démontré dans des études antérieures (paragraphe I.2, page 6). Nous avons trouvé une influence faible mais statistiquement significative du mode de reproduction des enceintes. Dans le cadre de notre protocole, il semble également raisonnable de penser que la perception des différences entre les enregistrements est aussi dépendante de la prise de son utilisée, mais nos expériences ne nous ont pas permis de quantifier cette éventuelle influence. Par contre, nous avons montré que la comparaison des corrélations entre les dissemblances perceptives et les différentes dissemblances objectives donne toujours les mêmes tendances et reste relativement indépendante des différents paramètres expérimentaux testés. La stabilité de notre protocole expérimental vis-à-vis de ces paramètres est primordiale et doit être établie avant que son utilisation puisse être exploitable.

La comparaison des corrélations entre les dissemblances objectives et perceptives, pour les analyses objectives et les métriques testées, nous a permis de mettre en évidence l'avantage de considérer les densités de sonie de nos signaux. En ayant recours à un modèle auditif plutôt que de conserver des analyses objectives classiques, nous nous rapprochons des résultats fournis par l'approche perceptive, et donc de notre but visant à obtenir une mesure objective plus pertinente des différences entre enceintes acoustiques. Cet avantage nécessite bien sûr d'être confirmé par des recherches complémentaires. D'autres méthodes d'analyses et d'autres définitions de métrique doivent être testées. Cet intérêt d'utiliser des modèles auditifs dans le cadre de l'évaluation de la reproduction sonore était déjà présent dans la littérature, mais à notre connaissance il n'avait pas été quantifié. Staffeldt [26][43][63] justifie l'utilisation d'un tel modèle par des corrélations "visuelles" entre le résultat de ses tests d'écoute et la forme des réponses en fréquence de ses enceintes. Klippel [10][48] et Bramsløw [9] utilisent la densité de sonie de leurs signaux comme base de leur méthode d'évaluation objective de la reproduction sonore. Ils choisissent ce type d'analyse a priori, en émettant l'hypothèse qu'elle donnera de meilleurs résultats que les techniques d'analyse classiques. Ils obtiennent effectivement de bons résultats, mais sans avoir testé

leur hypothèse de départ.

Pour compléter cette comparaison directe des dissemblances objectives et perceptives basée sur l'évaluation de leur corrélation, le chapitre suivant présente une comparaison cherchant à tenir compte du caractère multidimensionnel de notre perception de la reproduction sonore.

CHAPITRE IV

Comparaison multidimensionnelle des dissemblances : espaces perceptifs et objectifs

————— oOo —————

Parmi les analyses objectives que nous avons testées, les densités de sonie fournissent les dissemblances les plus corrélées avec les dissemblances perceptives. Il s'agit ensuite de savoir si elles sont suffisantes pour retrouver complètement ces dissemblances perceptives. Nous avons pour cela effectué une comparaison multidimensionnelle des dissemblances évaluées selon les approches perceptive et objective, en considérant les espaces résultant de l'analyse multidimensionnelle de ces dissemblances.

La comparaison des espaces que nous avons utilisée est pour le moment simplement "visuelle". Il sera nécessaire de définir un critère qui permette de quantifier la ressemblance entre deux espaces. La comparaison visuelle de deux espaces multidimensionnels semble tout de même moins délicate que celle de réponses fréquentielles d'enceintes variant simultanément selon de multiples critères.

La comparaison effectuée avait pour but de répondre à plusieurs questions. Quelles dimensions perceptives sont responsables des différences perçues par les auditeurs entre nos enregistrements? L'analyse objective permet-elle de retrouver ces dimensions? Si oui, le rapprochement entre dissemblances objectives et perceptives est-il conditionné par les différents paramètres expérimentaux? Enfin, notre interprétation de la nature des dissemblances perceptives dépend-elle de ces mêmes paramètres expérimentaux?

————— oOo —————

SOMMAIRE

IV.1 Influence de l'extrait musical et de la prise de son	81
IV.2 Influence du mode de reproduction des enceintes	86
IV.3 Influence de la pièce	89
IV.4 Conclusions intermédiaires	91
IV.4.1 Un seul espace, deux dimensions	91
IV.4.2 Bilan de la comparaison objectif/perceptif pour ces deux dimensions	92
IV.5 Influence des stimuli utilisés	94
IV.5.1 Influence de la tâche perceptive?	96
IV.5.2 Limitations de l'approche objective pour décrire les quatre dimensions	98
IV.6 Attributs objectifs préliminaires	99
IV.6.1 Balance grave-aigu	100
IV.6.2 Emergence du bas-médium	103
IV.6.3 Bilan sur les attributs objectifs préliminaires	106
IV.7 Bilan de la comparaison multidimensionnelle des dissemblances	107

IV.1 Première série d'expériences : influence de l'extrait musical et de la prise de son

Nous reprenons ici les données brutes de notre première série d'expériences impliquant trois extraits musicaux (Kan'nida, Mc Coy Tyner et Vivaldi) et trois prises de son (AB ORTF, MS et omnidirectionnelle). Avant de réaliser les analyses multidimensionnelles des matrices de dissemblances entre les enregistrements, nous avons vérifié qu'il n'existait pas de catégorie parmi ces enregistrements, grâce à une analyse en arbre hiérarchique de leurs dissemblances perceptives. La Figure IV.1 présente l'exemple du dendrogramme obtenu après analyse des résultats du test faisant intervenir l'extrait musical Kan'nida et la prise de son AB ORTF. Les dendrogrammes correspondant aux autres tests sont présentés en annexe C.4. Aucune de ces analyses n'a révélé de groupements nets entre les enregistrements. La Figure IV.1 indique que les enregistrements impliquant les enceintes 2, 5 et 12 sont jugés bien différents des autres, mais ils ne se ressemblent pas spécialement entre eux et ne forment donc pas une catégorie à part. Il peut être constaté sur ce dendrogramme que les enceintes 7 et 8, pourtant du même modèle, n'ont pas conduit aux reproductions jugées les plus semblables.

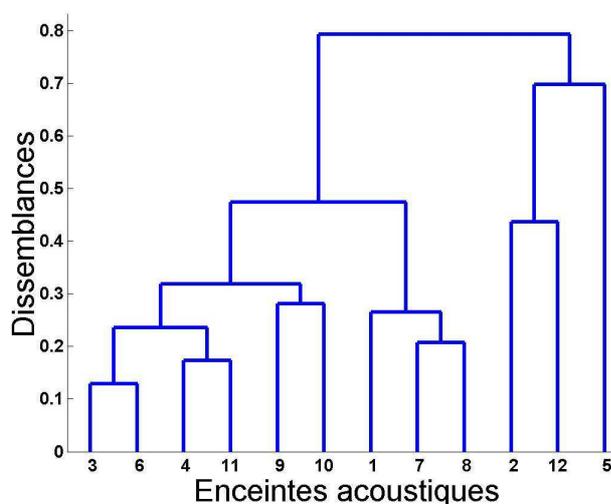


FIGURE IV.1 – Exemple de dendrogramme résultant de l'analyse en arbre hiérarchique des dissemblances entre enregistrements (Kan'nida, couple AB ORTF, première série d'expériences). Les enceintes 7 et 8 sont du même modèle.

L'analyse multidimensionnelle des dissemblances perceptives entre les enregistrements a ensuite été effectuée. Le principe de cette analyse et le programme utilisé pour la réaliser sont décrits dans l'annexe B.5. Afin de déterminer le nombre de dimensions le plus approprié pour décrire nos données, nous nous sommes basés sur les courbes de stress [39], l'interprétabilité des dimensions et les commentaires fournis par les auditeurs après les tests.

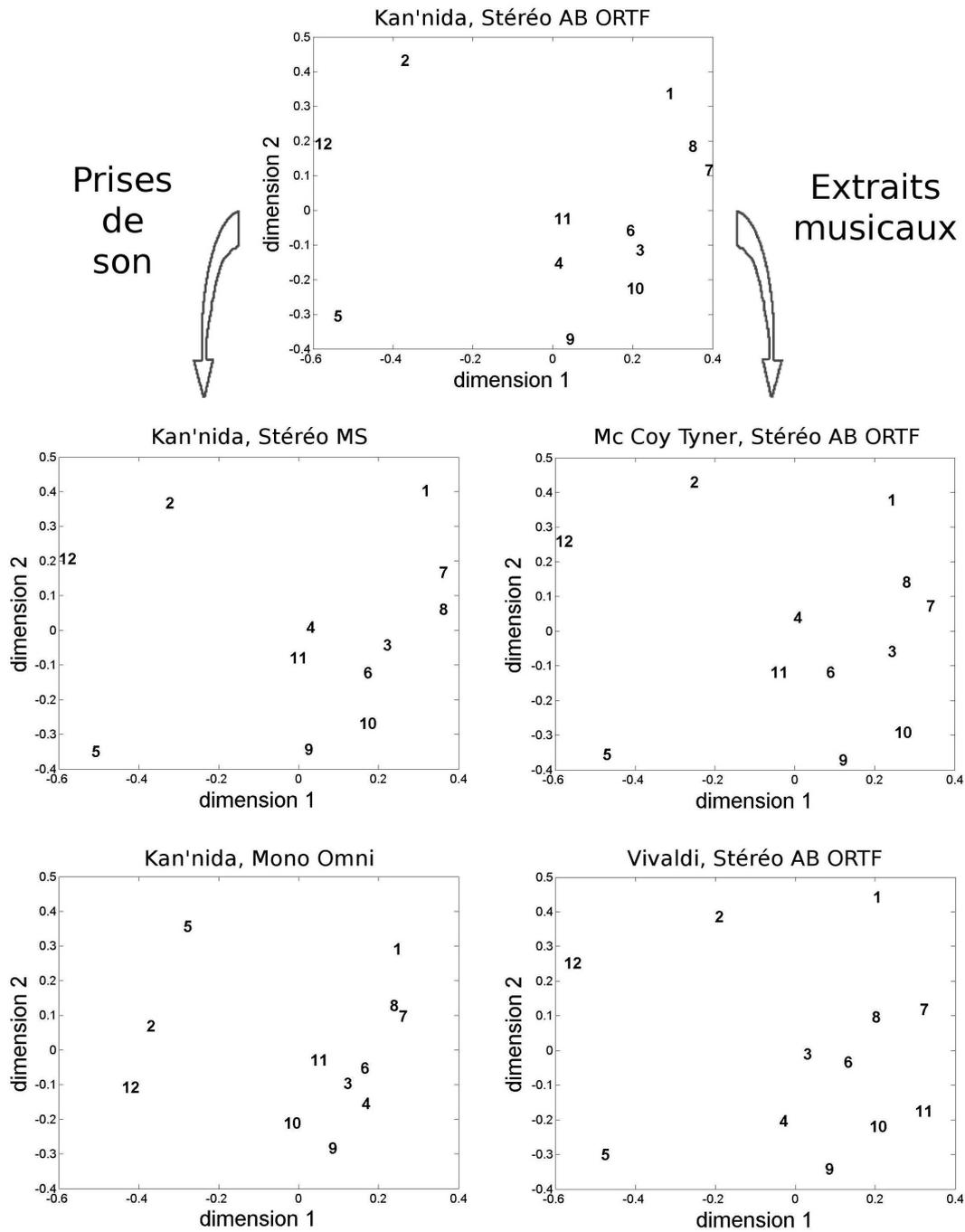


FIGURE IV.2 – *Espaces perceptifs obtenus lors de notre première série d'expériences impliquant trois prises de son et trois extraits musicaux. Chaque numéro correspond à une enceinte acoustique. Les enceintes 7 et 8 sont du même modèle.*

La Figure IV.2 présente les résultats obtenus pour les cinq tests psychoacoustiques de notre première série d'expériences. Les espaces perceptifs sont tous bidimensionnels, et montrent ainsi

que les auditeurs ont utilisé deux attributs perceptifs principaux pour différencier les enregistrements lors des tests psychoacoustiques. Chacun des douze numéros contenus dans ces espaces correspond à une enceinte acoustique.

Les axes de ces espaces ne sont pas fixés par notre programme d'analyse. Chaque espace peut être tourné ou dilaté. L'information importante réside dans les distances relatives entre les points représentant les stimuli. Par conséquent, afin de pouvoir plus facilement comparer les espaces perceptifs obtenus à partir des différents tests, nous les avons tous orientés et dilatés pour qu'ils coïncident au mieux avec un même espace de référence. L'espace correspondant à l'extrait musical Kan'nida et à la prise de son AB ORTF a été choisi arbitrairement comme référence pour cette première série d'expériences. Nous avons conservé l'orientation de ses axes obtenue par défaut avec notre programme. Cette orientation repose sur une analyse en composantes principales.

Il est important de rappeler que, comme toute mesure, la mesure psychoacoustique est bruitée. Ce bruit de mesure peut être mis en évidence par une analyse en bootstrap (annexe B.5.4). Lors de la comparaison des espaces, il faut donc garder à l'esprit que chaque stimulus est davantage représenté par une zone de l'espace plutôt que par un point précis. Il est également nécessaire de comparer les positions relatives des stimuli et non leur position absolue. En effet, ces espaces ont été obtenus pour des tests indépendants les uns des autres. D'un test à l'autre, les auditeurs n'ont pas forcément conservé la même échelle pour évaluer les dissemblances. Des différences de jugement liées à la prise de son ou à l'extrait musical ne peuvent donc pas être mises en évidence directement par des différences de positions absolues. Par contre, les différences de positions relatives des stimuli entre les espaces peuvent témoigner de l'influence de la prise de son ou de l'extrait musical utilisés, même si ces modifications peuvent également être liées au bruit de mesure.

Au regard de ces cinq espaces perceptifs, seule la position relative de l'enceinte 5 est considérablement modifiée, et cela uniquement au sein de l'espace correspondant à la prise de son monophonique omnidirectionnelle (Figure IV.2). Nous verrons lors de l'analyse des résultats de notre deuxième série d'expériences que ce comportement particulier ne semble pas dû à une différence d'évaluation des dissemblances perceptives lors du test utilisant les enregistrements monophoniques, mais plutôt à un artefact de représentation dû à une limitation de notre programme d'analyse multidimensionnelle.

Finalement, l'analyse des résultats de notre première série de tests psychoacoustiques montre que nous obtenons à peu près le même espace perceptif à deux dimensions, pour les trois prises de son et les trois extraits musicaux testés. La prise de son et l'extrait musical peuvent avoir une influence sur l'évaluation des dissemblances perceptives, mais cette influence reste limitée étant données les faibles modifications de positions relatives des stimuli entre les différents espaces. Il semble que l'évaluation de dissemblances fasse intervenir les mêmes dimensions perceptives malgré les modifications apportées à ces deux facteurs expérimentaux.

Il est intéressant de noter que les enceintes 7 et 8 qui sont du même modèle se retrouvent très proches dans tous les espaces perceptifs, même si elles ne sont pas forcément toujours les enceintes les plus proches.

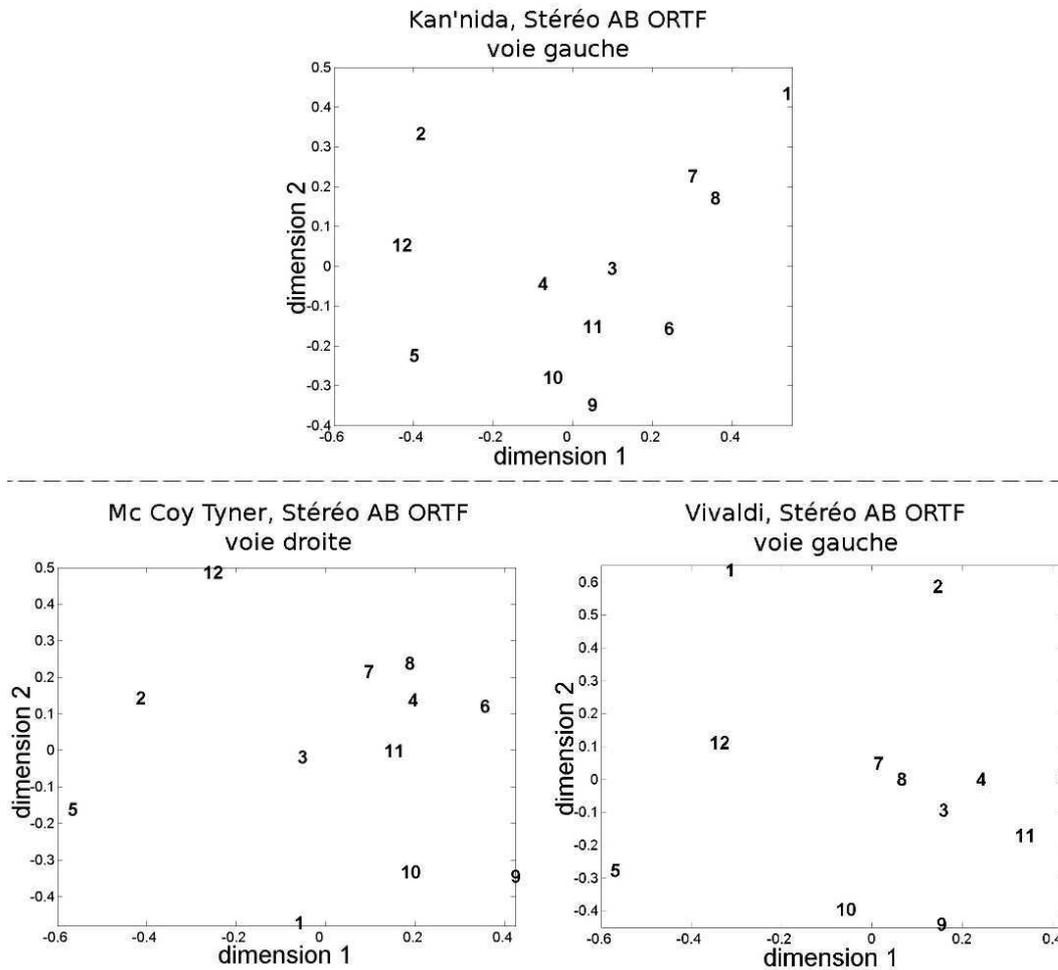


FIGURE IV.3 – Exemples d'espaces objectifs déterminés à partir de la densité de sonie totale de nos signaux. Ils sont à comparer aux espaces perceptifs de la Figure IV.2.

De manière analogue aux dissemblances perceptives, les dissemblances objectives provenant des différentes techniques d'analyse testées ont été soumises à l'analyse multidimensionnelle. Les espaces objectifs résultant de cette analyse ont été orientés et dilatés par rapport à la même référence que les espaces perceptifs, afin de faciliter leur comparaison. Etant donné la quantité importante d'espaces objectifs obtenus, seulement certains espaces seront présentés à titre d'exemple, afin d'illustrer les différentes tendances observées sur l'ensemble de nos enregistrements. Seules les analyses basées sur les différentes densités de sonie nous ont permis de retrouver les deux dimensions perceptives au niveau des espaces objectifs pour la totalité de nos enregistrements. Les tendances observées à partir des autres techniques d'analyse sont décrites dans

l'annexe C.5.

La densité de sonie totale permet de retrouver des espaces objectifs proches des espaces perceptifs pour les types d'enregistrements impliquant l'extrait musical Kan'nida. Les résultats sont par contre beaucoup moins satisfaisants pour l'analyse des enregistrements correspondant aux deux autres extraits musicaux. La Figure IV.3 fournit, pour chaque extrait musical impliqué dans les enregistrements, un exemple d'espace objectif obtenu à partir de la densité de sonie totale. Ces espaces sont à comparer aux espaces perceptifs correspondants (Figure IV.2). L'analyse globale sur l'ensemble du signal effectuée lors de la détermination de la densité de sonie totale des enregistrements n'apparaît pas suffisante.

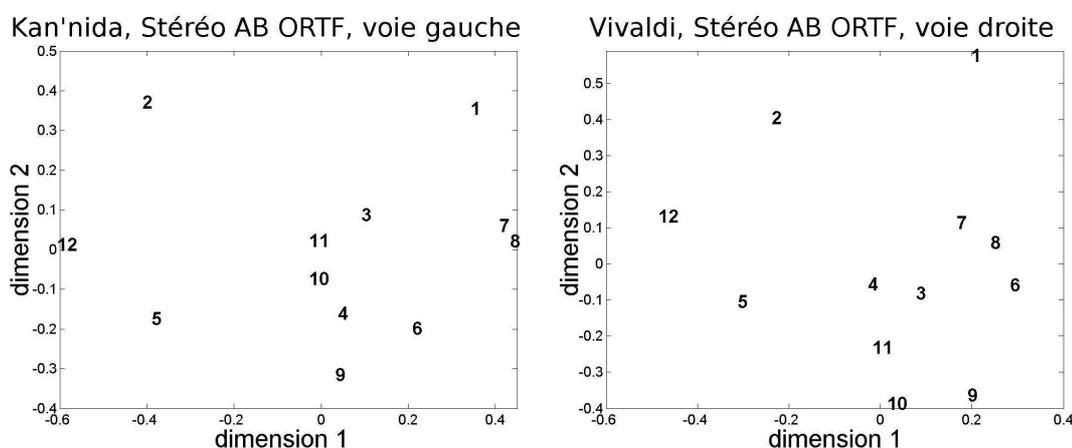


FIGURE IV.4 – Exemples d'espaces objectifs déterminés à partir de la densité de sonie 2 moyenne de nos signaux. Ils sont à comparer aux espaces perceptifs de la Figure IV.2.

Les densités de sonie 1 et 2 et leur moyennes temporelles donnent par contre des résultats intéressants pour tous nos signaux. Ces résultats sont équivalents pour les modèles 1 et 2, et pour les densités de sonie variant dans le temps et leur moyennes temporelles. Dans tous les cas et pour tous nos types d'enregistrements, nous obtenons un espace objectif bidimensionnel qui décrit bien l'espace perceptif correspondant. La Figure IV.4 présente les exemples obtenus à partir des signaux qui avaient précédemment conduit à la meilleure (Kan'nida, AB ORTF, voie gauche) et la moins bonne (Vivaldi, AB ORTF, voie droite) corrélation entre dissemblances objectives et perceptives (annexe C.2). Lors du choix des espaces à représenter, nous avons décidé de considérer la densité de sonie moyenne plutôt que la densité de sonie variant dans le temps, car elle utilise a priori moins d'information pour d'aussi bons résultats. La densité de sonie 2 moyenne a été préférée à la densité de sonie 1 moyenne car elle est basée sur un programme plus répandu [29].

Les phénomènes d'étalement spectral liés au système auditif et modélisés lors du calcul des différentes densités de sonie dépendent du contenu spectral des signaux considérés, ce contenu

pouvant évoluer au cours du temps. Alors que pour la densité de sonie totale, le modèle auditif est appliqué de façon globale sur l'ensemble du signal considéré comme stationnaire, le calcul des densités de sonie 1 et 2 est effectué en fonction du temps. Par rapport à la densité de sonie totale, ces analyses tiennent compte du fait que les phénomènes d'étalement spectral liés au système auditif dépendent de l'évolution temporelle des signaux non-stationnaires considérés. Cette première série d'expériences semble indiquer qu'il est nécessaire de prendre en compte cette évolution temporelle lors de l'évaluation objective de la reproduction sonore.

IV.2 Deuxième série d'expériences : influence du mode de reproduction des enceintes

L'ensemble des analyses précédentes a ensuite été appliqué à notre deuxième série d'expériences impliquant l'extrait musical Mc Coy Tyner, la prise de son AB ORTF et trois modes de reproduction pour les enceintes (monophonique, double monophonique et stéréophonique). Les dissemblances correspondant aux différents modes de reproduction, évaluées au cours de chacun des deux tests psychoacoustiques, ont été analysées séparément. Nous avons vérifié qu'il n'existait pas de catégorie parmi les nouveaux enregistrements. L'analyse multidimensionnelle des dissemblances perceptives a permis de retrouver des espaces perceptifs bidimensionnels (Figure IV.5). Ces espaces ont été tournés et dilatés pour coïncider avec l'espace perceptif impliquant le même extrait musical et la même prise de son lors de notre première série d'expériences. Ce dernier est à nouveau présenté sur la Figure IV.5 pour permettre la comparaison de ces deux expériences utilisant en majorité les mêmes enceintes.

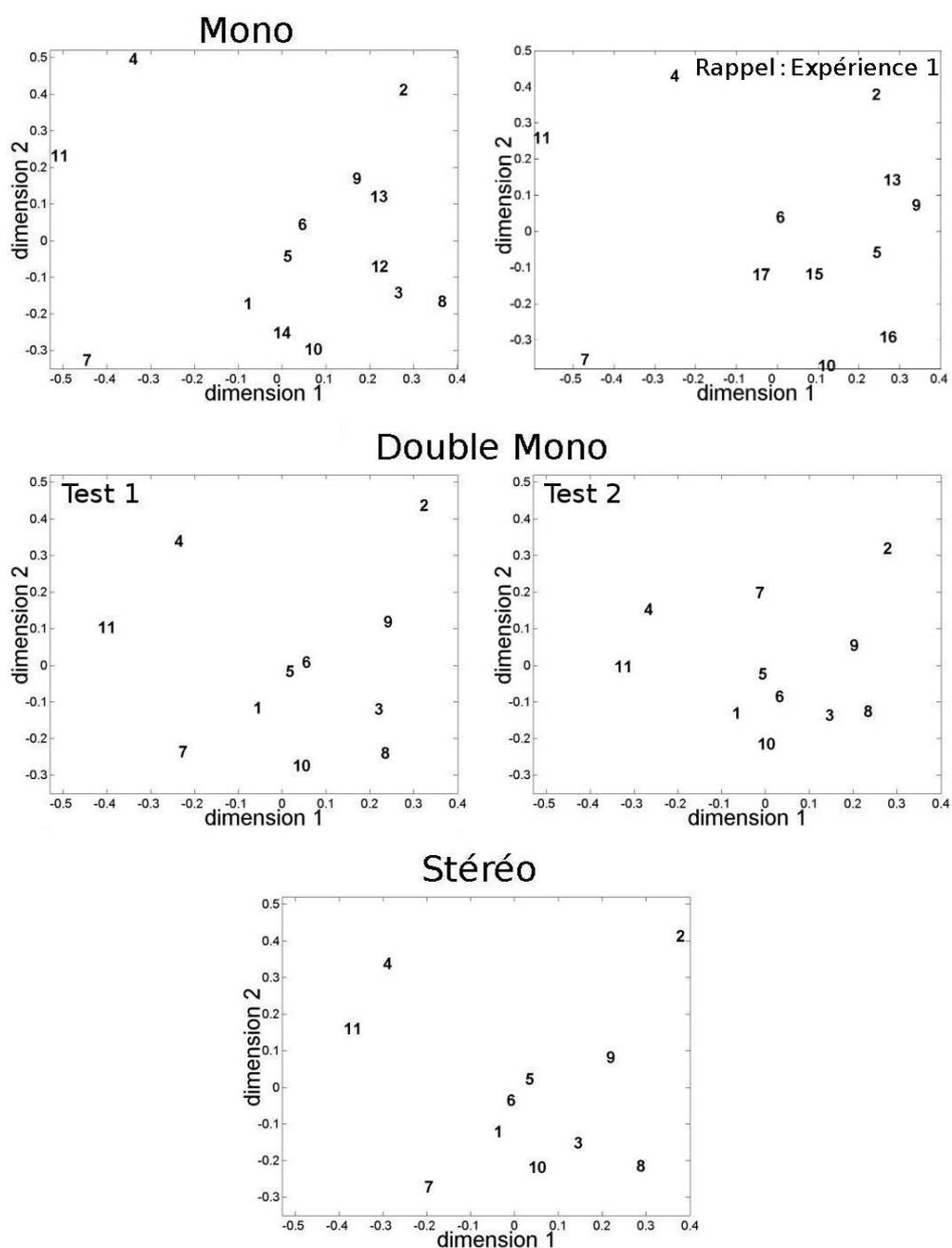


FIGURE IV.5 – *Espaces perceptifs obtenus lors de notre deuxième série d'expériences impliquant trois modes de reproduction. Chaque numéro correspond à une paire d'enceintes acoustiques (Double Mono, Stéréo) ou à l'enceinte droite de cette paire (Mono). Les enceintes 12, 13 et 14 sont respectivement les enceintes gauches des paires 3, 9 et 10. L'espace perceptif obtenu lors de notre première expérience, avec le même extrait musical (Mc Coy Tyner) et la même prise de son (AB ORTF), est présenté pour comparaison, la numérotation des enceintes ayant été modifiée pour faciliter cette comparaison.*

Les dissemblances entre les enregistrements de la reproduction double monophonique ont été évaluées deux fois, une première fois en même temps que celles correspondant à la reproduction monophonique lors du test 1, et une seconde fois avec celles correspondant à la reproduction stéréophonique lors du test 2. Les deux espaces perceptifs obtenus à partir de ces deux évaluations sont très similaires (Figure IV.5). Ils permettent ainsi de vérifier que les évaluations de dissemblances ont été reproductibles lors de nos tests psychoacoustiques, et qu'elles n'ont pas été dépendantes de la nature des autres stimuli impliqués dans le test. Seule la position relative de la paire d'enceintes 7 est considérablement modifiée sur l'espace correspondant au test 2. Ces enceintes correspondent à l'enceinte 5 de la Figure IV.2. L'analyse multidimensionnelle avait déjà révélée ce comportement singulier lors de notre première série d'expériences. Comme les enregistrements de la reproduction double monophonique ont été les mêmes pour les deux tests d'écoute, les résultats de ces deux tests peuvent être comparés directement. Les dissemblances perceptives impliquant la paire d'enceintes 7 sont très corrélées d'un test à l'autre, avec un coefficient de corrélation au dessus de 0,98. L'instabilité de la position de cette paire d'enceintes au sein des espaces perceptifs semble donc due à l'analyse multidimensionnelle, et non pas à une différence d'évaluation des dissemblances perceptives. C'est a priori un artefact de représentation dû à une limitation de notre programme d'analyse multidimensionnelle. La paire d'enceintes 7 peut en effet avoir une caractéristique particulière, entraînant une particularité perceptive des enregistrements correspondants. Le programme d'analyse multidimensionnelle que nous avons utilisé ne cherche à représenter que les dimensions perceptives communes aux différents stimuli et ne permet pas de prendre en compte une dimension particulière à un stimulus. Cette limitation de notre programme pourrait être corrigée par l'utilisation de programmes d'analyse multidimensionnelle plus poussés permettant la prise en compte de spécificités pour les stimuli (annexe B.5.2).

La Figure IV.5 montre finalement que la position des enceintes est très stable au sein des espaces perceptifs, qui sont très similaires pour les trois modes de reproduction considérés. Ces espaces ne semblent pas être affectés par l'effet du changement de mode reproduction révélé au chapitre précédent par l'analyse de variance. Cet effet était d'ailleurs prédit comme très faible. Il peut expliquer les légères modifications de positions relatives de certains stimuli entre les différents espaces, mais il ne modifie pas les dimensions perceptives principales impliquées dans les jugements des auditeurs lors des tests d'écoute. La Figure IV.5 montre également que ces dimensions sont très probablement les mêmes que les deux dimensions révélées lors de notre première série d'expériences.

Les dissemblances objectives déterminées sur les enregistrements de cette deuxième série d'expériences ont à leur tour été soumises à l'analyse multidimensionnelle. Les espaces objectifs résultant de cette analyse ont été orientés et dilatés pour coïncider au mieux avec l'espace perceptif correspondant, afin de faciliter leur comparaison. Comme lors de notre première série d'expériences, seules les densités de sonie nous ont permis de retrouver les dimensions percep-

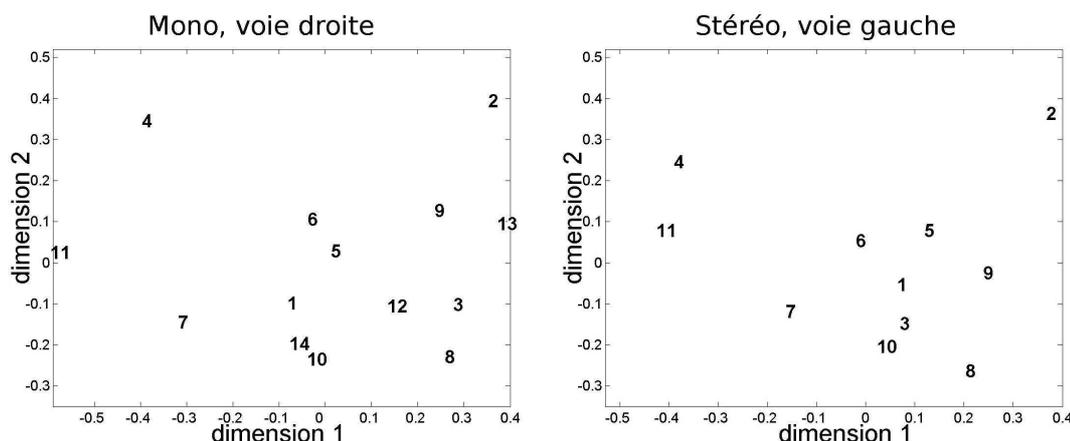


FIGURE IV.6 – Exemples d’espaces objectifs déterminés à partir de la densité de sonie 2 moyenne de nos signaux. Ils sont à comparer aux espaces perceptifs de la Figure IV.5.

tives au niveau des espaces objectifs. La densité de sonie totale a par contre cette fois eu des performances équivalentes à celles des autres densités de sonie. Les résultats sont à nouveau satisfaisants pour tous nos types d’enregistrements. Les espaces déterminés à partir de la densité de sonie 2 moyenne des signaux ayant précédemment conduit à la meilleure (mode Stéréo, voie gauche) et la moins bonne (mode Mono, voie droite) corrélation entre dissemblances objectives et perceptives sont présentés sur la Figure IV.6.

IV.3 Troisième expérience : influence de la pièce

Nous avons finalement réalisé l’analyse partielle des dissemblances perceptives mesurées lors de notre troisième expérience. Cette expérience a fait intervenir trente-sept enceintes dans une nouvelle salle d’écoute, le test psychoacoustique étant basé sur une tâche de classification libre. L’analyse partielle consiste à ne considérer que les dissemblances impliquant les onze enceintes qui avaient déjà été utilisées lors de notre première série d’expériences. Les résultats de cette analyse sont comparés à ceux de l’analyse partielle de notre première expérience, impliquant les mêmes onze enceintes dans une autre pièce, et utilisant le même extrait musical et la même prise de son. Notre but est ainsi d’estimer l’influence de la pièce sur les évaluations de dissemblances.

Aucune catégorie n’a été décelée parmi les enregistrements considérés. L’analyse multidimensionnelle des dissemblances perceptives a permis de retrouver à nouveau un espace perceptif bidimensionnel (Figure IV.7). L’espace perceptif résultant de l’analyse partielle des dissemblances provenant de la première expérience a également été déterminé, puis orienté et dilaté par rapport à l’espace précédent pour faciliter leur comparaison. Comme lors des deux précédentes séries d’expériences, nous observons un mouvement de l’enceinte 5 au sein de l’espace perceptif. Cette

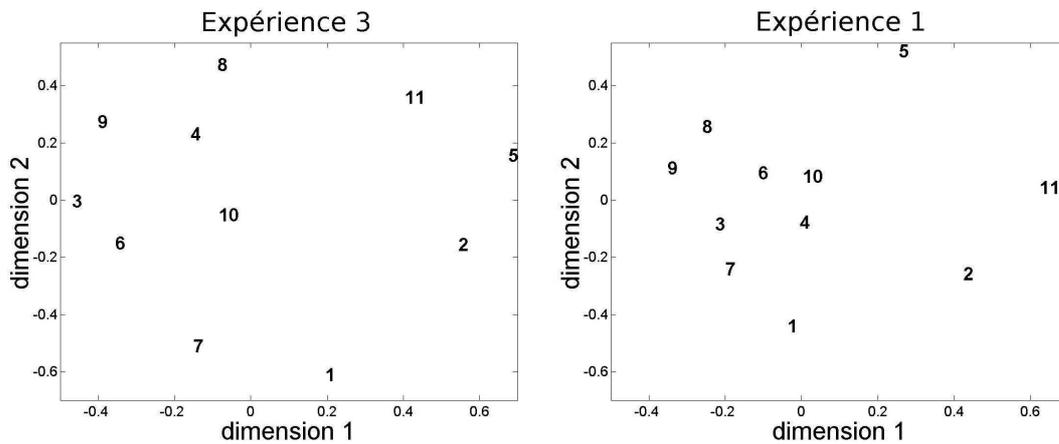


FIGURE IV.7 – *Espace perceptif obtenu lors de l'analyse partielle de notre troisième expérience. Chaque numéro correspond à une enceinte acoustique. L'espace perceptif correspondant à l'analyse partielle de notre première expérience impliquant les mêmes enceintes dans une autre pièce est présenté pour comparaison.*

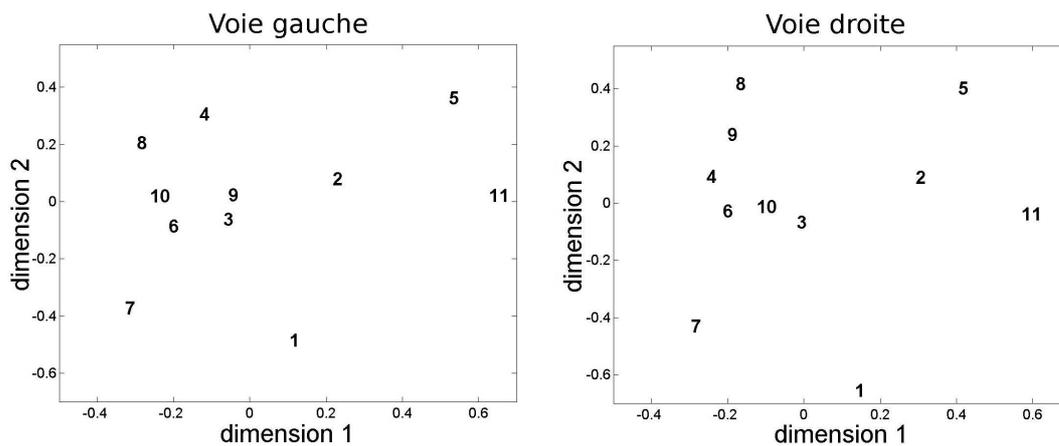


FIGURE IV.8 – *Espaces objectifs déterminés à partir de la densité de sonie 2 moyenne de chacune des deux voies de nos signaux. Ils sont à comparer à l'espace perceptif de la Figure IV.7 correspondant à cette troisième expérience et ayant été utilisé comme référence pour orienter et dilater les espaces objectifs.*

enceinte est numérotée 7 sur la Figure IV.5 et 5 sur la Figure IV.2. Ce comportement particulier a déjà été interprété dans le paragraphe précédent. Il est vraisemblablement lié à un artefact de représentation dû à notre programme d'analyse multidimensionnelle qui ne permet pas la prise en compte de dimension particulière à un stimulus. Les changements de pièce et de tâche perceptive, ainsi que la présence des nombreux autres stimuli impliqués dans le test d'écoute de la troisième expérience, peuvent expliquer les modifications de positions relatives de certains

stimuli. Les deux dimensions perceptives principales mises en jeu lors des tests d'écoute restent pourtant assez similaires, malgré les modifications apportées à tous ces facteurs.

L'analyse multidimensionnelle a ensuite été appliquée sur les dissemblances objectives déterminées sur les deux voies de nos enregistrements stéréophoniques. A nouveau, seuls les espaces objectifs obtenus à partir des densités de sonie nous ont permis de retrouver les dimensions de l'espace perceptif correspondant. La densité de sonie totale a fourni des résultats équivalents à ceux des autres densités de sonie. La Figure IV.8 présente les espaces déterminés à partir de la densité de sonie 2 moyenne.

IV.4 Conclusions intermédiaires

IV.4.1 Un seul espace, deux dimensions

L'ensemble des évaluations perceptives qui ont impliqué une douzaine d'enceintes et que nous avons réalisées dans différentes configurations expérimentales ont toutes abouti à la même conclusion : au moins deux dimensions perceptives principales ont été utilisées par les auditeurs pour différencier les enregistrements durant les tests psychoacoustiques, et ces dimensions étaient vraisemblablement les mêmes pour tous les tests. Bien sûr la position exacte des enceintes sur ces dimensions dépend des différents paramètres expérimentaux. L'extrait musical, la prise de son, le mode de reproduction des enceintes et la pièce ont tous un effet sur le signal soumis aux tests d'écoute. Leur influence est peut être révélée par les modifications de position relative des stimuli au sein des espaces perceptifs, mais ces modifications peuvent également être liées au bruit de mesure. Comme toute autre mesure, l'évaluation psychoacoustique étant à chaque fois effectuée dans des conditions différentes, elle ne conduit pas toujours à des résultats identiques. La détermination des conditions d'évaluation optimales pourra faire l'objet de recherches ultérieures. Les dimensions perceptives étant quant à elles relativement indépendantes de tous ces facteurs expérimentaux, notamment de l'extrait musical utilisé, elles semblent bien être caractéristiques des enceintes.

Au regard des indications fournies par les auditeurs lors des entretiens informels ayant suivi chaque test psychoacoustique, et après une écoute des enregistrements guidée par les espaces perceptifs obtenus, la première dimension perceptive qui est prépondérante semble déterminée par la balance spectrale ou l'équilibre grave-aigu des enregistrements. La deuxième dimension plus tenue serait liée à la présence plus ou moins importante de médium, et pourrait être associée à une notion de clarté du son. Ces interprétations des dimensions perceptives ont été confirmées lors de la recherche d'attributs objectifs ayant pour but de décrire ces dimensions (paragraphe IV.6, page 99).

Les dimensions mises en jeu dans nos tests sont comparables aux deux dimensions révé-

lées par les comparaisons par paires de Gabrielsson *et coll.* [19]. Les auteurs parlent en effet de "brightness-darkness" ou "balance between bass and treble" pour une dimension, et de "distinctness" ("clarity", "clearness") pour l'autre. Ces deux dimensions avaient également la même interprétation pour les trois extraits musicaux employés dans leurs tests. Nos dimensions peuvent aussi être comparées à d'autres résultats de la littérature. Gabrielsson et Sjögren [7] mettent en évidence la dimension "fullness-thinness", que Toole [14] utilise et définit plus tard comme relative à la quantité de basses fréquences par rapport celle des moyennes et hautes fréquences. Staffeldt [26] parle de "emphasized treble" et "emphasized bass" pour interpréter ces résultats. Les tests de Klippel [10] révèlent les critères "clearness" et "treble stressing" ("sharpness"), "general bass emphasis" ("volume"), "low bass emphasis", "brightness". Pour Olive [46], la balance spectrale "perçue" est un facteur important relié aux préférences des auditeurs. Une étude sémantique effectuée sur les commentaires libres de ses auditeurs lui révèle neuf facteurs, dont il interprète les deux principaux par les termes "bass quality, overall spectral balance and smoothness" et "mid-treble-colorations like bright, mid peak (megaphone) and colored". Finalement Bramsløw [9] met en évidence deux dimensions pertinentes du point de vue de la qualité de la reproduction sonore : "clearness" et "sharpness/fullness", et indique que cette deuxième dimension peut être interprétée comme le rapport du contenu des basses et des hautes fréquences.

Nos tests d'écoute n'ont pas révélé de nouvelles dimensions de la reproduction sonore, ce qui n'était pas leur but, par contre nos résultats semblent être en accord avec ceux de la littérature.

IV.4.2 Bilan de la comparaison objectif/perceptif pour ces deux dimensions

Comme la comparaison directe des dissemblances objectives et perceptives présentée au chapitre précédent, leur comparaison multidimensionnelle s'est avérée relativement indépendante des différents paramètres expérimentaux, les mêmes tendances ayant été observées dans toutes les configurations. La stabilité de notre protocole expérimental semble donc confirmée avec ce nouveau type de comparaison.

Parmi les techniques d'analyse que nous avons testées, seules les densités de sonie permettent de retrouver les dimensions perceptives principales impliquées dans les jugements des auditeurs. Pour tous nos types d'enregistrements, elles conduisent à un espace objectif bidimensionnel qui décrit assez bien l'espace perceptif correspondant. Les densités de sonie apparaissent comme des méthodes d'analyse pertinentes pour évaluer objectivement les deux dimensions perceptives mises en jeu lors de nos tests. Klippel a déjà utilisé la densité de sonie de ses signaux pour définir des attributs objectifs caractérisant les dimensions perceptives révélées par ses tests d'écoute sur des enceintes acoustiques [10][48]. Staffeldt [26] a aussi relié ces dimensions "emphasized treble" et "emphasized bass" à la forme des réponses en fréquence de ses enceintes, après avoir traité les réponses fréquentielles à l'aide d'un modèle auditif. La prise en compte de l'étalement spectral

dû au fonctionnement du système auditif semble déterminante pour retrouver les informations perçues par les auditeurs.

La densité de sonie totale, qui a conduit à des espaces objectifs proches des espaces perceptifs dans de nombreux cas, a tout de même fourni quelques mauvais résultats lors de notre première série d'expériences. L'analyse globale sur l'ensemble du signal effectuée lors de la détermination de la densité de sonie totale des enregistrements ne serait donc pas suffisante. Klippel [10][48] et Staffeldt [26] ont considéré uniquement la densité de sonie totale de leurs signaux, tandis que Bramsløw [9] a utilisé la densité de sonie en fonction du temps. Notre première série d'expériences semble indiquer qu'il est important de prendre en compte le fait que les phénomènes d'étalement spectral liés au système auditif dépendent de l'évolution temporelle des signaux considérés. Des recherches complémentaires restent à effectuer pour répondre à cette question cruciale. Lorsque Staffeldt [26][43][63] applique un modèle auditif directement sur la réponse en fréquence de ses enceintes pour évaluer leur reproduction sonore, il ne tient pas compte du fait que ces enceintes reproduisent des signaux musicaux non-stationnaires, et que l'évolution temporelle de ces signaux peut avoir un effet sur la perception des auditeurs. La nécessité de prendre en compte cette évolution temporelle remettrait en cause la pertinence de l'observation directe des réponses en fréquence des enceintes acoustiques.

Les densités de sonie 1 et 2 donnent des résultats équivalents. La densité de sonie 1 de nos signaux est calculée toutes les 100 ms, tandis que leur densité de sonie 2 est déterminée toutes les 10 ms. La résolution temporelle des modèles auditifs utilisés ne semble donc pas cruciale vis-à-vis des deux dimensions révélées pour le moment par nos expériences. Par rapport à la densité de sonie 1, la densité de sonie 2 ajoute la prise en compte de l'étalement temporel du signal dû au fonctionnement du système auditif. Le fait d'obtenir des résultats équivalents avec les deux modèles semble indiquer que la prise en compte de ce phénomène n'apparaît pas essentielle dans notre évaluation, ce qui apporte du crédit à l'hypothèse de Bramsløw [9] qui avait négligé cet effet lors de l'élaboration de son modèle d'évaluation de la qualité de la reproduction sonore.

Les densités de sonie 1 et 2 moyennes fournissent d'aussi bons résultats que les densités de sonie 1 et 2 variant dans le temps. Les dimensions perceptives semblent également liées à une impression moyenne des auditeurs sur l'ensemble de l'extrait [11][43][63][64]. Bramsløw a basé son modèle sur la moyenne temporelle de la densité de sonie de ses signaux : "there is one subjective rating over the entire period — as implicitly 'time-averaged' by the listener, and the objective estimate has likewise been collapsed across time according to the model assumptions" [9].

IV.5 Troisième expérience : influence des stimuli utilisés

Afin de se donner une chance de révéler davantage de dimensions perceptives lors de nos tests psychoacoustiques, notre troisième expérience a impliqué un panel d'enceintes beaucoup plus large que nos expériences précédentes. Ces dernières, avec douze stimuli, ne pouvaient pas mettre en évidence plus de deux dimensions perceptives. La dépendance du résultat de la comparaison objectif/perceptif vis-à-vis du nombre et de la nature des enceintes impliquées a pu être évaluée grâce à l'analyse complète de notre troisième expérience. Celle-ci a fait intervenir trente-sept enceintes, et une tâche de classification libre a été utilisée pour réaliser le test psychoacoustique.

La Figure IV.9 présente le dendrogramme obtenu après l'analyse en arbre hiérarchique des dissemblances entre enregistrements révélées lors du test d'écoute. Seules les enceintes 8 et 13 forment une catégorie à part vis-à-vis des autres enceintes. Ces deux modèles possèdent une bande passante très limitée et ont été incorporés dans l'expérience en tant que représentants d'une reproduction sonore de mauvaise qualité, a priori très différents des autres modèles. Ils nuisent cependant à l'homogénéité de nos stimuli, et même s'ils ne représentent que deux stimuli parmi trente-sept, il convient d'être vigilant lors de l'interprétation des dimensions révélées par l'analyse multidimensionnelle. Aucune autre catégorie n'apparaît nettement parmi les enceintes restantes. Toute partition qui pourrait être effectuée parmi les enregistrements semble dépendre de façon continue du degré de similarité entre les enregistrements qui serait choisi pour effectuer cette partition.

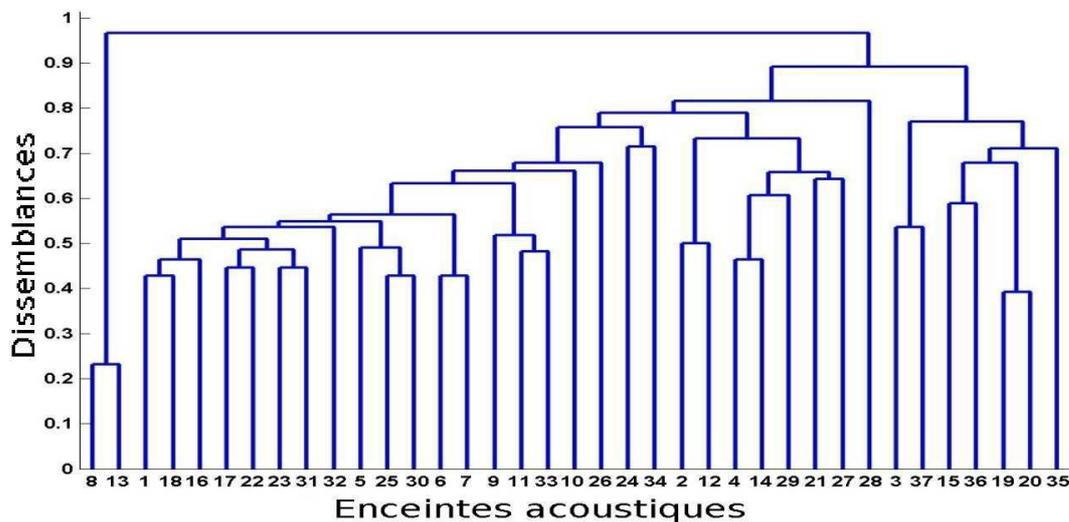


FIGURE IV.9 – Dendrogramme résultant de l'analyse en arbre hiérarchique des dissemblances entre enregistrements lors de notre troisième expérience. Les enceintes 19 et 20 sont du même modèle.

L'analyse multidimensionnelle des dissemblances perceptives a permis de révéler quatre di-

mensions (Figure IV.10), correspondant aux quatre attributs perceptifs principaux utilisés par les auditeurs pour différencier les enregistrements lors des tests psychoacoustiques. Sans avoir effectué d'analyse sémantique complète des entretiens avec les auditeurs, nous avons utilisé leurs commentaires pour essayer d'identifier ces dimensions, en nous aidant d'une écoute des enregistrements basée sur l'espace perceptif. L'orientation des axes de l'espace résultant de l'analyse multidimensionnelle n'étant pas fixée, nous avons conservé l'orientation obtenue par défaut avec notre programme. Celle-ci repose sur une analyse en composantes principales. Nous avons écouté les enregistrements en nous déplaçant le long des différentes dimensions pour tenter de les interpréter.

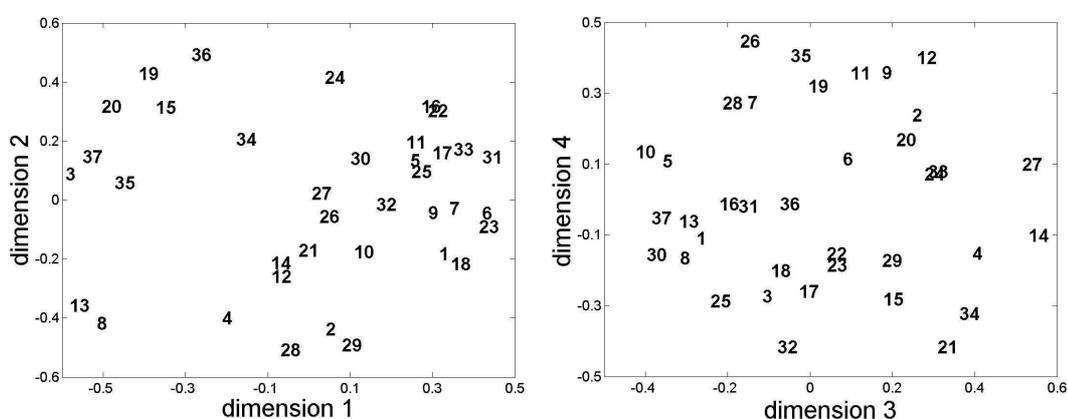


FIGURE IV.10 – Espace perceptif obtenu lors de notre troisième expérience. Chaque numéro correspond à une enceinte acoustique. Les enceintes 19 et 20 sont du même modèle.

La dimension 1 est prépondérante et paraît déterminée par l'équilibre grave-aigu ou la balance spectrale de nos signaux. Elle a donc la même interprétation que la première dimension perceptive de nos expériences précédentes, cependant, son comportement est différent. En effet, lorsque que nous nous déplaçons de gauche à droite sur cette dimension, nous trouvons les enregistrements manquant de basses, puis ceux ayant trop de basses et enfin les stimuli "équilibrés", alors que lors de nos expériences précédentes, nous avons les signaux manquant de basses, puis les équilibrés et enfin ceux ayant trop de basses. La dimension 3 correspond à la deuxième dimension perceptive de nos expériences précédentes, liée à la présence plus ou moins importante de médium, et pouvant être associée à une notion de clarté du son. Ces interprétations des dimensions perceptives sont étayées par des attributs objectifs pouvant décrire ces dimensions. Ces derniers sont présentés dans la suite de ce chapitre.

Deux nouvelles dimensions perceptives ont été révélées par rapport à nos expériences précédentes. La dimension 2 serait liée à la réverbération, à la sensation d'espace évoqué à l'écoute des enregistrements. Elle résulterait donc de l'interaction des enceintes avec la salle. Les tests

d'écoute de Gabrielsson et Sjögren [7], puis de Klippel [10], ont révélé une dimension du même type, nommée "feeling of space" par les auteurs. Bramsløw [9] parle lui de "spaciousness" pour interpréter ces résultats. Notre dimension 4 est principalement décelée sur les cymbales et peut caractériser la "douceur" ou "dureté" du son. Elle serait comparable à la "sharpness/hardness-softness" de Gabrielsson et Sjögren [7].

Il est à noter que notre dimension 2, à caractère plus spatial que timbral, semble avoir davantage de poids sur les jugements des auditeurs que la dimension de clarté du son reliée à l'émergence du médium. Elle constitue en effet le deuxième axe fixé par l'analyse en composantes principales qui a déterminé l'orientation de l'espace que nous avons considéré. Ceci nuancerait donc l'affirmation de Staffeldt selon laquelle "the room impression and width of the stereophonic sound image are of less importance to the subjective evaluation of the loudspeakers" [26]. Même avec des tests d'écoute réalisés au casque, à partir d'enregistrements de reproductions monophoniques, l'impression d'espace et l'effet de salle sont apparus comme l'un des critères les plus importants lors de l'évaluation des dissemblances par les auditeurs. Cette importance a été confirmée par les commentaires libres de nos auditeurs. Ce résultat diffère de ceux d'Olive [46] qui a montré que 94% des commentaires de ses auditeurs étaient reliés au timbre, tandis que seulement 3% semblaient caractéristiques de la composante spatiale de la reproduction sonore. Ceci pourrait s'expliquer par la nature des enceintes impliquées dans ses tests. En effet, la dimension "spatiale" révélée par notre dernière expérience n'était pas non plus apparue lors de nos tests précédents. Si les stimuli ne varient pas suffisamment selon une dimension donnée, celle-ci n'est pas révélée par les tests d'écoute, car ce n'est pas un des critères principaux utilisés par les auditeurs pour différencier les stimuli. Les dimensions perceptives pouvant être mises en évidence par un test d'écoute dépendent bien sûr des stimuli impliqués dans ce test.

Gabrielsson *et coll.* [11] rappellent à propos de l'évaluation de l'importance relative des différentes dimensions révélées par leurs tests [22] : "This is indicated by the correlations between the perceptual scales and fidelity, which, in the present case, were highest for clarity and spaciousness, followed by fullness, brightness, softness, and absence of extraneous sounds." Même si le classement exact du poids des différentes dimensions diffère légèrement, les quatre dimensions perceptives impliquées dans nos tests psychoacoustiques semblent en accord avec ces résultats.

IV.5.1 Influence de la tâche perceptive ?

Nous n'avons pas pu expliquer la différence observée au niveau du comportement de la première dimension perceptive. De nouvelles données expérimentales sont nécessaires pour cela. Nous ne savons pas si ce comportement n'a pas été révélé lors de nos précédentes expériences du fait d'un manque de stimuli, ou bien si ce comportement est dû à l'influence de la tâche de classification libre employée dans le test psychoacoustique de la troisième expérience.

Cette dimension perceptive pourrait être caractéristique du "degree of discolouration" considéré par Klippel [10][48], qui évalue l'écart de la reproduction par rapport à une reproduction idéale "équilibrée". Ce serait le cas si nous pouvions affirmer que les reproductions avec trop de basses sont perçues plus équilibrées que les reproductions manquant de basses. Mais, les auditeurs ont également pu utiliser des critères de qualité ou de préférence pour grouper les enregistrements, plutôt que de se baser sur leur similarité comme il leur était demandé dans la consigne du test. Les reproductions manquant de basses auraient été considérées comme de très mauvaise qualité, les reproductions équilibrées correspondraient à la meilleure qualité, tandis que celles ayant trop de basses seraient de qualité intermédiaire.

La présence des enceintes 8 et 13, conduisant à une reproduction sonore de particulièrement mauvaise qualité, peut avoir favorisé un tel biais lors du test psychoacoustique. Ce sont les enceintes qui ont été le plus souvent groupées ensemble (Figure IV.9), alors que les enregistrements correspondant ne semblent pourtant pas être les plus semblables à l'écoute, si ce n'est qu'ils sont d'aussi mauvaise qualité. En incorporant ces modèles d'enceinte, nous avons voulu suivre les exemples rencontrés dans la littérature qui ont utilisé des reproductions devant être symptomatiques d'une mauvaise qualité pour servir de référence [14][19][20][22]. Mais ces études étaient justement souvent basées sur des évaluations de qualité, de fidélité. Suivre leur exemple ne s'est finalement peut être pas avéré judicieux dans le cadre de notre expérience.

Toole [14] a montré que les évaluations de fidélité des enceintes sont influencées par la présence ou non de modèles de mauvaise qualité au sein du même test d'écoute. Il peut en être de même pour des évaluations de dissemblances. Est-ce que nos deux enceintes de mauvaise qualité n'ont pas entraîné une tâche d'identification qui a remplacé l'évaluation directe de la similarité? Guyot [49] indique que lors de ces expériences portant sur des sons de l'environnement: "l'identification de la source était un facteur prédominant dans la formation des catégories. Dans la plupart des situations, lorsque la source est reconnue, on ne cherche pas à qualifier le son. C'est pourquoi, les objets sonores ne sont pas uniquement catégorisés sur la base de similarités perceptives, mais aussi sur celle de similarités des sources". Le dendrogramme de la Figure IV.9 montre que les enceintes 8 et 13 nuisent à l'homogénéité de notre corpus de stimuli, et ceci peut avoir généré un biais lors de l'évaluation des dissemblances [24].

L'interprétation de la première dimension perceptive révélée par nos tests reste donc à explorer. Elle est liée à l'équilibre grave-aigu de nos enregistrements, mais est-elle caractéristique de l'équilibre ou de la qualité des reproductions testées? La tâche de classification libre a-t-elle eu une influence? Favorise-t-elle les jugements basés sur une notion de qualité plutôt que de similarité, ou bien aurions-nous observé la même tendance avec un test de comparaison par paires? Les enceintes 8 et 13 ont été les deux enceintes le plus souvent groupées ensemble, mais est-ce que c'est parce qu'elles se ressemblent ou parce qu'elles sont très différentes des autres? Une comparaison par paires confirmerait sûrement que ces deux enceintes sont très différentes des

autres, mais pas forcément qu'elles conduisent aux deux enregistrements les plus semblables. Un test basé sur la classification libre offre davantage de liberté aux auditeurs qu'un test de comparaison par paires. Les réponses peuvent être moins spontanées car l'auditeur peut écouter et ré-écouter les stimuli, modifier les groupements préalablement établis, revenir sur ses jugements. Il peut donc apparaître plus difficile de contrôler ce qui fait réellement l'objet de l'évaluation. Dans notre expérience, nous souhaitons que ce soit la similarité entre les enregistrements. De nouvelles données expérimentales devront être collectées pour répondre à ces questions.

IV.5.2 Limitations de l'approche objective pour décrire les quatre dimensions

Les dissemblances objectives entre les trente-sept enregistrements, calculées à partir des différentes techniques d'analyse testées, ont également été soumises à l'analyse multidimensionnelle. Aucune technique d'analyse n'a cette fois permis de trouver un espace objectif semblable à l'espace perceptif à quatre dimensions. La densité de sonie a conduit à des espaces objectifs bidimensionnels insuffisants pour décrire correctement ces quatre dimensions.

Afin de tenter d'améliorer notre analyse objective, en plus des densités de sonie considérées précédemment, nous avons calculé la valeur moyenne des densités de sonie obtenues sur chacune des deux voies de nos signaux. Nous souhaitons ainsi prendre en compte le fait que les auditeurs écoutent simultanément les deux voies de nos enregistrements, et effectuent un jugement global de dissemblance après cette écoute. La moyenne droite-gauche permet d'augmenter très légèrement la corrélation entre dissemblances objectives et perceptives — avec un coefficient de corrélation de 0,74 pour la densité de sonie 2 moyenne, contre 0,70 précédemment (Figure III.7, page 75) — mais les tendances observées restent les mêmes. L'espace objectif obtenu à partir de la moyenne droite-gauche de la densité de sonie 2 moyenne de nos enregistrements est présenté sur la Figure IV.11.

Les deux dimensions objectives mises en évidence par les densités de sonie sont les mêmes que lors de nos expériences précédentes. Elles sont liées à l'équilibre grave-aigu et à l'émergence du médium. La deuxième dimension liée au médium permet de décrire assez bien la troisième dimension perceptive. Par contre, du fait de la différence de comportement observée au niveau de la première dimension perceptive, la première dimension objective n'est plus appropriée pour la décrire. La première dimension objective a en effet le même comportement que les premières dimensions perceptives et objectives de nos expériences précédentes. Les deuxième et quatrième dimensions perceptives ne sont pas mises en évidence au sein des dissimilarités objectives telles que nous les avons calculées.

Cette troisième expérience nous a donc permis de mettre en évidence les limitations de notre analyse objective qui devra être affinée. La résolution temporelle des modèles auditifs utilisés pourrait être insuffisante pour décrire les dimensions perceptives 2 et 4. La dimension 2 serait liée à la réverbération contenue dans les enregistrements, tandis que la dimension 4 est

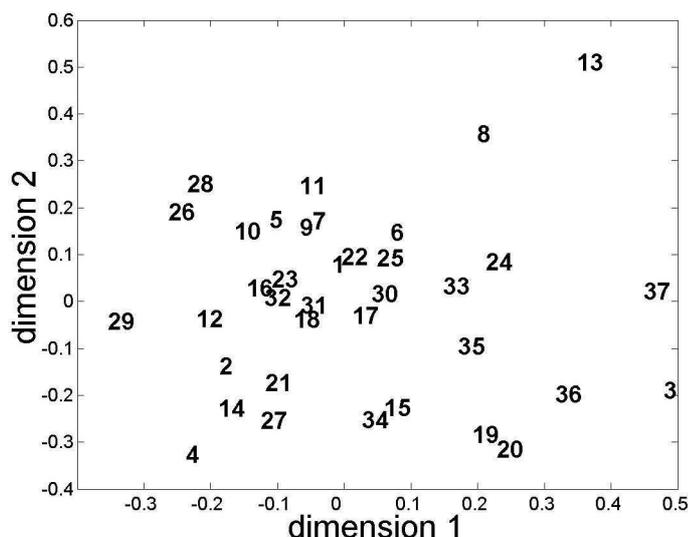


FIGURE IV.11 – *Espace objectif déterminé à partir de la moyenne droite-gauche de la densité de sonie 2 moyenne de nos signaux. Il est à comparer à l'espace perceptif de la Figure IV.10 correspondant à cette troisième expérience.*

détectée principalement sur les transitoires des cymbales. Ces interprétations semblent indiquer qu'il serait intéressant de pouvoir considérer la structure temporelle fine des signaux, afin de détecter l'information pertinente vis-à-vis de ces deux dimensions. La métrique que nous avons définie à partir des densités de sonie pour évaluer leur dissemblance mérite sans doute d'être améliorée. La moyenne temporelle telle que nous la calculons sur les dissemblances déterminées aux différentes instants élimine peut être beaucoup d'information qui pourrait s'avérer utile.

IV.6 Attributs objectifs préliminaires

Afin de mieux identifier les dimensions perceptives révélées par nos tests, nous avons cherché à les décrire à l'aide d'attributs objectifs préliminaires. Cette recherche n'a été que partielle et ne se veut en rien exhaustive. Elle avait simplement pour but de compléter notre écoute et l'analyse des commentaires de nos auditeurs.

L'orientation des axes de nos espaces multidimensionnels n'est pas fixée par notre programme d'analyse. Ceci nous a autorisé à tourner et dilater les différents espaces par rapport à une même référence, afin de faciliter leur comparaison. Pour les espaces qui ont servi de référence, nous avons conservé l'orientation des axes par défaut, qui repose sur une analyse en composantes principales. Lors de la recherche d'attributs objectifs décrivant les dimensions, l'orientation choisie a son importance. Une recherche plus approfondie d'attributs objectifs nécessiterait de considérer plusieurs orientations, afin de définir les meilleurs attributs possibles.

Les commentaires des auditeurs d'une part, et l'écoute des enregistrements réalisée en se déplaçant le long des différentes dimensions d'autre part, nous ont permis d'avoir une idée du type d'attributs que nous devons rechercher pour interpréter plus objectivement ces dimensions. La densité de sonie 2 moyenne s'étant avérée une des techniques d'analyse objective les plus pertinentes pour différencier nos enregistrements, c'est elle que nous avons utilisée pour définir ces attributs objectifs.

Pour notre première série d'expériences, nous avons utilisé directement les espaces objectifs pour rechercher des attributs objectifs décrivant nos dimensions. Les espaces objectifs et perceptifs étant très ressemblants, les attributs déterminés sont directement reliés aux dimensions perceptives révélées lors de nos tests psychoacoustiques. Notre deuxième série d'expériences a conduit à des dimensions perceptives et objectives similaires à la première série d'expériences (Figure IV.5, page 87). Aucune recherche supplémentaire d'attributs objectifs n'a été effectuée sur ces données. Pour notre troisième expérience, les espaces perceptif et objectif sont différents. Des attributs objectifs ont donc été recherchés pour chacun de ces espaces. Ceci nous a notamment permis d'étudier la différence de comportement des premières dimensions perceptives et objectives. Cette différence de comportement limitant la pertinence de notre analyse objective, il était important de mieux la comprendre. L'analyse objective n'ayant pas permis de retrouver les dimensions 2 et 4 de l'espace perceptif de notre troisième expérience, nous n'avons pour le moment pas cherché d'attributs décrivant ces dimensions. Cette recherche devra être effectuée, car interpréter plus objectivement ces dimensions nous permettra a priori de mieux déceler les limitations de notre analyse objective, et d'en proposer ainsi des améliorations.

Notre premier critère de choix lors de la recherche d'attributs objectifs a été un critère de monotonie. Un attribut est jugé adéquat s'il est lié par une relation monotone à la dimension qu'il est censé décrire. Le critère de corrélation entre l'attribut et la dimension a aussi été considéré, mais il est plus sévère puisqu'il impose à cette relation d'être linéaire. Concernant les données de notre première série d'expériences, l'interprétation préalable des dimensions étant la même pour les trois extraits musicaux et les trois prises de son, nous avons cherché à définir des attributs qui étaient appropriés pour l'ensemble des signaux testés. Seuls les attributs ayant conduit aux meilleurs résultats selon l'ensemble de ces critères sont décrits dans ce document.

IV.6.1 Balance grave-aigu

La Figure IV.12 présente les densités de sonie 2 moyenne de trois enregistrements de notre première série d'expériences, la voie gauche des enregistrements AB ORTF utilisant l'extrait musical Kan'nida ayant été choisie à titre d'exemple. Les trois signaux considérés sont alignés parallèlement à la première dimension de l'espace objectif correspondant, obtenu à partir de la densité de sonie 2 moyenne. Le premier attribut que nous avons défini à partir de cette densité

de sonie 2 moyenne est la "balance grave-aigu" de nos enregistrements. Cet attribut objectif est calculé en considérant le niveau dans les trois premières bandes de Bark, c'est-à-dire sur la gamme de fréquence 20-280 Hz, et celui dans les dix dernières bandes de Bark, c'est-à-dire sur la gamme de fréquence 1800-15500 Hz. Ces régions fréquentielles apparaissent en trait plein sur la Figure IV.12. Les niveaux sont obtenus en intégrant la densité de sonie 2 moyenne sur chacune de ces deux gammes de tonie. La balance grave-aigu est ensuite déterminée en prenant le rapport de ces deux niveaux. Les gammes de fréquence considérées ont été déterminées de manière à fournir les meilleurs résultats possibles, selon les critères que nous nous étions fixés, afin de décrire la première dimension de nos différentes expériences, pour l'ensemble des signaux testés.

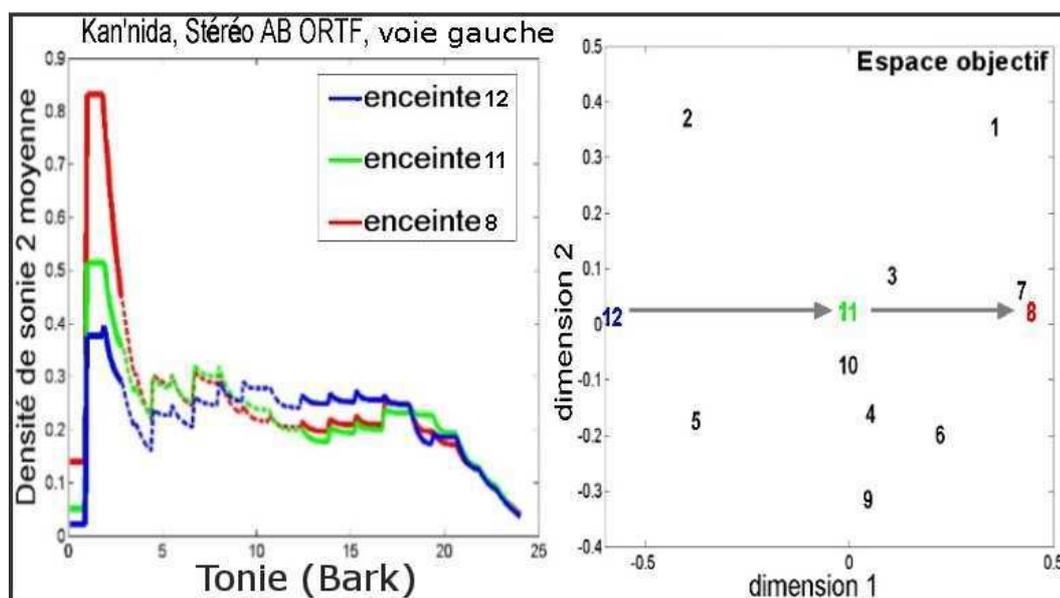


FIGURE IV.12 – Première série d'expériences : balance grave-aigu de nos enregistrements et première dimension de l'espace objectif correspondant (densité de sonie 2 moyenne, Kan'nida, AB ORTF, voie gauche). Chaque numéro correspond à une enceinte acoustique. La balance grave-aigu est définie comme le rapport des niveaux dans les bandes de Bark 1 à 3 (20-280 Hz) et 12 à 21 (1800-15500 Hz).

Considérant notre première série d'expériences, la relation entre la balance grave-aigu de nos enregistrements et la première dimension de l'espace objectif est monotone pour pratiquement tous nos types de signaux. Cette relation est même parfois linéaire. Le coefficient de corrélation entre l'attribut et la dimension varie entre 0,78 et 0,97 pour les différents types d'enregistrement (annexe C.6). Il vaut en moyenne 0,91 sur l'ensemble des signaux. La Figure IV.13 présente deux cas à titre d'exemple : le plus favorable présentant une relation linéaire (Kan'nida, prise omnidirectionnelle), et le plus défavorable pour lequel la monotonie n'est plus assurée (Vivaldi, AB ORTF, voie droite).

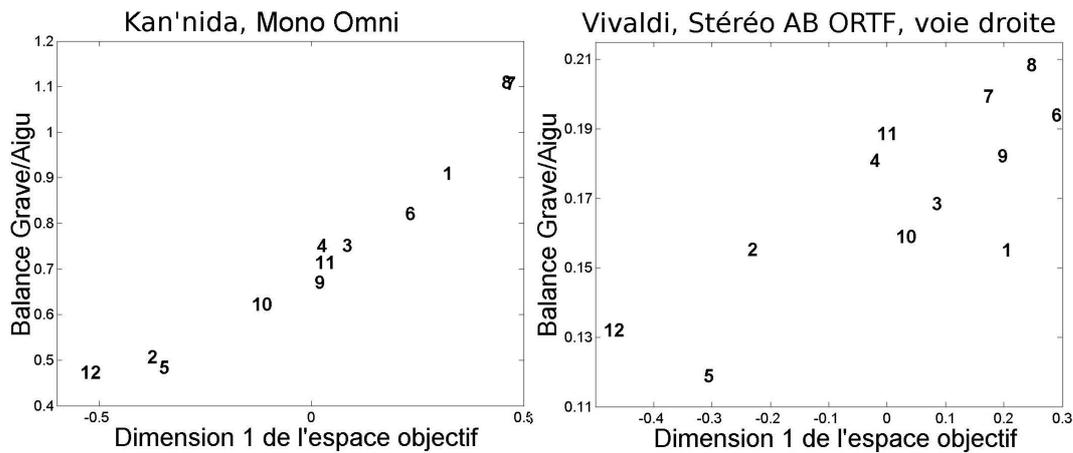


FIGURE IV.13 – Première série d'expériences : évaluation de la monotonie de la relation entre la balance grave-aigu de nos enregistrements et la première dimension de l'espace objectif correspondant (densité de sonie 2 moyenne). Chaque numéro correspond à une enceinte acoustique. Les deux exemples présentés sont ceux ayant conduit à la meilleure (Kan'nida, prise omnidirectionnelle) et la moins bonne (Vivaldi, AB ORTF, voie droite) corrélation entre l'attribut et la dimension.

La première dimension de l'espace perceptif à quatre dimensions obtenu lors de notre troisième expérience a été interprétée de la même façon que la première dimension des expériences précédentes, mais elle n'a pas le même comportement. En considérant la balance grave-aigu de nos enregistrements, cette différence de comportement est mise en évidence. La Figure IV.14 présente les variations de la balance grave-aigu des enregistrements lorsque nous nous déplaçons selon la première dimension de l'espace perceptif. Lors d'un déplacement de la gauche vers la droite sur cette dimension, les enregistrements manquent dans un premier temps de basses, avec par contre une amplification des aigus. Ensuite, au milieu de la dimension, se trouvent les enregistrements ayant trop de basses et comparativement pas assez d'aigus. Finalement, l'extrémité droite de la dimension contient les enregistrements "équilibrés". Ces observations confirment notre interprétation précédemment basée sur l'écoute des enregistrements. Le comportement de la première dimension de l'espace objectif basé sur la densité de sonie 2 moyenne est quant à lui similaire à celui des premières dimensions perceptives et objectives de nos expériences précédentes. Cette dimension de l'espace objectif dépend quasiment linéairement de la balance grave-aigu (Figure IV.14), avec un coefficient de corrélation de 0,90.

Gabrielsson *et coll.* [64] ont eux aussi relié leur dimension "brightness versus dullness" aux niveaux d'amplification des basses et des hautes fréquences, mais ces niveaux étaient évalués sur les réponses fréquentielles de leurs systèmes de reproduction. La dimension "brightness" augmente lorsque l'amplification de la réponse est déplacée vers les aigus, tandis qu'inversement, la

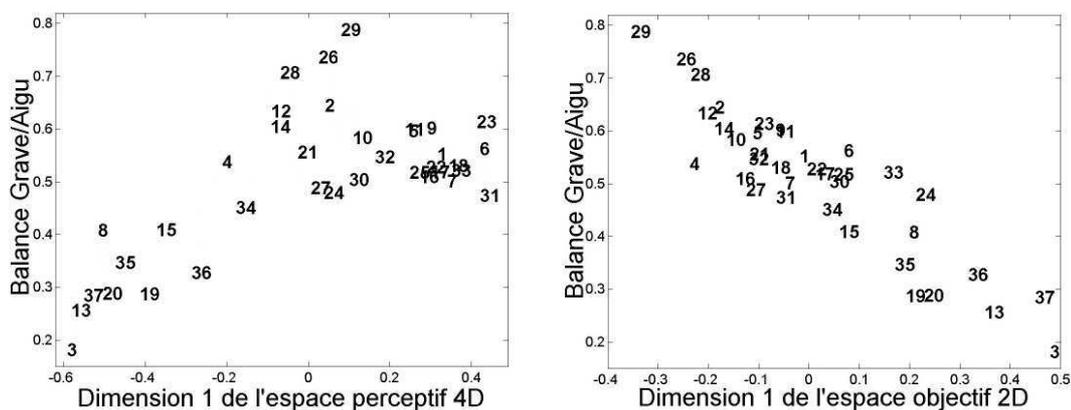


FIGURE IV.14 – *Troisième expérience : évaluation de la monotonie des relations entre la balance grave-aigu de nos enregistrements et les premières dimensions des espaces perceptif et objectif (moyenne droite-gauche de la densité de sonie 2 moyenne). Chaque numéro correspond à une enceinte acoustique.*

dimension "dullness" est favorisée lorsque ce déplacement s'effectue vers les basses fréquences. La dimension "fullness" a aussi été associée à l'amplification de ces basses fréquences. Les liens entre attributs et dimensions perceptives reposent ici uniquement sur des corrélations visuelles, c'est-à-dire des tendances observées en comparant les formes de réponse fréquentielle et les évaluations selon les dimensions perceptives.

Klippel [10][48] a utilisé la densité de sonie totale de ses signaux pour définir des attributs objectifs décrivant les résultats de ses tests d'écoute. Ces attributs ont été calculés à partir du "discolouration spectrum" des signaux, qui correspond à la différence entre les densités de sonie des signaux et celle d'une reproduction idéale "équilibrée". Les attributs reposent sur différentes pondérations appliquées sur ce "discolouration spectrum", favorisant soit les basses fréquences, soit les aigus. Il a ainsi obtenu des attributs bien corrélés avec les dimensions "treble stressing", "general bass emphasis", "low bass emphasis", avec des coefficients de corrélation entre 0,87 et 0,90. Il a ensuite décrit la dimension "brightness", présentée comme une évaluation globale incluant "general bass emphasis" et "treble stressing", par une combinaison linéaire des deux attributs correspondants. Cette dernière dimension semble avoir le même type d'interprétation que notre balance grave-aigu.

IV.6.2 Emergence du bas-médium

Le deuxième attribut objectif que nous avons considéré évalue "l'émergence du bas-médium" de nos enregistrements. A la différence de la balance grave-aigu présentée au paragraphe précédent, nous n'avons pas pu déterminer une définition unique de l'attribut qui soit valable pour

toutes nos expériences. Une définition a été utilisée pour décrire la deuxième dimension impliquée dans notre première série d'expériences. Nous avons alors recherché une définition qui soit appropriée pour l'ensemble des signaux testés. Cette définition a dû être ensuite modifiée pour la description de la troisième dimension perceptive de notre troisième expérience. Même si la définition précise de l'attribut n'est pas la même pour ces deux séries d'expériences, l'évaluation porte dans les deux cas sur l'émergence du bas-médium.

Pour notre première série d'expériences, l'émergence du bas-médium 1 a été définie sur la densité de sonie 2 moyenne de nos enregistrements comme le niveau dans les bandes de Bark 5 à 9, correspondant à la gamme de fréquence 355-1120 Hz. Ce niveau est calculé en intégrant la densité de sonie 2 moyenne sur cette gamme de tonie. Comme les enregistrements sont égalisés en sonie, ce niveau du bas-médium correspond à l'émergence de cette gamme de fréquence par rapport au reste du spectre perçu. La Figure IV.15 illustre cette définition à partir de l'exemple des densités de sonie 2 moyenne de la voie gauche des enregistrements AB ORTF utilisant l'extrait musical Vivaldi. Les trois signaux considérés sont alignés parallèlement à la deuxième dimension de l'espace objectif correspondant, obtenu à partir de la densité de sonie 2 moyenne.

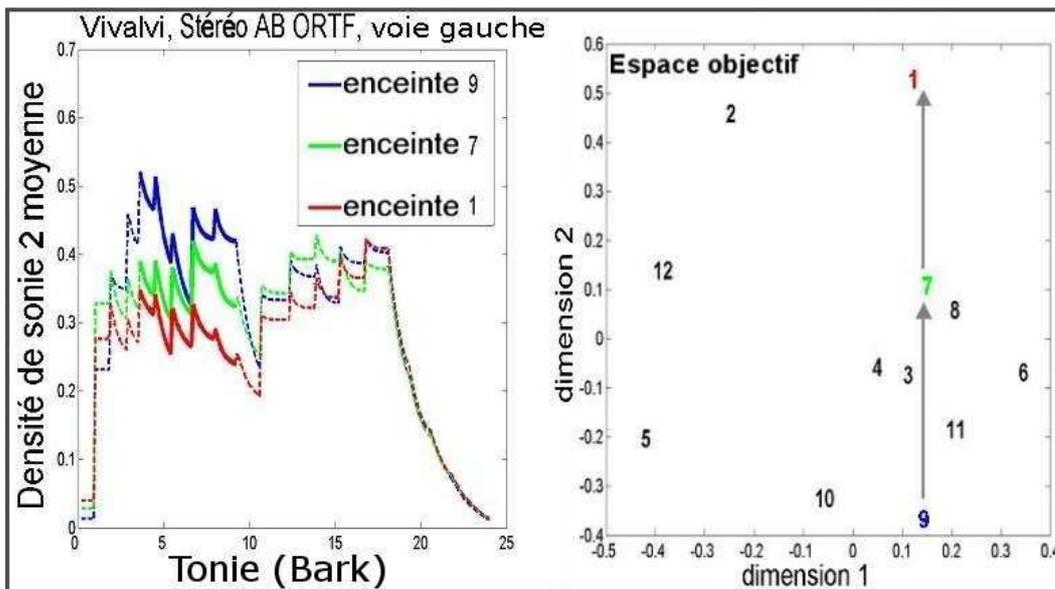


FIGURE IV.15 – Première série d'expériences : émergence du bas-médium 1 de nos enregistrements et deuxième dimension de l'espace objectif correspondant (densité de sonie 2 moyenne, Vivaldi, AB ORTF, voie gauche). Chaque numéro correspond à une enceinte acoustique. L'émergence du bas-médium est définie comme le niveau dans les bandes de Bark 5 à 9 (355-1120 Hz, 3,5-9,2 Barks).

La monotonie de la relation entre l'émergence du bas-médium 1 et la deuxième dimension de l'espace objectif est presque toujours respectée, cette relation pouvant même être linéaire. Le

coefficient de corrélation entre l'attribut et la dimension varie entre 0,65 et 0,96 pour les différents types d'enregistrement. Il vaut en moyenne 0,88 sur l'ensemble des signaux. La Figure IV.16 présente deux cas à titre d'exemple : le plus favorable présentant une relation linéaire (Mc Coy Tyner, voie droite de la prise AB ORTF), et le plus défavorable qui constitue une exception pour laquelle l'attribut ne semble pas assez discriminant (Kan'nida, voie gauche de la prise AB ORTF).

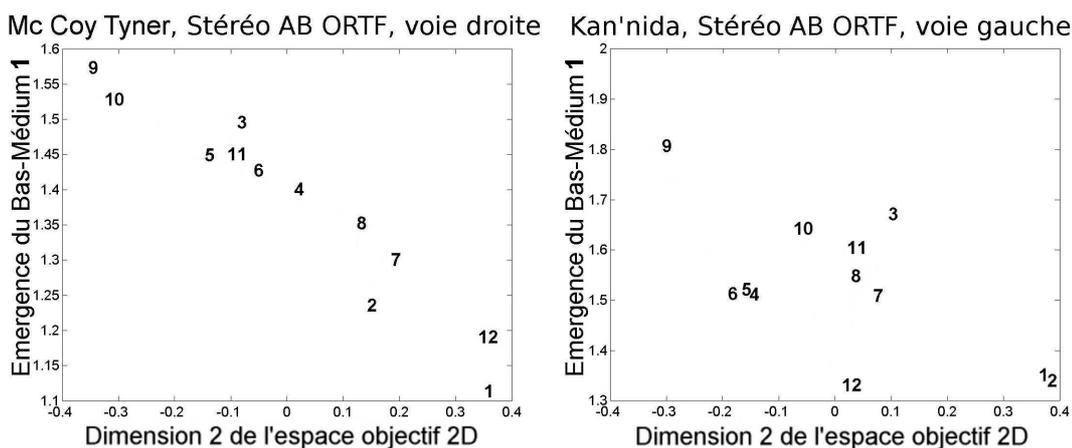


FIGURE IV.16 – *Première série d'expériences : évaluation de la monotonie de la relation entre l'émergence du bas-médium 1 de nos enregistrements et la deuxième dimension de l'espace objectif correspondant (densité de sonie 2 moyenne). Chaque numéro correspond à une enceinte acoustique. Les deux exemples présentés sont ceux ayant conduit à la meilleure (Mc Coy Tyner, voie droite de la prise AB ORTF) et la moins bonne (Kan'nida, voie gauche de la prise AB ORTF) corrélation entre l'attribut et la dimension.*

Pour notre troisième expérience, l'émergence du bas-médium 2 a cette fois été déterminée à partir du niveau dans les bandes de Bark 4 à 11, correspondant à la gamme de fréquence 280-1800 Hz. Le niveau est calculé en intégrant la densité de sonie 2 moyenne sur cette gamme de tonie. L'attribut objectif est obtenu en faisant le rapport de ce niveau et du niveau calculé sur le reste de la densité de sonie 2 moyenne. La relation entre l'émergence du bas-médium 2 et la troisième dimension perceptive est assez monotone (Figure IV.17). Le coefficient de corrélation entre l'attribut et la dimension perceptive vaut 0,76.

La Figure IV.17 présente également les variations de l'émergence du bas-médium 2 pour un déplacement le long de la deuxième dimension de l'espace objectif obtenu à partir de la densité de sonie 2 moyenne. La relation entre l'attribut et la dimension objective est monotone, et leur coefficient de corrélation s'élève à 0,87. D'autres attributs objectifs peuvent être utilisés pour décrire plus précisément la deuxième dimension objective (annexe C.6), mais nous recherchons le lien entre la deuxième dimension objective et la troisième dimension perceptive. L'émergence

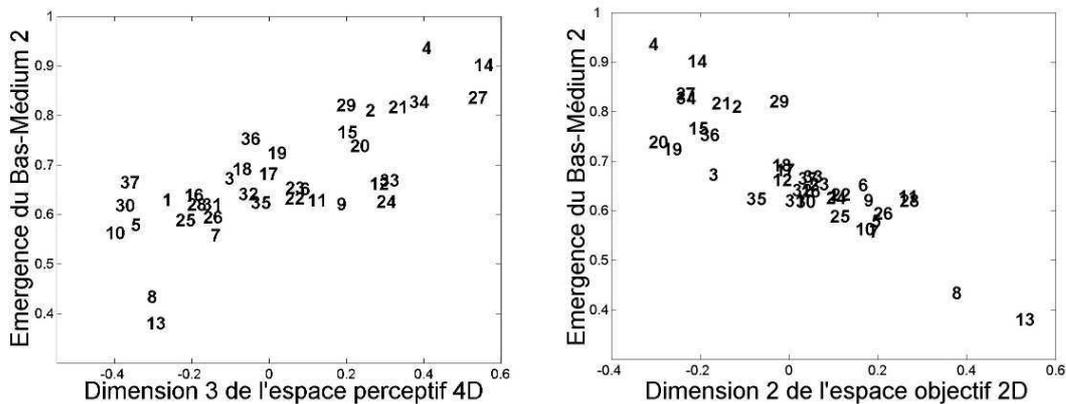


FIGURE IV.17 – Troisième expérience : évaluation de la monotonie des relations entre l'émergence du bas-médium 2 de nos enregistrements et, d'une part, la troisième dimension de l'espace perceptif, d'autre part, la deuxième dimension de l'espace objectif (moyenne droite-gauche de la densité de sonie 2 moyenne). Chaque numéro correspond à une enceinte acoustique.

du bas-médium 2 semble relier ces deux dimensions.

Même si elles n'ont pas des définitions identiques, les émergences du bas-médium 1 et 2 font intervenir la même gamme de fréquence. Bech [73] indique que la région fréquentielle comprise entre 500 Hz et 2 kHz est déterminante vis-à-vis des changements de timbre de la reproduction sonore. Gabrielsson *et coll.* [64] relie leur dimension "clarity" à une certaine amplification du médium, entre 500 et 4000 Hz, sur la réponse en fréquence de leur systèmes de reproduction.

IV.6.3 Bilan sur les attributs objectifs préliminaires

La recherche d'attributs objectifs préliminaires que nous avons effectuée avait pour but d'identifier de manière objective les dimensions perceptives révélées lors de nos tests, de mieux comprendre le comportement de la première dimension perceptive, et finalement de montrer qu'il est moins difficile de trouver des attributs objectifs décrivant les dimensions perceptives une fois qu'une technique d'analyse pertinente a été identifiée, et que nous savons comment observer les signaux.

La recherche approfondie d'attributs objectifs décrivant les critères perceptifs utilisés par les auditeurs lors de nos tests d'écoute reste à effectuer. Le fait que la définition précise de nos attributs préliminaires change au cours de nos différentes expériences montre qu'ils ne sont pas assez robustes. Les attributs proposés lors de notre troisième expérience apparaissent plus pertinents car ils sont définis sur beaucoup de stimuli. Par contre, ils pourraient être moins caractéristiques des enceintes que ceux de notre première série d'expériences, la détermination de ces derniers ayant fait intervenir des extraits musicaux différents.

Pour que ces attributs aient un sens et une utilité, leur définition doit être établie sur un corpus important de stimuli, afin qu'elle soit le plus indépendante possible de ces stimuli et le plus caractéristique possible de la perception de la reproduction sonore en général. Les recherches dans le domaine de la qualité sonore visent à caractériser de manière générale notre perception des événements sonores, mais ne disposent pour cela que d'un corpus de stimuli forcément limité. Les attributs objectifs définis, comme les dimensions perceptives révélées lors de tests psychoacoustiques, semblent souvent dépendre des stimuli impliqués [7][28][46][74][76]. "Il est fort probable que les dimensions sur lesquelles les auditeurs se concentrent sont fonction de l'ensemble de sons utilisé dans l'expérience, autrement dit que leurs représentations sont peut être codées par rapport au contexte des stimuli fournis au cours d'une séance de test" McAdams [74]. Pour réduire cet effet de contexte, Olive [46] indique qu'une solution consiste à augmenter la taille du panel d'enceintes considéré.

Afin de pouvoir généraliser nos résultats concernant la perception de la reproduction sonore, nous avons ainsi tenté d'utiliser un panel de stimuli le plus large possible lors de notre troisième expérience. Cette expérience a soulevé des questions concernant la nature de la première dimension perceptive, prépondérante, et l'influence de la tâche utilisée lors du test psychoacoustique. Il convient de répondre à ces questions avant de chercher à définir plus précisément des attributs objectifs caractérisant les dimensions mises en évidence.

Les analyses objectives que nous avons menées sont basées sur l'analyse de signaux musicaux reproduits par les enceintes acoustiques. Les attributs objectifs préliminaires auraient pu être optimisés pour chaque type de signal que nous avons utilisé. Pourtant, nous avons vu que notre interprétation des dimensions perceptives était la même pour trois extraits musicaux différents (Figure IV.2, page 82). Nous avons donc cherché des attributs qui étaient appropriés pour tous les signaux à la fois. Il paraît pertinent de rechercher une mesure qui soit plus caractéristique des enceintes et moins dépendante de l'extrait musical utilisé. C'est cette mesure qui fait l'objet de notre recherche. Nous avons montré qu'il était important de tenir compte du fonctionnement complexe du système auditif pour différencier les signaux issus d'enceintes différentes de manière pertinente, il s'agit maintenant de déterminer une méthode de mesure plus caractéristique du système mesuré et moins dépendante du signal de mesure lui-même. Des attributs objectifs caractéristiques des enceintes et directement reliés à la perception du champ sonore qu'elles produisent pourront alors être obtenus.

IV.7 Bilan de la comparaison multidimensionnelle des dissemblances

L'analyse multidimensionnelle nous a permis de comparer plus en détail nos dissemblances perceptives et objectives. L'ensemble de nos expériences impliquant une douzaine d'enceintes a

révélé deux dimensions perceptives liées à la balance grave-aigu et à l'émergence du médium de nos enregistrements. Les espaces objectifs obtenus à partir des densités de sonie de nos enregistrements décrivent assez bien ces deux dimensions perceptives, pour toutes les conditions expérimentales testées. Notre première série d'expériences semble indiquer qu'il est nécessaire de modéliser la dépendance des phénomènes auditifs vis-à-vis de l'évolution temporelle des signaux analysés. Ce résultat reste à confirmer, car il remettrait en cause la pertinence de l'observation directe des réponses en fréquence des enceintes acoustiques.

Bien sûr les évaluations perceptives et objectives sont influencées par les différents paramètres expérimentaux — l'extrait musical, la prise de son, le mode de reproduction des enceintes et la salle d'écoute — mais cette influence s'est avérée limitée. Les dimensions principales utilisées par les auditeurs lors de nos tests psychoacoustiques semblent bien caractéristiques des enceintes acoustiques. Ces dimensions ont notamment été les mêmes pour trois extraits musicaux différents. Notre but est donc maintenant de déterminer une évaluation objective des enceintes qui reste pertinente, mais moins dépendante des signaux de mesure utilisés et plus caractéristique des enceintes.

Considérer un panel d'enceintes beaucoup plus large nous a permis de mettre en évidence deux nouvelles dimensions perceptives, liées à la sensation d'espace et au caractère plus ou moins "doux" de nos enregistrements. Les limites de notre analyse objective ont ainsi été révélées. Cette analyse reste en effet pour le moment insuffisante pour prendre en compte ces nouveaux critères de différence. L'analyse objective devra donc être affinée afin de décrire complètement la perception de nos auditeurs.

Notre dernière expérience a également soulevé des questions concernant l'influence de la tâche perceptive impliquée dans nos tests psychoacoustiques. Nous devons répondre à ces questions pour pouvoir exploiter les résultats de tests perceptifs faisant intervenir un nombre important de stimuli. Ces tests sont nécessaires pour définir des attributs objectifs pertinents. La définition précise d'un attribut objectif est délicate, et souvent dépendante de la nature des stimuli considérés. De nombreuses données expérimentales sont nécessaires avant d'arriver à une telle définition. Notre approche pour déterminer de tels attributs consiste d'abord à identifier la meilleure technique d'analyse permettant d'observer nos signaux de manière pertinente vis-à-vis de la perception. Ensuite seulement nous utilisons cette analyse pour rechercher des attributs objectifs décrivant nos dimensions perceptives.

Ces perspectives sont décrites plus en détails dans le chapitre suivant qui présente les différentes pistes envisagées pour la suite de ce projet de recherche, dont le travail présenté dans ce document constitue une première étape.

CHAPITRE V

Perspectives

————— oOo —————

Notre étude a soulevé un certain nombre de questions. Ce chapitre contient des propositions d'expériences visant à y répondre. Il évoque également le potentiel de notre protocole expérimental, et certaines études que nous envisageons de réaliser en l'utilisant.

————— oOo —————

SOMMAIRE

V.1	Analyse du signal audio	111
V.1.1	Autres modèles et autres métriques	111
V.1.2	Méthodes de comparaison objectif/perceptif	112
V.1.3	Intérêt de la recherche d'attributs objectifs	113
V.2	Psychoacoustique	113
V.2.1	Influence de la tâche perceptive	114
V.2.2	Perception de la reproduction sonore	115
V.3	Mesures de réponses impulsionnelles	116
V.4	Enceintes acoustiques	117
V.4.1	Dynamique	117
V.4.2	Interaction avec la salle	118
V.4.3	Lien avec les paramètres physiques	121
V.5	Bilan	121

V.1 Analyse du signal audio

Beaucoup reste à faire concernant la recherche de l'information pertinente contenue dans nos signaux, vis-à-vis de la perception de leur dissemblances par les auditeurs lors des tests psychoacoustiques. Les densités de sonie permettent de retrouver deux des dimensions perceptives principales révélées par ces tests, mais au moins deux dimensions restent encore à mettre en évidence au niveau de l'analyse objective.

V.1.1 Autres modèles et autres métriques

Nous avons testé un certain nombre de méthodes d'analyse, mais d'autres sont à considérer. Par exemple, de nouveaux modèles auditifs [86] pourront être employés. Etant donné le caractère temporel de l'interprétation que nous avons donnée aux dimensions perceptives qu'il nous reste à mettre en évidence par l'analyse objective, la résolution temporelle de ces modèles nous a semblé être un paramètre important. Nous avons calculé la densité de sonie 2 de nos signaux avec une nouvelle résolution temporelle de 2 ms. Les corrélations entre les dissemblances objectives ainsi obtenues et les dissemblances perceptives n'ont pas été meilleures que celles obtenues à partir de la densité de sonie 2 calculée toutes les 10 ms, ou de la densité de sonie 1 calculée toutes les 100 ms. La métrique que nous avons utilisée sur les densités de sonie variant dans le temps, impliquant notamment une moyenne arithmétique des dissemblances instantanées, élimine peut être les différences potentielles d'information fournies par ces trois résolutions temporelles. Cette intégration temporelle pourrait aussi masquer des différences potentielles entre une densité de sonie variant dans le temps et sa moyenne temporelle. Nous avons choisi de réaliser cette intégration en utilisant une moyenne arithmétique, car nous ne voulions pas donner plus d'importance à un instant qu'à un autre. Mais rien ne prouve que les auditeurs font de même lorsqu'ils évaluent les dissemblances entre les enregistrements. Si le caractère plus ou moins "doux" du son est plus perceptible sur les coup de cymbales, alors tous les instants ne sont pas équivalents. De même, certaines régions fréquentielles peuvent avoir plus d'importance que d'autres pour les auditeurs. Il serait donc intéressant de considérer différentes pondérations de l'information contenue dans le plan "temps-fréquence perçu".

D'autres métriques doivent bien sûr être testées. Nous avons défini une seule métrique par technique d'analyse, or, si une analyse objective n'est pas pertinente, la métrique utilisée peut être autant responsable que la technique d'analyse. C'est vrai pour les densités de sonie, mais aussi pour les autres analyses. Les dissemblances associées aux techniques d'analyse classiques qui ont été considérées dans cette étude sont toutes quadratiques vis-à-vis des grandeurs comparées (paragraphe II.5.2, page II.5.2). Cette manière d'évaluer les dissemblances s'est avérée moins pertinente que celle définie sur les densités de sonie, mais cela ne veut pas dire que le spectre ou la transformée temps-fréquence des signaux ne sont pas adaptés pour retrouver les dissemblances

perceptives. Des essais ont été réalisés en calculant les dissemblances à partir de la différence des modules du spectre plutôt que de considérer le carré de cette différence. Les corrélations obtenues avec les dissemblances perceptives sont plus élevées, comparables aux résultats fournis par les densités de sonie. La comparaison multidimensionnelle reste à effectuer. Cette comparaison est plus critique. En effet, elle a mis en évidence des différences entre les résultats obtenus à partir de la densité de sonie totale et ceux obtenus à partir des autres densités de sonie, différences qui n'apparaissaient pas lors de l'évaluation des corrélations. Il est donc important de vérifier lors de cette comparaison multidimensionnelle si la nouvelle dissemblance définie sur le module du spectre donne toujours des résultats équivalents à celle définie sur les densités de sonie. Il semble également intéressant d'essayer de calculer des dissemblances à partir du logarithme du spectre. Les dissemblances devant alors être évaluées en prenant les rapports de logarithme, il sera important de définir un critère pour seuiller le logarithme du spectre, de manière à ne pas évaluer les dissemblances là où il n'y a pas de signal utile. Les métriques impliquées dans cette étude ont toutes été définies de façon parallèle sur les techniques d'analyse classiques, de manière à pouvoir comparer ces techniques. Il est maintenant important de pouvoir comparer différents types de métrique.

Les densités de sonie mettent en évidence au moins une partie de l'information que nous recherchons dans les signaux, mais il est peut être possible de retrouver cette information par des analyses moins complexes. En testant la densité de sonie totale et les densités de sonie 1 et 2, nous avons commencé à étudier la pertinence, vis-à-vis de notre étude, des différentes étapes de traitement impliquées dans les modèles auditifs. La modélisation des phénomènes d'étalement temporel au sein du système auditif ne semble pas primordiale, tandis que la prise en compte du caractère non-stationnaire des signaux analysés apparaît nécessaire. Il serait maintenant intéressant de définir des métriques sur des grandeurs intermédiaires entre le spectre et la densité de sonie, telles que le diagramme d'excitation de la membrane basilaire [87] (annexe E.3). Nous recherchons l'analyse la plus simple qui permette de retrouver les différences perçues par les auditeurs, il s'agit donc de conserver uniquement les étapes de l'analyse qui sont nécessaires.

V.1.2 Méthodes de comparaison objectif/perceptif

Le critère de corrélation utilisé lors de la comparaison directe des dissemblances objectives et perceptives permet de quantifier le lien entre ces dissemblances, mais il impose également une contrainte de linéarité. Rien ne nous permet pourtant de supposer que la relation entre ces dissemblances doit être linéaire.

La comparaison multidimensionnelle des dissemblances s'affranchit de cette contrainte de linéarité. La comparaison visuelle des espaces objectifs et perceptifs est une première étape. Il serait maintenant intéressant de définir un critère qui permette de quantifier la ressemblance

entre ces espaces.

La recherche de relations non-linéaires entre les dissemblances objectives et perceptives compléterait également cette étude. Les valeurs de corrélation obtenues lors de nos expériences indiquent que nous avons effectué un rapprochement entre dissemblances objectives et perceptives, mais d'autres types de rapprochement sont peut-être possibles.

V.1.3 Intérêt de la recherche d'attributs objectifs

Alors que la détermination d'une méthode d'analyse pertinente des signaux est utile pour définir des attributs objectifs décrivant les dimensions perceptives, inversement, la recherche de tels attributs pour les dimensions non expliquées par l'analyse objective peut permettre de comprendre les limitations de cette analyse, dans le but de l'améliorer. Il serait donc utile de chercher à décrire les dimensions perceptives 2 et 4 de notre troisième expérience (Figure IV.10, page IV.10). La dimension 2 semble caractéristique de l'interaction des enceintes avec la salle, et serait liée à la réverbération, à la sensation d'espace évoqué à l'écoute des enregistrements. Pour Klippel [10], "this sensation cannot be explained by the stationary measured sound pressure response, but it is caused by energy parts of the direct and diffuse field arriving one after another". Il décrit ainsi sa dimension "feeling of space" à partir de la différence de niveaux des champs direct et réverbéré simulés dans ses conditions d'écoute. Un attribut du type de ceux utilisés en acoustique des salles pour évaluer l'impression spatiale, basés sur la corrélation interaurale des signaux [88], pourrait aussi être pertinent. Notre dimension 4 peut caractériser la "douceur" ou "dureté" du son. "Softness is favored by a certain emphasis on lower frequencies, while sharpness is associated with more or less steeply rising responses toward higher frequencies or marked resonance peaks at higher frequencies" Gabrielsson *et coll.* [64]. Nous n'avons pas réussi à définir d'attribut de ce type qui décrive correctement notre dimension 4, à partir de la densité de sonie 2 moyenne de nos signaux. Cette dimension perceptive étant subtile et principalement décelée sur les cymbales, l'analyse objective devrait peut-être se limiter à certaines parties ciblées du signal pour se donner une chance de la décrire correctement.

V.2 Psychoacoustique

La comparaison des résultats de nos expériences a soulevé la question de l'influence de la tâche employée lors des tests psychoacoustiques. Nous devons répondre à cette question concernant notre outil d'évaluation, avant de pouvoir étudier davantage la perception de la reproduction sonore.

V.2.1 Influence de la tâche perceptive

Le comportement de la première dimension perceptive révélée par nos tests psychoacoustiques n'est pas la même suivant que la tâche utilisée repose sur des comparaisons par paires impliquant une douzaine de stimuli, ou sur une tâche de classification libre portant sur trente-sept stimuli. Nous ne savons pas si cette différence provient de la tâche perceptive ou des stimuli. Lors d'une étude portant sur l'émotion musicale, Bigand *et coll.* [81] ont obtenu des résultats comparables avec les deux types de test impliquant une même sélection de stimuli. Ils indiquent que la corrélation entre les matrices de dissemblances obtenues pour chaque test est significative, avec un coefficient de corrélation de 0,65, et que les espaces résultant de l'analyse multidimensionnelle de ces dissemblances sont dans l'ensemble assez similaires. De la même façon, nous souhaitons pouvoir comparer notre expérience basée sur la classification libre avec un test utilisant des comparaisons par paires sur nos trente-sept enregistrements.

La comparaison de toutes les paires possibles de stimuli serait trop longue pour que la tâche puisse être réalisée par les auditeurs. Certains stimuli "redondants" pourraient être éliminés, mais pour que le test devienne faisable il faudrait diviser par deux le nombre de stimuli. Les dimensions perceptives pouvant être mises en évidence par un test d'écoute dépendent bien sûr des stimuli impliqués dans ce test. Si nous éliminons trop de stimuli, les dimensions obtenues précédemment peuvent être modifiées. La comparaison des résultats des deux tests ne nous permettra alors pas de dire si cette modification est due au changement de tâche perceptive.

Bigand *et coll.* [81] ont pu réaliser les deux tests avec vingt-sept stimuli, après avoir réduit la durée de leurs extraits musicaux à une seconde. Il leur a d'abord fallu montrer que cette réduction ne modifiait pas leurs résultats. Le test de comparaison par paires durait ainsi une heure, ce qui est encore considérable. Fastl et Ellermeier [89] évoquent également la possibilité de ne pas faire juger toutes les paires de stimuli aux auditeurs. Certaines paires ne sont pas présentées, en fonction des réponses antérieures de l'auditeur : si la paire A-B a été jugée très semblable et que la paire A-C a été jugée très dissemblable, alors la paire B-C n'est pas présentée et est aussi considérée comme très dissemblable. Les espaces obtenus par analyse multidimensionnelle de ces dissemblances seraient par contre moins stables que lorsque toutes les paires sont évaluées [89]. Faure [90] cite également une méthode proposée par Tsogo [91][92] qui consiste à diviser les auditeurs en plusieurs groupes, chaque groupe écoutant tous les stimuli, mais évaluant seulement une partie des paires. Tsogo a montré que présenter deux tiers des paires à chaque groupe d'auditeurs était suffisant pour obtenir une bonne représentation multidimensionnelle, les évaluations moyennes de chaque groupe étant rassemblées dans une matrice de dissemblances globale.

Ces différentes pistes devront être explorées, afin de mieux définir l'influence de notre outil d'évaluation des dissemblances perceptives.

V.2.2 Perception de la reproduction sonore

Déterminer l'influence de la tâche utilisée dans nos tests d'écoute devrait nous permettre de mieux interpréter la nature de notre première dimension perceptive, de savoir si elle est plutôt liée à une notion de qualité, de déséquilibre spectral, ou de balance spectrale. Il faudra également identifier la "spécificité" de l'enceinte qui se déplace parfois beaucoup au sein des espaces perceptifs (enceinte 5 sur la Figure IV.2, page IV.2, et 7 sur la Figure IV.5, page IV.5). Ce déplacement a en effet été interprété comme une particularité perceptive de l'enregistrement correspondant, qui n'est pas partagée par les autres stimuli, et que le programme d'analyse multidimensionnelle utilisé ne sait pas prendre en compte.

Notre protocole expérimental peut aussi être utilisé pour étudier d'autres dimensions perceptives de la reproduction sonore. Nos tests n'ont certainement pas révélé toutes les dimensions potentielles. D'autres extraits musicaux pourraient notamment faire émerger de nouvelles dimensions. Le choix des "bons" extraits est nécessaire pour réaliser une recherche exhaustive de ces dimensions.

"Lorsqu'ils comparent des sons présentés les uns après les autres, les auditeurs ne sont capables d'utiliser ou de prêter attention simultanément qu'à un nombre limité de dimensions (...) Certaines dimensions ou propriétés peuvent acquérir une plus grande importance psychologique que d'autres, et de ce fait, influencer les évaluations de similarités" McAdams [74]. Après avoir décrit objectivement les dimensions perceptives, il est envisageable d'essayer de les égaliser. Cette égalisation pourrait autoriser la mise en évidence de dimensions plus subtiles, qui auraient été masquées par les dimensions prépondérantes avant leur égalisation. Au passage, nous pourrions également vérifier la pertinence des attributs objectifs utilisés : deux modèles d'enceintes égalisés devront se rapprocher au sein de l'espace perceptif, selon la dimension correspondant à l'attribut égalisé [43][52].

Il peut être utile de considérer des stimuli de synthèse pour explorer les dimensions perceptives, comme dans les études évoquées au paragraphe I.6. Il sera alors particulièrement important de veiller à la pertinence des dimensions étudiées, qui peuvent être créées par l'expérimentateur, et ne seraient peut être pas apparues en comparant des systèmes réels.

Les espaces objectifs, s'ils ont été préalablement validés par comparaison à des espaces perceptifs, peuvent être utilisés pour émettre des hypothèses sur notre perception de la reproduction sonore dans différentes conditions. Klippel [10][48] teste ainsi l'influence des conditions d'écoute et des extraits musicaux considérés à partir d'évaluations selon ses attributs objectifs. Il est ensuite possible de valider les hypothèses émises par des tests psychoacoustiques.

Même si une dimension telle que la réverbération ou la sensation d'espace semble avoir eu une influence importante lors de notre dernière expérience, nous ne prétendons pas pouvoir étudier la composante spatiale de la reproduction sonore à l'aide de notre protocole expérimental actuel.

Une approche du même type, basée sur les différences perçues, pourrait être intéressante pour une telle étude. Le choix du "restituteur" à utiliser pour les tests d'écoute est par contre très délicat.

Notre protocole expérimental permet a priori d'étudier un certain nombre dimensions de la perception sonore. L'importance relative des différentes dimensions perceptives peut être quantifiée, de manière à ce qu'un constructeur soit informé de l'ordre de priorité des critères qu'il doit considérer lors de l'élaboration de son produit.

V.3 Mesures de réponses impulsionnelles

Les résultats de notre première série d'expériences semblent indiquer que l'analyse globale sur l'ensemble du signal, effectuée lors de la détermination de la densité de sonie totale des enregistrements, n'est pas suffisante, et qu'il faut tenir compte du caractère non-stationnaire des signaux lors de l'analyse objective. D'un autre côté, les dimensions perceptives mises en jeu lors des tests d'écoute apparaissent caractéristiques des enceintes plutôt que des extraits musicaux utilisés. Il est donc intéressant d'étudier comment nous pouvons utiliser l'information contenue dans les réponses impulsionnelles ou fréquentielles des enceintes acoustiques, et dans quelle mesure il est possible de s'affranchir des extraits musicaux lors de l'analyse objective.

Lors des séances d'enregistrements, en plus des extraits musicaux, nous avons utilisé des signaux de mesure — des séquences MLS pour Maximum Length Sequence [93] dont le spectre est celui d'un bruit blanc, et des sinus glissants [94][95]. Même si ces enregistrements n'ont pas encore été exploités pour évaluer les réponses impulsionnelles des enceintes, nous souhaitons réaliser un test d'écoute à partir de ces signaux de bruit. Ceci revient en fait à faire "écouter directement" les réponses fréquentielles des enceintes aux auditeurs. Les dissemblances perceptives obtenues seront ensuite comparées à celles impliquant des extraits musicaux. Les séquences de bruit correspondent à un signal stationnaire. La comparaison des résultats obtenus à partir des différents signaux permettra d'étudier l'influence du caractère non-stationnaire des signaux musicaux. Est-ce que les dimensions perceptives mises en jeu sont les mêmes pour un signal de bruit que pour des signaux musicaux? Si oui, il devrait être possible de définir une évaluation objective des dissemblances qui soit indépendante de l'extrait musical. Si non, il est nécessaire de prendre en compte le caractère non-stationnaire des signaux musicaux. La comparaison objective/perceptive sur ce test est également intéressante. La densité de sonie totale et les densités de sonie moyennes évaluées au cours du temps doivent fournir les mêmes résultats sur un signal stationnaire de bruit. Est ce que ce résultat sera en accord avec les dissemblances perceptives résultant du test psychoacoustique? Savoir qu'une partie au moins des dimensions perceptives peut être expliquée en faisant abstraction du caractère non-stationnaire des signaux musicaux serait déjà une information importante.

Lors de tests perceptifs consistant à évaluer différentes modifications fréquentielles de signaux selon des dimensions perceptives choisies, Gabrielsson *et coll.* [64] ont trouvé des résultats similaires pour des signaux de bruit rose et de jazz. Ils indiquent que les spectres de ces signaux étaient proches. Nos séquences de bruit enregistrées devront être filtrées avant le test, par un filtre dont le gabarit correspond à celui d'un "extrait musical moyen" [10][48]. Ceci permettra de s'assurer que la gamme fréquentielle du signal utile soit à peu près la même pour les tests impliquant les différents signaux, et que nous n'étudions pas simplement l'influence de différences de spectre du signal d'excitation des enceintes. Même si ces différences ne seront pas complètement annulées, il s'agira de les minimiser autant que possible.

Lors de cette expérience, il sera important de garder à l'esprit que la différence de signaux considérée peut influencer la perception des auditeurs, mais aussi le comportement des enceintes acoustiques soumises à des sollicitations différentes.

Une deuxième expérience est envisagée. Nous voulons utiliser les signaux de mesure enregistrés pour évaluer la réponse impulsionnelle de nos enceintes dans les conditions d'enregistrement. Ces réponses seront convoluées par les extraits musicaux originaux utilisés lors des enregistrements, afin de "simuler" ces enregistrements. Les enregistrements simulés ne font pas intervenir l'enceinte mais sa réponse impulsionnelle, c'est-à-dire à son approximation linéaire. Comparer les enregistrements réels et simulés permettra donc de comparer l'enceinte acoustique à son filtre linéaire équivalent, et d'étudier ainsi l'influence potentielle de ses composantes non-linéaires.

Ces deux expériences ont pour but d'étudier la possibilité de caractériser les enceintes acoustiques indépendamment de la musique qu'elles reproduisent. Est-ce qu'il est possible de définir une "réponse en fréquence perçue" des enceintes? Est-ce que ce concept a un sens?

V.4 Enceintes acoustiques

Notre protocole expérimental paraît propice à l'étude de la dynamique des enceintes et de leur interaction avec la salle. Le lien entre perception, signal acoustique et paramètres physiques de l'enceinte peut également être exploré.

V.4.1 Dynamique

Une enceinte acoustique n'est pas un système linéaire, et ne se comporte pas de la même façon suivant le niveau de sollicitation auquel elle est soumise. La dynamique d'une enceinte est par définition le rapport entre le niveau le plus fort et le niveau le plus faible qu'elle est capable de reproduire linéairement. Ce terme est également employé pour caractériser la capacité de l'enceinte à répondre de la même façon pour les forts et faibles niveaux de sollicitation [30][31]. Il serait intéressant d'étudier la perception du champ acoustique rayonné par des enceintes, en fonction du niveau auquel elles sont sollicitées.

Nous souhaitons pour cela enregistrer ce champ acoustique produit à différents niveaux de sollicitation. Les enregistrements seront ensuite égalisés avant les tests d'écoute, pour que ce soit les différences de comportement en fonction du niveau, et pas les différences de niveau, qui fassent l'objet de l'étude. Il faudra également réaliser les enregistrements dans un environnement où le bruit de fond est faible. Les enceintes n'étant pas utilisées au même niveau sonore, le rapport signal/bruit ne sera pas le même pour tous les enregistrements, et nous devons éviter que des différences de niveau de bruit de fond n'apparaissent lors de l'égalisation de sonie. Il sera ensuite possible de comparer les dissemblances relatives entre enceintes en fonction du niveau sonore de la reproduction, et également de comparer une même enceinte fonctionnant à plusieurs niveaux. Il conviendra de ne pas impliquer les petites enceintes qui sont conçues pour ne fonctionner qu'à faible niveau.

Staffeldt [26] a réalisé des tests à deux niveaux de reproduction différents et n'a pas trouvé d'effet lié à ce paramètre. Il semble pourtant qu'au moins certaines enceintes se comportent différemment en fonction du niveau. Alors qu'à un niveau suffisamment élevé elles ont une réponse "équilibrée", ceci ne sera pas forcément le cas à des niveaux plus faibles, notamment concernant la restitution des basses fréquences. Les différences intervenant aux très basses fréquences ne pourront par contre a priori pas être mises en évidence par notre protocole utilisant une reproduction au casque, ces fréquences étant davantage perçues par l'ensemble du corps que seulement par les tympans.

Les tests d'écoute sur les enceintes acoustiques nécessitent tous une égalisation du niveau de reproduction, qui permet d'éviter que des différences de sonie ne masquent des différences plus fines. Le niveau sonore commun utilisé ne correspond pas forcément pour toutes les enceintes à leur niveau de fonctionnement optimal. Cette limitation peut être contournée par notre protocole expérimental, si la contrainte liée à un faible niveau de bruit de fond lors des enregistrements est respectée, les enceintes pouvant alors toutes être utilisées à leur niveau nominal. Nous n'avons pas su profiter de cet avantage lors de nos expériences, mais ceci pourra être fait dans le futur.

V.4.2 Interaction avec la salle

La dimension perceptive liée à la réverbération mise en évidence lors de notre troisième expérience indique qu'il y a une interaction des enceintes avec la pièce, et que cette interaction est un critère important utilisé par les auditeurs pour différencier nos enregistrements. Nous pouvons étudier cette interaction en réalisant des enregistrements à l'aide de microphones placés à différents endroits dans la pièce.

Pendant les séances d'enregistrement, en plus des microphones de prise de son utilisés dans le cadre des tests psychoacoustiques, nous avons disposé plusieurs microphones de mesure dans la salle d'écoute. Ces microphones sont visibles sur les photographies des expériences présentées

dans l'annexe A.4. Ils peuvent être utilisés pour mettre en évidence l'influence de la salle sur les différences relatives entre enceintes. Les enceintes n'interagissant pas toutes de la même façon avec la pièce, ces différences relatives peuvent varier avec la position "d'écoute" dans la salle. Nous n'avons pas réalisé d'étude complète sur la question, mais un exemple montrant l'intérêt de notre protocole expérimental pour effectuer une telle étude est présenté. Il consiste à évaluer, pour chaque microphone, les dissemblances objectives entre les enregistrements impliquant les différentes enceintes. Les espaces objectifs obtenus à partir de ces dissemblances déterminées en différents points de la pièce sont ensuite comparés.

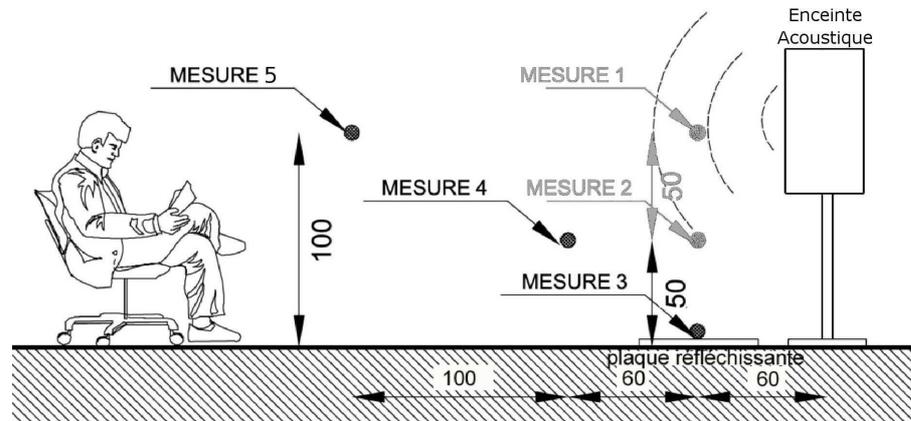


FIGURE V.1 – Microphones de mesure utilisés lors de notre première série d'expériences, en plus des différentes prises de son testées dans le cadre des tests psychoacoustiques. Les microphones de prise de son étaient placés au niveau du microphone de mesure 5 (annexe A.4).

Nous considérons ici les enregistrements réalisés dans le cadre de notre première série d'expériences (paragraphe III.1.1, page III.1.1). La Figure V.1 indique la position des microphones de mesure qui étaient placés dans le plan contenant l'enceinte et les microphones de prise de son. Le microphone de mesure 5 était situé dans la zone d'écoute où se trouvaient également les microphones de prise de son. Nos analyses objectives ont été appliquées aux signaux provenant des microphones 3, 4 et 5. Seuls les enregistrements correspondant à l'extrait musical Kan'nida ont été analysés.

La densité de sonie 2 moyenne des signaux a été évaluée, et les dissemblances objectives entre ces densités ont été calculées. La Figure V.2 présente les espaces objectifs obtenus par l'analyse multidimensionnelle de ces dissemblances. Aucun test psychoacoustique n'a été réalisé sur les signaux impliqués. L'espace perceptif obtenu à partir du même extrait musical et de la prise de son AB ORTF est tout de même présenté à titre de comparaison. Nous constatons que les deux dimensions des espaces objectifs semblent pouvoir être interprétées de manière similaire aux dimensions perceptives. Le changement important de position de l'enceinte 5, entre l'espace

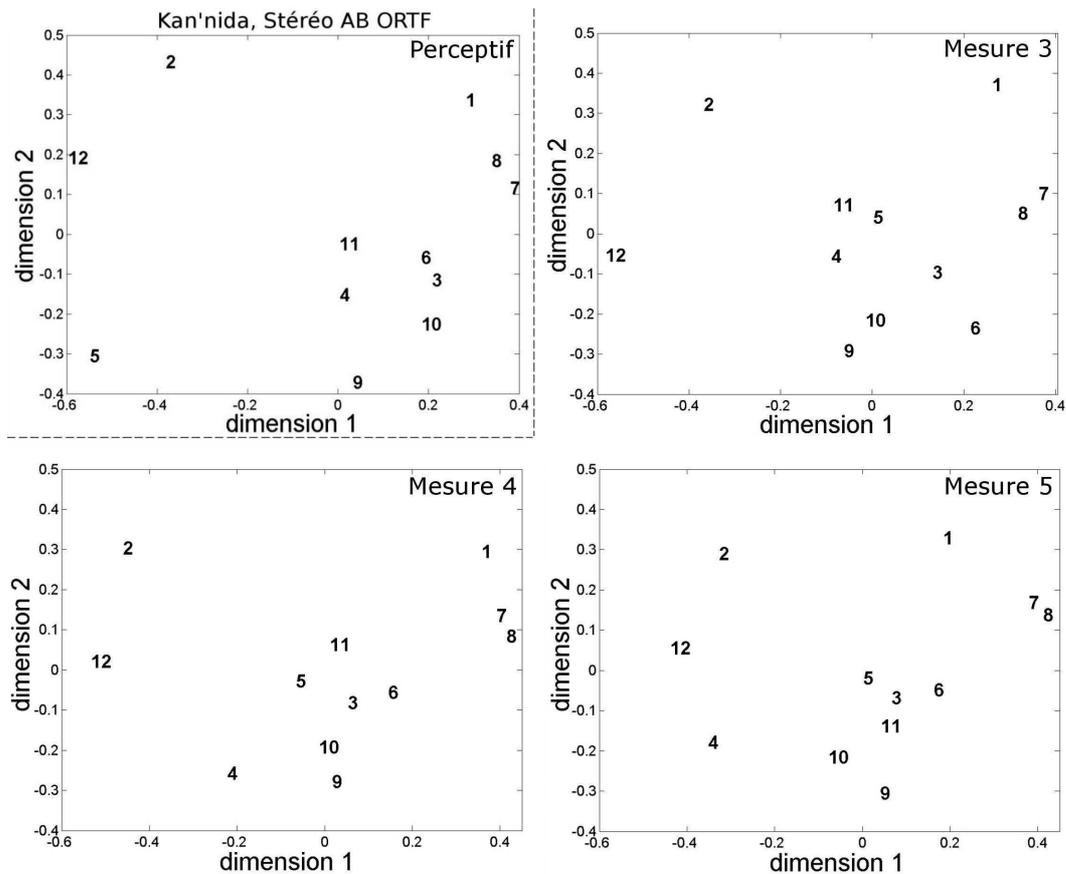


FIGURE V.2 – *Espaces objectifs obtenus à partir de la densité de sonie 2 moyenne des enregistrements utilisant l'extrait musical Kan'nida, provenant des microphones de mesure 3, 4 et 5 de notre première série d'enregistrements. L'espace perceptif obtenu à partir du même extrait musical et de la prise de son AB ORTF est rappelé pour comparaison.*

perceptif et les trois espaces objectifs, peut s'expliquer par la spécificité associée à cette enceinte lors de nos expériences précédentes (chapitre IV).

Nous nous intéressons ici à la comparaison des positions relatives des enceintes au sein des espaces objectifs, en fonction de la position de mesure considérée dans la salle d'écoute. Les positions relatives des enceintes 2, 5, 7, 8 ou 9 restent très stables lorsque nous changeons de position de mesure. La position de l'enceinte 4 et dans une moindre mesure celles des enceintes 1 et 6 semblent par contre plus influencées par la position du microphone considéré. Ces mouvements au sein de l'espace objectif pourraient être caractéristiques du fait que les différentes enceintes n'interagissent pas toutes de la même façon avec la pièce.

Bien sûr une étude beaucoup plus complète est nécessaire afin d'étudier cette interaction des enceintes avec la salle d'écoute. Cette étude pourrait être réalisée à partir d'enregistrements effectués à plusieurs positions, dans plusieurs pièces. Lors des tests d'écoute sur les enceintes,

c'est l'ensemble enceinte-salle qui est évalué. Le poids relatif des caractéristiques propres aux enceintes et à leur interaction avec la salle devra être déterminé.

V.4.3 Lien avec les paramètres physiques

Notre étude s'est concentrée sur le lien entre le signal acoustique délivré par les enceintes lors des enregistrements et la perception qu'ont eu les auditeurs de ces enregistrements. Nous ne sommes pas encore remonté aux paramètres physiques des enceintes responsables des différences de reproduction et de perception.

Certaines études ont employé des enceintes dont les caractéristiques physiques pouvaient être modifiées [7][10][19][48][56], principalement dans le but d'explorer les dimensions perceptives liées à la reproduction sonore. De tels prototypes pourrait également être utilisés pour faire le lien entre les dimensions perceptives, les caractéristiques du signal acoustique émis, et les paramètres physiques de l'enceinte.

Notre but à terme est de définir un outil d'analyse objective qui permette au constructeur d'évaluer l'effet "perçu" des modifications qu'il apporte à ces produits. Il sera alors possible de hiérarchiser l'importance relative des modifications apportées. Il s'agira ensuite de faire le lien avec les appréciations de qualité globale de ces produits (Figure II.1, page II.1).

V.5 Bilan

Notre recherche s'est effectuée à l'intersection de plusieurs domaines complémentaires. Dans ce chapitre, les expériences envisagées ont été présentées par domaine — analyse du signal audio, psychoacoustique, mesures de réponses impulsionnelles et caractéristiques des enceintes acoustiques. Les recherches au sein de chacun de ces domaines sont dépendantes les unes des autres, les avancées dans un domaine posant des questions dans un autre.

Le test d'écoute que nous souhaitons réaliser à partir des enregistrements de séquences MLS constitue un bon exemple de cette complémentarité. Avant de pouvoir réaliser ce test, nous devons déterminer l'influence potentielle de la tâche perceptive qui sera utilisée lors du test psychoacoustique. L'intérêt de cette expérience utilisant des reproductions de signaux de bruit est de comparer la perception des enceintes sur des signaux stationnaire et non-stationnaires. Cette comparaison apportera des informations concernant le degré de complexité de l'analyse du signal audio qu'il est nécessaire d'utiliser pour retrouver les différences perçues par les auditeurs. La possibilité d'observer directement les réponses impulsionnelles des enceintes afin de les caractériser dépendra de cette complexité.

Le travail présenté dans ce document constitue le premier pas de ce projet de recherche qui reste à poursuivre au sein de ces différents domaines.

Conclusion

Les recherches présentées dans ce document portent sur la confrontation d'évaluations objectives et perceptives de différences relatives entre enceintes acoustiques. Ce travail constituant le lancement de cette thématique de recherche au sein du laboratoire, nous avons dû commencer par définir un protocole expérimental approprié et par tester l'influence des différents paramètres expérimentaux impliqués. Nous avons décidé de nous concentrer dans un premier temps sur l'évaluation des différences de restitution du timbre par les enceintes acoustiques. Notre but étant le rapprochement objectif/perceptif, nous n'avons pas tenté d'apporter de nouvelles connaissances concernant la perception de la reproduction sonore ou les mesures objectives d'enceintes acoustiques. Le lien entre ces deux domaines a par contre fait l'objet de notre recherche.

Etant données les difficultés rencontrées dans la littérature pour rapprocher a posteriori des mesures physiques et perceptives au départ disjointes, nous avons tenté de garder nos évaluations objectives et perceptives le plus proche possible en les menant en parallèle. Nous avons pour cela mis au point un protocole expérimental qui nous permet de répondre aux exigences de contrôle des tests d'écoute et de comparer rapidement des enceintes acoustiques placées à la même position dans une salle d'écoute. Il consiste à enregistrer le champ acoustique rayonné par les différentes enceintes. Les enregistrements sont ensuite soumis à la fois à des tests perceptifs réalisés au casque et à différentes techniques d'analyse objective. Les tests psychoacoustiques ont pour but d'évaluer les dissemblances perceptives entre les enregistrements. De manière parallèle, nous avons défini une "métrique" associée à chaque technique d'analyse objective, afin de calculer pour chacune d'elles les dissemblances objectives entre les enregistrements. Certaines de nos analyses objectives font intervenir des modèles auditifs permettant de déterminer la densité de sonie de nos signaux en tenant compte du fonctionnement complexe du système auditif. Une métrique a ainsi été proposée pour évaluer objectivement la dissemblance de deux densités de sonie.

Les dissemblances objectives et perceptives ont été systématiquement comparées au cours d'études impliquant de nombreuses conditions expérimentales différentes. Cette comparaison a consisté dans un premier temps à évaluer la corrélation entre dissemblances objectives et perceptives. Les espaces objectifs et perceptifs résultant de l'analyse multidimensionnelle de ces dissemblances ont ensuite été comparés. Le but de ces comparaisons était de rechercher une

technique d'analyse et une métrique associée permettant de retrouver les différences perçues par les auditeurs, c'est-à-dire d'extraire des signaux l'information pertinente d'un point de vue perceptif.

Nos évaluations de dissemblances sont effectuées avec une prise de son et un extrait musical donnés, sur des enceintes utilisées dans une salle selon un certain mode de reproduction. Chacun de ces paramètres expérimentaux influence en partie le résultat de l'évaluation. Par contre, nous avons montré que la comparaison objectif/perceptif restait suffisamment indépendante de ces paramètres, les mêmes tendances ayant été observées lors de toutes nos expériences.

Parmi les analyses objectives testées jusqu'à présent, les densités de sonie ont systématiquement fourni les dissemblances objectives les plus corrélées avec les dissemblances perceptives. Nous avons ainsi pu quantifier l'intérêt d'utiliser un modèle auditif en vue d'une discrimination plus pertinente de la reproduction sonore. L'analyse multidimensionnelle des dissemblances perceptives a révélé dans un premier temps deux dimensions principales utilisées par les auditeurs lors des tests psychoacoustiques. Les densités de sonie ont permis de retrouver ces deux dimensions pour la totalité de nos signaux. D'autres analyses et d'autres métriques doivent bien sûr être testées, mais ce premier rapprochement des dissemblances objectives et perceptives est encourageant. Notre première expérience semble indiquer que l'analyse globale sur l'ensemble du signal, effectuée lors de la détermination de la densité de sonie totale des enregistrements, ne serait pas suffisante. La dépendance des phénomènes d'étalement spectral liés au fonctionnement du système auditif vis-à-vis de l'évolution temporelle des signaux devrait donc être prise en compte. Ceci remettrait en cause la pertinence de l'observation directe des réponses en fréquence des enceintes acoustiques. Ce résultat reste à confirmer par de nouvelles données expérimentales.

Notre protocole basé sur des enregistrements présente l'avantage de ne pas limiter le nombre d'enceintes pouvant être impliquées au sein d'un même test. Le vaste panel d'enceintes considéré lors de notre dernière expérience nous a permis de mettre en évidence deux dimensions perceptives supplémentaires. Les densités de sonie ont une nouvelle fois fourni les dissemblances objectives les plus corrélées avec les dissemblances perceptives, mais elles ont conduit à des espaces objectifs insuffisants pour décrire les deux nouvelles dimensions perceptives. Cette expérience nous a donc donné l'occasion d'explorer les limitations de notre analyse objective. Celle-ci devra être affinée pour décrire complètement la perception de nos auditeurs.

Après avoir vérifié que la comparaison objectif/perceptif ne dépendait pas des paramètres expérimentaux testés, nous avons également constaté que les dimensions perceptives impliquées dans les jugements des auditeurs gardaient la même interprétation lors des différentes expériences. Ces dimensions ont notamment été les mêmes pour trois extraits musicaux différents. Bien sûr les évaluations perceptives sont influencées par les paramètres expérimentaux — c'est aussi le cas des évaluations objectives. Mais cette influence s'est avérée limitée. Les dimensions principales utilisées par les auditeurs lors de nos tests psychoacoustiques semblent bien carac-

téristiques des enceintes acoustiques. Il s'agit donc maintenant de déterminer une évaluation objective des enceintes qui reste pertinente, mais qui soit moins dépendante des signaux utilisés et plus caractéristique des enceintes.

Des attributs objectifs préliminaires ont été proposés pour décrire la nature de nos deux premières dimensions perceptives. Notre approche pour déterminer de tels attributs consiste à identifier dans un premier temps la technique d'analyse qui semble la plus appropriée pour observer les signaux. Ensuite seulement nous utilisons cette analyse pour rechercher des attributs objectifs décrivant nos dimensions perceptives.

Notre dernière expérience, basée sur une tâche de classification libre plutôt que sur des comparaisons par paires, a soulevé des questions concernant l'influence de la tâche perceptive impliquée dans nos tests psychoacoustiques. Nous devons répondre à ces questions pour pouvoir exploiter les résultats de tests perceptifs faisant intervenir un nombre important de stimuli. Ces tests sont nécessaires pour définir des attributs objectifs qui soient pertinents.

Ces travaux ont été présentés lors de quatre conférences [96][97][98][99]. Ils ont ouvert un certain nombre de perspectives, et plusieurs expériences sont d'ores et déjà envisagées pour tirer profit des avantages de notre protocole.

BIBLIOGRAPHIE

- [1] Acta Acustica Special Issue, « Aurally-adequate sound-quality evaluation », *Acta Acustica*, **83**(5), pp. 747–826, 1997.
- [2] Acta Acustica Special Issue, « Auditory quality of systems », *Acta Acustica*, **90**(6), pp. 999–1130, 2004.
- [3] IEC Publication 60268-5, « Sound system equipment - Part 5 : Loudspeakers », International Electrotechnical Commission, Geneva, Switzerland, 1989.
- [4] Norme NF 61305 5, « Équipements et systèmes audio grand public haute fidélité. Méthodes pour mesurer et spécifier les performances - Partie 5 : Haut-parleurs », Association Française de Normalisation (AFNOR), 2005.
- [5] IEC Publication 60581-7, « High fidelity audio equipment and systems. Minimum performance requirements - Part 7 : Loudspeakers », International Electrotechnical Commission, Geneva, Switzerland, 1986.
- [6] Norme NF C97-405, « Matériels et systèmes haute fidélité. Enceintes acoustiques - Règles », Association Française de Normalisation (AFNOR), 1980.
- [7] A. Gabrielsson, H. Sjögren, « Perceived sound quality of sound-reproducing systems », *J. Acoust. Soc. Am.*, **65**(4), pp. 1019–1033, 1979.
- [8] F. Toole, « Loudspeaker measurements and their relationship to listener preferences : Part 1 », *J. Audio Eng. Soc.*, **34**(4), pp. 227–235, 1986.
- [9] L. Bramsløw, « An objective estimate of the perceived quality of reproduced sound in normal and impaired hearing », *Acta Acustica*, **90**, pp. 1007–1018, november/december 2004.
- [10] W. Klippel, « Multidimensional relationship between subjective listening impression and objective loudspeaker parameters », *Acustica*, **70**, pp. 45–54, march 1990.
- [11] A. Gabrielsson, B. Lindström, O. Till, « Loudspeaker frequency response and perceived sound quality », *J. Acoust. Soc. Am.*, **90**(2, Pt. 1), pp. 707–719, 1991.
- [12] S. Lipshitz, J. Vanderkooy, « The great debate : Subjective evaluation », *J. Audio Eng. Soc.*, **29**(7/8), pp. 482–491, 1981.

- [13] F. Toole, « Listening tests - Turning opinion into facts », *J. Audio Eng. Soc.*, **30**(6), pp. 431–445, 1982.
- [14] F. Toole, « Subjective measurements of loudspeaker: sound quality and listener performance », *J. Audio Eng. Soc.*, **33**(1/2), pp. 2–32, 1985.
- [15] S. Bech, « Perception of timbre of reproduced sound in small rooms: Influence of room and loudspeaker position », *J. Audio Eng. Soc.*, **42**(12), pp. 999–1007, 1994.
- [16] S. Olive, P. Schuck, S. Sally, M. Bonneville, « The effect of loudspeaker placement on listener preference ratings », *J. Audio Eng. Soc.*, **42**(9), pp. 651–669, 1994.
- [17] F. Toole, « Subjective measurements of loudspeakers - A comparison of stereo and mono listening », in *AES 74th Convention*, no. 2023 (H-5), october 1983.
- [18] F. Toole, S. Olive, « The modification of timbre by resonances: perception and measurement », *J. Audio Eng. Soc.*, **36**(3), pp. 122–142, 1988.
- [19] A. Gabrielsson, U. Rosenberg, H. Sjögren, « Judgments and dimension analyses of perceived sound quality of sound-reproducing systems », *J. Acoust. Soc. Am.*, **55**(4), pp. 854–861, 1974.
- [20] H. Eisler, « Measurement of perceived acoustic quality of sound-reproducing systems by means of factor analysis », *J. Acoust. Soc. Am.*, **39**(3), pp. 484–492, 1966.
- [21] A. Illényi, P. Korpássy, « Correlation between loudness and quality of stereophonic loudspeakers », *Acustica*, **49**(4), pp. 334–336, 1981.
- [22] A. Gabrielsson, B. Lindstrom, « Perceived sound quality of high-fidelity loudspeakers », *J. Audio Eng. Soc.*, **33**(1/2), pp. 33–53, 1985.
- [23] T. Nousaine, « Can you trust your ears? », in *AES 91st Convention*, no. 3177 (L-3), october 1991.
- [24] P. Susini, S. McAdams, S. Winsberg, « A multidimensional technique for sound quality assessment », *Acta Acustica*, **85**, pp. 650–656, 1999.
- [25] I. Boulet. *La sonie des sons impulsionsnels: perception, mesures et modèles*. Thèse de Doctorat, Université de la Méditerranée - Aix-Marseille II, 2005.
- [26] H. Staffeldt, « Correlation between subjective and objective data for quality loudspeakers », *J. Audio Eng. Soc.*, **22**, pp. 402–415, july/august 1974.
- [27] S. Bech, « Selection and training of subjects for listening tests on sound-reproducing equipment », *J. Audio Eng. Soc.*, **40**(7/8), pp. 590–610, 1992.
- [28] S. Olive, « Differences in performance and preference of trained versus untrained listeners in loudspeaker tests: a case study », *J. Audio Eng. Soc.*, **51**, pp. 806–825, september 2003.
- [29] H. Fastl, « The psychoacoustics of sound-quality evaluation », *Acta Acustica*, **83**(5), pp. 754–764, 1997.

-
- [30] IEC Publication 60268-13, « Sound system equipment - Part 13 : Listening tests on loudspeakers », International Electrotechnical Commission, Geneva, Switzerland, 1998.
- [31] AES20-1996, « AES recommended practice for professional audio - Subjective evaluation of loudspeakers », *J. Audio Eng. Soc.*, **44**(5), pp. 382–400, 1996.
- [32] AES Staff Writer, « New horizons in listening test design », *J. Audio Eng. Soc.*, **52**, pp. 65–73, january/february 2004.
- [33] F. Rumsey, « Subjective assessment of spatial attributes of reproduced sound », in *AES 15th International Conference*, pp. 122–135, october/november 1998.
- [34] F. Rumsey, « Spatial quality evaluation for reproduced sound : terminology, meaning, and a scene-based paradigm », *J. Audio Eng. Soc.*, **50**, pp. 651–666, september 2002.
- [35] S. Bech, « Spatial aspects of reproduced sound in small rooms », *J. Acoust. Soc. Am.*, **103**, pp. 434–445, january 1998.
- [36] S. Bech, « Methods for the subjective evaluation of spatial characteristics of sound », in *AES 16th International Conference*, pp. 487–504, april 1999.
- [37] S. Zielinski, F. Rumsey, S. Bech, « Effects of down-mix algorithms on quality of surround sound », *J. Audio Eng. Soc.*, **51**, pp. 780–798, september 2003.
- [38] D. Griesinger, « The psychoacoustics of apparent source width, spaciousness and envelopment in performance spaces », *Acta Acustica*, **83**, pp. 721–731, 1997.
- [39] I. Borg, P. Groenen, *Modern multidimensional scaling. Theory and applications*. Springer, 1997.
- [40] R. Guski, « Psychological methods for evaluating sound quality and assessing acoustic information », *Acta Acustica*, **83**, pp. 765–774, 1997.
- [41] W. Dillon, M. Goldstein, *Multivariate analysis. Methods and applications*. John Wiley & Sons, 1984.
- [42] T. Nakayama, T. Miura, O. Kosaka, M. Okamoto, T. Shiga, « Subjective assessment of multichannel reproduction », *J. Audio Eng. Soc.*, **19**, pp. 744–751, october 1971.
- [43] H. Staffeldt, « Differences in the perceived quality of loudspeaker sound reproduction caused by the loudspeaker-room-listener interactions », in *AES 90th Convention*, no. 3046 (H-1), february 1991.
- [44] F. Toole, « Binaural record/reproduction systems and their use in psychoacoustic investigations », in *AES 91st Convention*, no. 3179 (L-6), october 1991.
- [45] K. Gundry, S. Olive, « Comments on "Differences in performance and preference of trained versus untrained listeners in loudspeaker tests : a case study" and author's reply », *J. Audio Eng. Soc.*, **52**, pp. 261–262, march 2004.
- [46] S. Olive, « A multiple regression model for predicting loudspeaker preference using objec-

- tive measurements: Part 1 - Listening test results », in *AES 116th Convention*, no. 6113, may 2004.
- [47] B. Moore, C. Tan, « Perceived naturalness of spectrally distorted speech and music », *J. Acoust. Soc. Am.*, **114**, pp. 408–419, july 2003.
- [48] W. Klippel, « Assessing the subjectively perceived loudspeaker quality on the basis of objective parameters », in *AES 88th Convention*, no. 2929 (J5), march 1990.
- [49] F. Guyot. *Etude de la perception sonore en termes de reconnaissance et d'appréciation qualitative : une approche par la catégorisation*. Thèse de Doctorat, Université du Maine, 1996.
- [50] J. Kruskal, M. Wish, *Multidimensional scaling*. Sage Publications, 1978.
- [51] J. Marozeau. *L'effet de la fréquence fondamentale sur le timbre*. Thèse de Doctorat, Université Pierre et Marie Curie - Paris VI, 2004.
- [52] R. M. Aarts, « A new method for the design of crossover filters », *J. Audio Eng. Soc.*, **37**, pp. 445–454, june 1989.
- [53] U. Jekosch, « Basic concepts and terms of "quality" reconsidered in the context of product-sound quality », *Acta Acustica*, **90**, pp. 999–1006, november/december 2004.
- [54] S. Bech, « Requirements for low-frequency sound reproduction, Part 1 : The audibility of changes in passband amplitude ripple and lower system cutoff frequency and slope », *J. Audio Eng. Soc.*, **50**(7/8), pp. 564–580, 2002.
- [55] J. Pedersen, A. Mäkivirta, « Requirements for low-frequency sound reproduction, Part 2 : Generation of stimuli and listening system equalization », *J. Audio Eng. Soc.*, **50**(7/8), pp. 581–593, 2002.
- [56] S. Olive, « A multiple regression model for predicting loudspeaker preference using objective measurements: Part 2 - Development of the model », in *AES 117th Convention*, no. 6190, october 2004.
- [57] R. Bücklein, « The audibility of frequency response irregularities », *J. Audio Eng. Soc.*, **29**, pp. 126–131, march 1981.
- [58] C. Janse, A. Kaiser, « Time-frequency distributions of loudspeakers: The application of the Wigner distribution », *J. Audio Eng. Soc.*, **31**, pp. 198–223, april 1983.
- [59] C. Janse, A. Kaiser, « The Wigner distribution: A valuable tool for investigating transient distortion », *J. Audio Eng. Soc.*, **32**, pp. 868–882, november 1984.
- [60] R. Heyser, « The delay plane, objective analysis of subjective properties: Part 1 », *J. Audio Eng. Soc.*, **21**, pp. 690–701, november 1973.
- [61] E. Paulus, E. Zwicker, « Programme zur automatischen bestimmung der lautheit aus terzpegeln oder frequenzgruppenpegeln », *Acustica*, **27**(5), pp. 253–266, 1972.

-
- [62] E. Zwicker, H. Fastl, C. Dallmayr, « BASIC-Program for calculating the loudness of sounds from their 1/3-oct band spectra according to ISO 532 B », *Acustica*, **55**, pp. 63–67, 1984.
- [63] H. Staffeldt, « Measurement and prediction of the timbre of sound reproduction », *J. Audio Eng. Soc.*, **32**(6), pp. 410–414, 1984.
- [64] A. Gabrielsson, B. Hagerman, T. Bech-Kristensen, G. Lundberg, « Perceived sound quality of reproductions with different frequency responses and sound level », *J. Acoust. Soc. Am.*, **88**(3), pp. 1359–1366, 1990.
- [65] E. Zwicker, R. Feldtkeller, *Psychoacoustique : l'oreille, récepteur d'information*. Masson, 1981.
- [66] E. Zwicker, H. Fastl, *Psychoacoustics : facts and models*. Springer, 1999.
- [67] M. Botte, G. Canévet, L. Demany, C. Sorin, *Psychoacoustique et perception auditive*. Inserm-SFA-CNET, 1989.
- [68] F. Toole, « Loudspeaker measurements and their relationship to listener preferences : Part 2 », *J. Audio Eng. Soc.*, **34**(5), pp. 323–348, 1986.
- [69] J. Kates, « A perceptual criterion for loudspeaker evaluation », *J. Audio Eng. Soc.*, **32**(12), pp. 938–945, 1984.
- [70] E. Zwicker, U. Zwicker, « Audio engineering and psychoacoustics : Matching signals to the final receiver, the human auditory system », *J. Audio Eng. Soc.*, **39**, pp. 115–126, march 1991.
- [71] M. Burkhard, K. Genult, « Merging subjective and objective acoustical measurements », in *AES 11th International Conference*, pp. 105–113, may 1992.
- [72] S. Bech, « Timbral aspects of reproduced sound in small rooms. 1 », *J. Acoust. Soc. Am.*, **97**, pp. 1717–1726, march 1995.
- [73] S. Bech, « Timbral aspects of reproduced sound in small rooms. 2 », *J. Acoust. Soc. Am.*, **99**, pp. 3539–3549, june 1996.
- [74] S. McAdams, « La reconnaissance de sources et d'événements sonores », in *Penser les sons. Psychologie cognitive de l'audition (S. McAdams et E. Bigand)*, Presses universitaires de France, 1994.
- [75] L. Pols, L. Van Der Kamp, R. Plomp, « Perceptual and physical space of vowel sounds », *J. Acoust. Soc. Am.*, **46**(2), pp. 458–466, 1969.
- [76] O. Houix. *Catégorisation auditive des sources sonores*. Thèse de Doctorat, Université du Maine, 2003.
- [77] G. Lemaitre. *Etude perceptive et acoustique de nouveaux avertisseurs sonores automobiles*. Thèse de Doctorat, Université du Maine, 2004.
- [78] C. Bonnet, *Manuel pratique de psychophysique*. Armand Colin, 1986.

- [79] W. Gaver, « What in the world do we hear? : an ecological approach to auditory event perception », *Ecological psychology*, **5**(1), pp. 1–29, 1993.
- [80] J. Grey, « Multidimensional perceptual scaling of musical timbres », *J. Acoust. Soc. Am.*, **61**, pp. 1270–1277, may 1977.
- [81] E. Bigand, S. Vieillard, F. Madurell, J. Marozeau, A. Dacquet, « Multidimensional scaling of emotional responses to music : the effect of musical expertise and of the duration of the excerpts », *Cognition and Emotion*, **19**, pp. 1113–1139, 2005.
- [82] British Standard ISO 226 2003, « Acoustics - Normal equal-loudness level contours », BSi, 2003.
- [83] E. Zwicker, H. Fastl, « A portable loudness-meter based on ISO 532B », in *11th International Congress on Acoustics*, pp. 135–137, 1983.
- [84] G. Canévet, D. Habault, S. Meunier, F. Demirdjian, « Auditory perception of sounds radiated by a fluid-loaded vibrating plate excited by a transient point force », *Acta Acustica*, **90**, pp. 181–193, 2004.
- [85] C. Vogel, V. Maffiolo, J.-D. Polack, M. Castellengo, « Validation subjective de la prise de son en extérieur », in *4ème Congrès Français d'Acoustique, Marseille, France*, pp. 307–310, 1997.
- [86] B. Moore, B. Glasberg, T. Baer, « A model for the prediction of thresholds, loudness, and partial loudness », *J. Audio Eng. Soc.*, **45**, pp. 224–240, april 1997.
- [87] B. Moore. Communication personnelle, septembre 2005.
- [88] M. Morimoto, « The relation between spatial and cross-correlation measures », in *15th International Congress on Acoustics*, pp. 581–584, june 1995.
- [89] H. Fastl, W. Ellermeier. Communication personnelle, septembre 2005.
- [90] J. Faure. *Influence de paramètres structuraux d'une plaque rayonnante sur la perception sonore*. Thèse de Doctorat, Université Claude Bernard Lyon I, 2003.
- [91] L. Tsogo, M.-H. Masson, A. Bardot, « A metric multidimensional scaling method for large objects sets », in *1997 IEEE International Conference on Systems, Man, and Cybernetics*, 1997.
- [92] A. Bardot, V. Roussarie, L. Tsogo, M.-H. Masson, « Pertinence of sensory evaluation of vehicle sounds : analysis of a multidimensional scaling representation with the classical sensory profiling analysis », in *Inter-Noise 2002*, august 2002.
- [93] D. Rife, J. Vanderkooy, « Transfert-function measurement with Maximum-Length Sequences », *J. Audio Eng. Soc.*, **37**, pp. 419–444, june 1989.
- [94] A. Farina, « Simultaneous measurement of impulse response and distortion with swept-sine technique », in *AES 108th Convention*, no. 5093 (D-4), february 2000.

-
- [95] S. Müller, P. Massarani, « Transfer-function measurement with sweeps », *J. Audio Eng. Soc.*, **49**, pp. 443–471, june 2001.
- [96] M. Lavandier, P. Herzog, S. Meunier, « Perceptual and physical evaluation of loudspeakers », in *Joint congress CFA/DAGA'04*, vol. 2, pp. 1105–1106, march 2004.
- [97] M. Lavandier, P. Herzog, S. Meunier, « Perceptual and physical evaluation of loudspeakers », in *AES 117th Convention*, no. 6240, october 2004.
- [98] M. Lavandier, S. Meunier, P. Herzog, « Perceptual and physical evaluation of differences among a large panel of loudspeakers », in *Forum Acusticum 2005*, no. 430-0, pp. 1689–1694, september 2005.
- [99] M. Lavandier, B. Guyot, S. Meunier, P. Herzog, « The influence of stereophony on the restitution of timbre by loudspeakers », in *AES 119th Convention*, no. 6619, october 2005.
- [100] C. Hugonnet, P. Walder, *Théorie et pratique de la prise de son stéréophonique*. Editions Eyrolles, 1995.
- [101] R. Streicher, W. Dooley, « Basic stereo microphone perspectives - A review », *J. Audio Eng. Soc.*, **33**, pp. 548–556, july/august 1985.
- [102] T. Wonnacott, R. Wonnacott, *Statistique. Economie, gestion, sciences, médecine*. Economica, 1991.
- [103] W. Rand, « Objective criteria for the evaluation of clustering methods », *Journal of the American Statistical Association*, **66**, pp. 846–850, 1971.
- [104] H. Abdi, *Introduction au traitement statistique des données expérimentales*. Grenoble: PUG, 1987.
- [105] S. Winsberg, J. Carroll, « A quasi-nonmetric method for multidimensional scaling via an extended euclidean model », *Psychometrika*, **54**, pp. 217–229, june 1989.
- [106] J. Max, J.-L. Lacoume, *Méthodes et techniques de traitement du signal*. Dunod, 2000.
- [107] A. Oppenheim, R. Schaffer, *Digital signal processing*. Prentice Hall, 1975.
- [108] P. Welch, « The use of the Fast Fourier Transform for the estimation of power spectra : a method based on time averaging over short modified periodograms », *IEEE Trans. Audio Electroacoustics*, **AU-15**, pp. 70–73, june 1967.
- [109] P. Flandrin, *Temps-fréquence*. Traité des nouvelles technologies. Série traitement du signal., Hermès, 1993.

ANNEXE A

Matériel utilisé lors des expériences

————— oOo —————

Cette annexe présente l'ensemble du matériel utilisé lors des séances d'enregistrements et des tests psychoacoustiques. Les signaux considérés dans toute de notre étude sont échantillonnés à 44100 Hz.

————— oOo —————

SOMMAIRE

A.1	Prises de son testées	iii
A.2	Panel d'enceintes considéré	iii
A.3	Matériel d'enregistrement	v
A.4	Photographies des séances d'enregistrements	vi
A.5	Evaluation de la durée de réverbération des salles d'écoute	vii
A.6	Dispositif expérimental lors des tests psychoacoustiques	viii

A.1 Prises de son testées

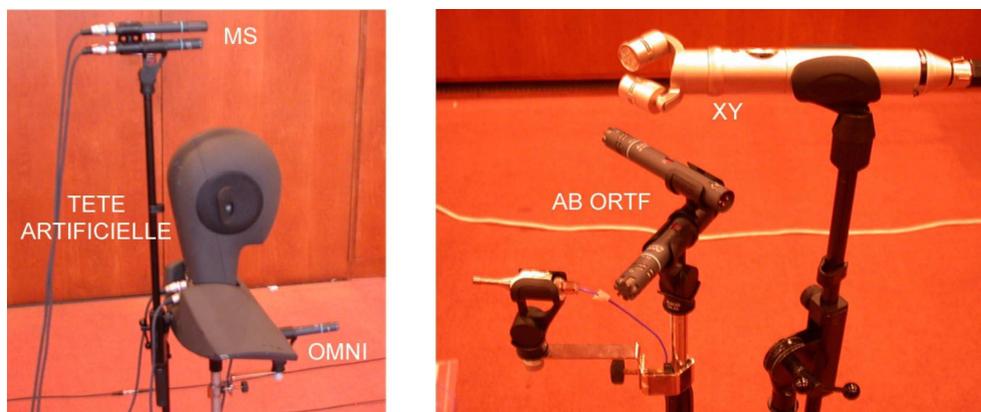


FIGURE A.1 – *Les différentes prises de son testées*

La Figure A.1 présente les prises de son mises en oeuvre lors de notre première série d'expériences : la tête artificielle, les couples stéréophoniques AB ORTF, XY et MS et une prise de son monophonique omnidirectionnelle. La Figure A.2 rappelle le principe des trois couples stéréophoniques et leurs différences. Le couple AB ORTF utilise deux capsules cardioïdes espacées de 17 cm dans le même plan horizontal, leur axes de prise de son formant un angle de 110° . Le couple XY est également composé de deux capsules cardioïdes, mais elles sont cette fois coïncidentes. Elles sont placées l'une au dessus de l'autre, avec un angle de 90° entre leurs axes de prise de son. Le couple MS est formé de deux capsules coïncidentes superposées. L'une est cardioïde et orientée vers la source, elle fournit le signal M pour "middle". L'autre est bidirectionnelle avec ses deux lobes de directivité orientés latéralement, à 90° de l'axe de la capsule cardioïde. Elle fournit le signal S pour "side". Les voies stéréophoniques de gauche et de droite sont obtenues par multiplexage en faisant la somme et la différence de ces deux signaux M et S. La voie monophonique cardioïde centrale M peut être ré-obtenue en sommant ces deux voies stéréophoniques gauche et droite. Nous avons ainsi eu accès à une prise de son monophonique cardioïde.

Le lecteur souhaitant davantage d'information sur ces différentes techniques de prise de son pourra se référer à [100][101].

A.2 Panel d'enceintes considéré

La Figure A.3 montre l'ensemble des enceintes qui ont été impliquées dans nos séances d'enregistrements. Ce panel a fait intervenir des enceintes classiques de type colonne (Audience T3, Audio Référence AR86V, BC Acoustique Tibre, B&W DM603 S3, Ballade 234A3, Cabasse Altura, JM Lab 710 Onyx, Mordaunt Short Avant 906, Tannoy Mercury Mx3) ou bibliothèque (A2T

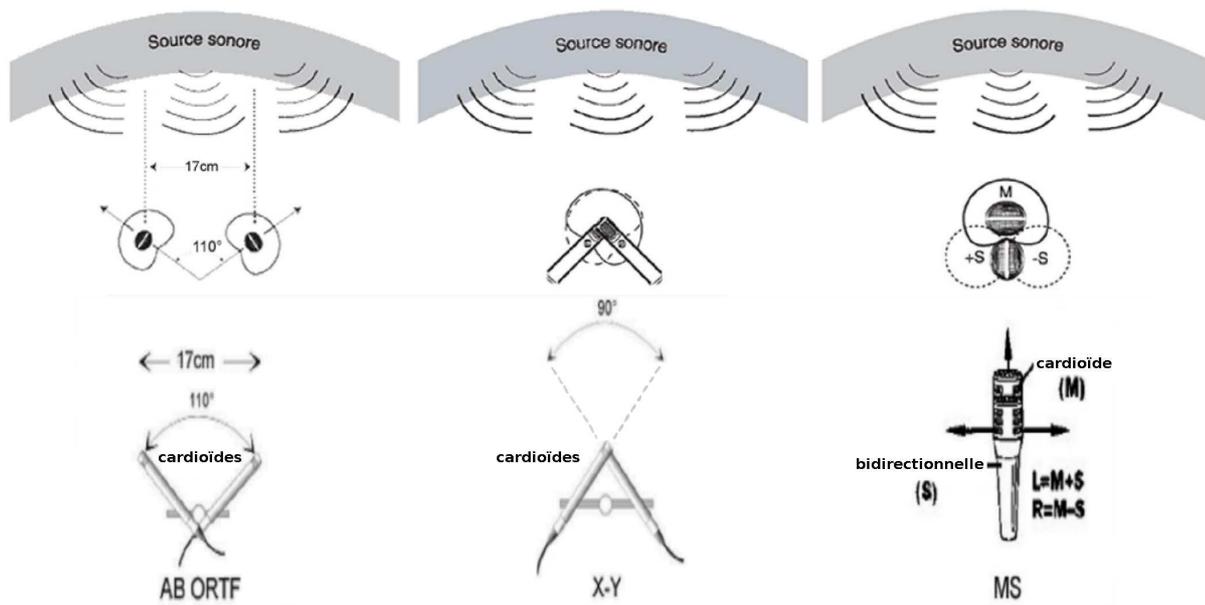


FIGURE A.2 – Principe des trois couples stéréophoniques utilisés

BV4027, Audio Référence AR16V, BC Acoustique Tamise, B&W DM601 S3, JBL 4312, Jamo Alto, Mission 760i, Studio Lab SLB 102, Tannoy 600, Tannoy 800, Tannoy Mercury M1, Waterfall), des enceintes actives de studio (Genélec 1029A, Genélec 1030A, Genélec 1031A, Mackie HR 524) ou multimedia (Altec Lansing Fx602, Fostex 6301B, Yamaha MS101II), un panneau électrostatique (Martin Logan Sequel II), une enceinte de sonorisation (Hortus Audio Pro Mann serie), la technologie NXT (Mission X-Space), des haut-parleurs large bande (Anthony Gallo A'Diva, Supravox), un modèle très haut de gamme (Mosquito Néo) et des enceintes censées représenter des reproductions à bande passante limitée du fait de leur utilisation non conventionnelle (BC Acoustique Tennessee sans tweeter, satellite Bose).

Notre troisième expérience nous a permis d'inclure toutes ces enceintes dans le même test d'écoute, une deuxième JBL 4312 constituant alors la trente-septième enceinte. Notre première série d'expérience a fait intervenir douze enceintes (Audience T3, Fostex 6301B, Genélec 1031A, JBL 4312, Jamo Alto, Mackie HR 524, deux Studio Lab SLB 102, Tannoy 600, Tannoy 800, Tannoy Mercury M1, Yamaha MS101II). Onze paires d'enceintes ont été utilisées dans la deuxième série d'expériences (A2T BV4027, Audience T3, BC Acoustique Tibre, Fostex 6301B, Genélec 1031A, JBL 4312, Jamo Alto, Mordaunt Short Avant 906, Studio Lab SLB 102, Tannoy 600, Yamaha MS101II). Pour trois de ces paires (BC Acoustique Tibre, Studio Lab SLB 102, Tannoy 600), les enceintes gauche et droite ont été impliquées dans les enregistrements en reproduction monophonique.



FIGURE A.3 – Panel d'enceintes considéré lors de nos expériences

A.3 Matériel d'enregistrement

Ce paragraphe dresse la liste du matériel utilisé pendant les séances d'enregistrements.

- Prises de son :
 - Capsule omnidirectionnelle : AKG Blue Line CK-92,
 - Capsules cardioïdes pour les couples AB ORTF et MS : AKG Blue Line CK-91,
 - Capsule bidirectionnelle pour le couple MS : AKG Blue Line CK-94,
 - Corps de microphone : AKG Blue Line SE-300B,
 - Couple stéréophonique XY : Rode NT4,
 - Tête artificielle : Head Acoustics HRS II.2,

- Préamplificateur des microphones : Tascam MX-4,
- Préamplificateur de la tête artificielle : MCN Audio U1126.
- Enregistrement :
 - Carte son : RME DIGI9652 et Fostex VC-8, avec une fréquence d'échantillonnage choisie à 44100 Hz,
 - Logiciel : Cool Edit Pro.
- Chaîne de reproduction haute-fidélité :
 - Amplificateur Vecteur I-4.2,
 - Lecteur Compact Disc Vecteur L-3.2,
 - Pour les enceintes actives : Adaptateur de niveau Behringer Ultralink Pro MX882 (première série d'expériences), Pré-amplificateur Rotel RC972 (deuxième série d'expériences) ou Pré-amplificateur Electrocompaniet EC 4.7(troisième expérience).

A.4 Photographies des séances d'enregistrements

Les Figures A.4, A.5, A.6 et A.7 présentent quelques photographies de nos différentes séances d'enregistrements. Des microphones de mesure sont indiqués sur les Figures A.4 et A.5, et apparaissent également sur un support en forme de croix sur les Figures A.6 et A.7. Ils ont été utilisés pour effectuer des mesures de réponses impulsionnelles des enceintes en différents points de la pièce. Ces mesures n'ont pas encore été exploitées et sont seulement évoquées dans le chapitre V.

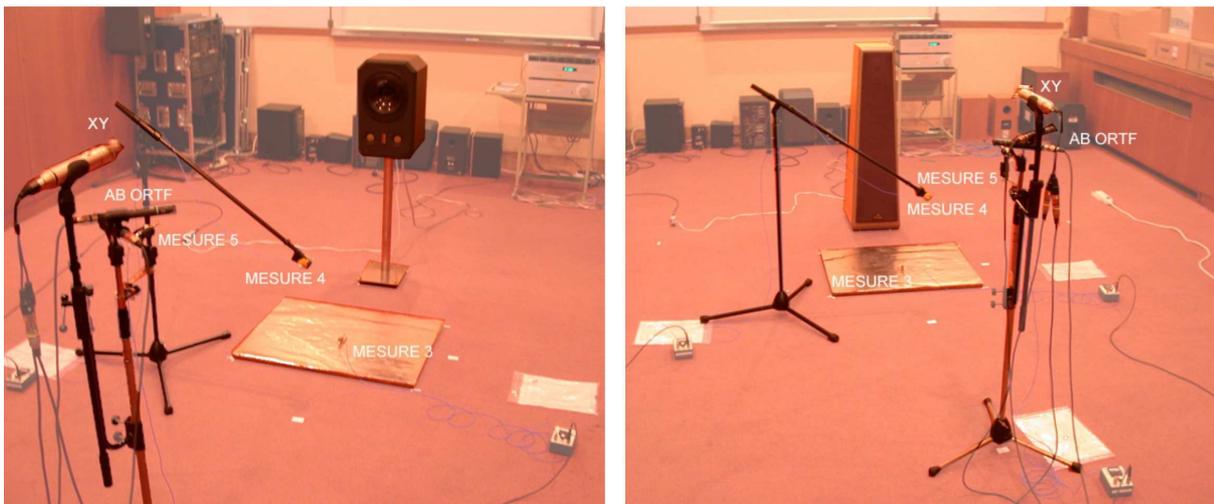


FIGURE A.4 – Première série d'expériences

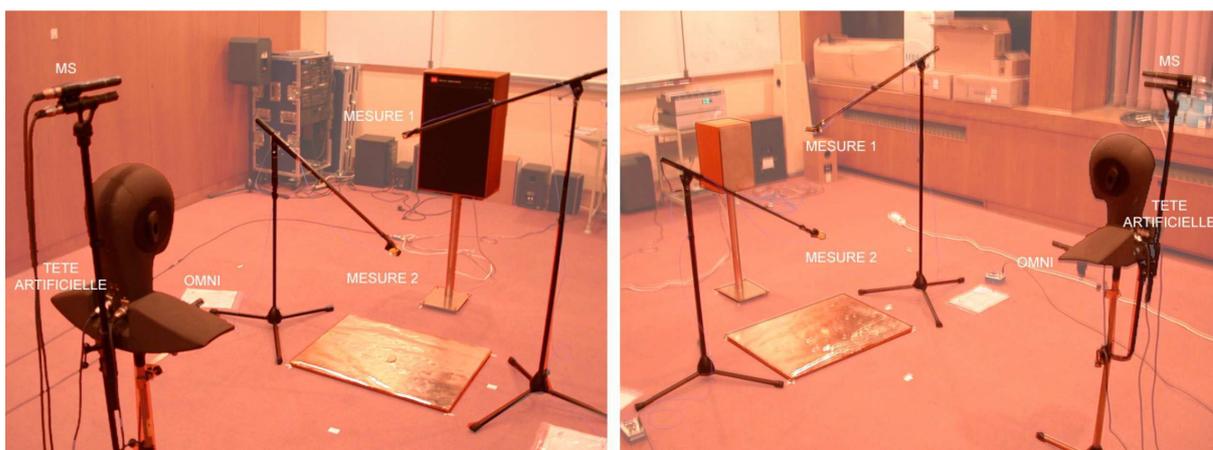


FIGURE A.5 – Première série d'expériences



FIGURE A.6 – Deuxième série d'expériences

A.5 Evaluation de la durée de réverbération des salles d'écoute

La durée de réverbération des salles d'écoute n'a été évaluée qu'à une seule position dans chaque pièce. Cette évaluation a donc été très sommaire, une évaluation complète aurait nécessité d'effectuer la mesure pour plusieurs positions de la source et du microphone.

Dans la première salle (Figure III.1, page 60), la mesure a été réalisée au niveau des microphones de prise de son. Dans la deuxième salle (Figure III.6, page 73), elle a été effectuée à 1m en face de l'enceinte, entre les microphones de prise de son et l'enceinte. Dans les deux cas, la source était une enceinte Genélec 1031A, située à la même position que les enceintes lors des enregistrements, et le microphone était omnidirectionnel. Le signal utilisé était un bruit rose brutalement interrompu. La durée de réverbération de la pièce a été déterminée en évaluant le durée nécessaire pour que le niveau sonore décroisse de 60 dB après extinction de la source.



FIGURE A.7 – Troisième expérience

Les Tableaux A.1 et A.2 contiennent les résultats de cette évaluation réalisée par bande de tiers d'octave entre 500 et 5000 Hz.

Fréquence centrale des tiers d'octave (en Hz)	500	630	800	1000	1250	1600	2000	2500	3150	4000	5000
Tr60 (en s.)	0,49	0,36	0,49	0,46	0,53	0,68	0,55	0,45	0,49	0,59	0,54

TABLEAU A.1 – Durée de réverbération dans la première salle d'écoute

Fréquence centrale des tiers d'octave (en Hz)	500	630	800	1000	1250	1600	2000	2500	3150	4000	5000
Tr60 (en s.)	0,39	0,40	0,31	0,38	0,43	0,38	0,35	0,41	0,41	0,49	0,44

TABLEAU A.2 – Durée de réverbération dans la deuxième salle d'écoute

A.6 Dispositif expérimental lors des tests psychoacoustiques

Les tests psychoacoustiques ont tous été réalisés dans une cabine audiométrique. Le casque utilisé était un Stax SR Lambda Professional. Les réponses en fréquence du casque et des dif-

férents microphones de prise de son n'ont pas été pas compensées. Par contre, les comparaisons entre les enregistrements ont toujours été réalisées à l'aide du même casque et pour une prise de son donnée.

Les tests de comparaison par paires ont été effectués à l'aide d'une chaîne de reproduction Tucker&Davis. Les auditeurs quantifiaient alors les similarités entre les enregistrements à l'aide d'un terminal VT320. Le test de classification libre faisait lui intervenir une interface Matlab pour présenter les enregistrements aux auditeurs. La carte son utilisée dans ce cas pour la restitution était la même que celle qui a servi lors des enregistrements (paragraphe A.3, page vi).

Les Figures A.8 et A.9 présentent la consigne fournie aux auditeurs pour chaque type de test.

TEST DE SIMILARITE (comparaison par paires)

Le but de ce test est de juger la similarité entre 12 versions d'un même extrait musical présentées par paires. Pour cela, vous disposez d'un curseur que vous pourrez déplacer à l'aide des deux flèches du clavier

(← →) :

- ◆ Plus le curseur se trouve vers le côté **gauche**, plus les sons vous paraissent **semblables**.
- ◆ Plus le curseur se trouve vers le côté **droit**, plus les sons vous paraissent **dissemblables**.
- ◆ Le curseur autour du **centre** signifie que les sons vous paraissent **moyennement semblables**.

Très]-----x-----[Très
semblables dissemblables

Vous avez la possibilité de réécouter la paire en saisissant le numéro 0 du clavier. Ensuite, il vous suffira de saisir la touche return pour valider votre choix et entendre la paire suivante.

Le test débute par une première écoute des 12 extraits que vous aurez à comparer par la suite.

Merci beaucoup de votre participation

FIGURE A.8 – Consigne fournie aux auditeurs lors des tests de comparaison par paires

La consigne de la Figure A.9 évoque la présence potentielle de "bruits parasites (craquements, bruit de fond)" dans certains enregistrements. Les enregistrements impliquant les trente-sept enceintes de ce test ont été réalisés en plusieurs séances étalées sur plusieurs mois. Le niveau de reproduction n'ayant pas été exactement le même durant ces différentes séances, nous craignons qu'après égalisation du niveau des enregistrements, des différences de niveau de bruit de fond puissent apparaître. Cependant, aucun des cinquante-six auditeurs ayant effectué le test n'a évoqué la présence de bruits parasites ou de bruit de fond, et l'analyse de leurs jugements de dissemblances ne laisse en rien supposé qu'ils aient pu être influencé par un tel facteur, même inconsciemment.

TEST DE SIMILARITE (classification libre)

Le but de ce test est de juger de la **similarité/ressemblance** entre 37 versions d'un même extrait musical. Ces 37 versions sont représentées par des croix réparties au hasard sur l'écran devant vous. Vous pouvez **écouter** une version en **cliquant gauche** avec la souris sur la croix correspondante et vous pouvez **déplacer** cette croix en **cliquant droite** sur la souris.

La tâche consiste à regrouper les versions que vous trouvez très semblables:

◆ **Deux versions (croix) jugées très semblables seront placées dans le même paquet de croix**

◆ **Deux versions jugées différentes seront placées dans des paquets différents**

Ce sont bien les **différences entre les versions et non pas vos préférences** qui nous intéressent. Il ne s'agit pas de classer les mauvais d'un côté et les bons de l'autre, mais bien de **grouper les versions qui se ressemblent**. Certaines versions peuvent contenir des bruits parasites (craquements, bruit de fond), essayez autant que possible de faire abstraction de ces parasites lorsque vous comparez les versions.

Vous pouvez utiliser le nombre de paquets que vous voulez et écouter/déplacer les versions autant de fois que vous le désirez. La façon dont vous placez les paquets les uns par rapport aux autres sur l'écran n'a aucune importance. Répondez ce qui vous semble naturel, il n'y a ni bonnes, ni mauvaises réponses.

Lorsque vous avez classé toutes les versions/croix, faites signe à l'expérimentateur pour qu'il valide le test.

Merci beaucoup de votre participation

FIGURE A.9 – Consigne fournie aux auditeurs lors du test de classification libre

ANNEXE B

Méthodes d'analyse statistique utilisées

————— oOo —————

Cette annexe décrit le principe des analyses statistiques utilisées lors de notre étude, en présentant notamment le type d'information qu'elles permettent d'extraire des données. Nous ne reviendrons pas sur les fondements théoriques de ces différentes méthodes d'analyse. Des références sont fournies pour le lecteur souhaitant une description plus exhaustive.

————— oOo —————

SOMMAIRE

B.1	Corrélation	xiii
B.2	Comparaison de partitions: indice de Rand	xiii
B.3	Analyse en arbre hiérarchique	xiv
B.4	Analyse de variance (ANOVA)	xv
B.5	Analyse multidimensionnelle (MDS)	xvii
B.5.1	Modèle utilisé	xvii
B.5.2	Détermination du nombre de dimensions	xviii
B.5.3	Orientation des axes	xx
B.5.4	Stabilité de l'espace perceptif	xx

B.1 Corrélation

Le coefficient de corrélation ρ de deux variables x et y permet de quantifier le degré de linéarité de la relation entre ces variables [102]. Il vaut 1 ou -1 si la relation est parfaitement linéaire, et 0 si les deux variables sont complètement indépendantes. Nous l'avons estimé à l'aide de la formule :

$$\rho\{x,y\} = \langle(x - \langle x \rangle)(y - \langle y \rangle)\rangle / \sqrt{\langle(x - \langle x \rangle)^2\rangle\langle(y - \langle y \rangle)^2\rangle}$$

B.2 Comparaison de partitions : indice de Rand

Lors de notre troisième expérience impliquant la tâche de classification libre, les différences de jugement entre auditeurs peuvent provenir de différences de classement des enregistrements, mais aussi du nombre de classes qu'ils ont utilisées. La dissemblance entre deux auditeurs ne doit pas être évaluée directement par la corrélation de leurs réponses. En effet, un auditeur A peut avoir utilisé neuf classes qui sont l'exacte subdivision des trois classes définies par un auditeur B. Les jugements des auditeurs A et B ne sont pas sans rapport, bien au contraire, les auditeurs ont simplement considéré un degré de similarité différent entre les enregistrements pour effectuer les regroupements. L'évaluation directe de la corrélation entre les matrices binaires résultant des partitions réalisées par les auditeurs (paragraphe II.4.2, page II.4.2) rendrait difficilement compte de ce type de correspondance. Pour évaluer la correspondance entre différentes partitions d'un même ensemble d'éléments, le coefficient de corrélation est remplacé par l'indice de Rand [103], cité par Houix [76].

Le détail du calcul de l'indice de Rand de deux partitions est rappelé par Houix [76]. Il est basé sur l'établissement d'un tableau de contingence recensant le nombre de stimuli qui sont communs aux différentes classes des deux partitions. La correspondance entre les deux partitions est évaluée à partir de la façon dont les paires de stimuli sont classées. Pour une paire de stimuli donnée, quatre types de classement sont possibles :

- 1- les stimuli sont placés dans la même classe pour les deux partitions,
- 2- les stimuli sont placés dans la même classe pour la première partition mais dans des classes différentes pour la deuxième,
- 3- les stimuli sont placés dans des classes différentes pour la première partition mais dans la même classe pour la deuxième,
- 4- les stimuli sont placés dans des classes différentes pour les deux partitions.

Les types 1 et 4 correspondent à un accord des deux partitions pour la classification de la paire de stimuli considérée, tandis que les types 2 et 3 indiquent un désaccord. L'indice de Rand est calculé à partir du nombre d'occurrence des classements des différents types entre les deux

partitions. Lorsque les partitions sont identiques, l'indice de Rand vaut 1. Il tend vers 0 lorsque les partitions sont complètement indépendantes.

B.3 Analyse en arbre hiérarchique

Nous avons utilisé l'analyse en arbre hiérarchique pour deux raisons lors de notre étude : afin de vérifier la présence ou non de classes aux stratégies de réponses différentes parmi les auditeurs, avant de considérer la matrice moyenne de leurs évaluations, et pour tester l'homogénéité de nos stimuli en mettant ou non en évidence la présence de catégories parmi les enregistrements, avant d'interpréter les espaces résultant de l'analyse multidimensionnelle. Dans les deux cas l'analyse en arbre hiérarchique a pour but d'étudier la structure de données de dissemblances, et de faire apparaître d'éventuelles classes parmi les éléments dont les dissemblances ont été évaluées.

Le programme que nous avons utilisé permet de tracer un dendrogramme [51][76][77] à partir d'une matrice de dissemblances entre éléments (auditeurs ou enregistrements, dans notre cas). La Figure B.1 reprend l'exemple de la Figure III.2 déjà présentée au chapitre III, afin d'illustrer le processus de construction d'un dendrogramme et la lecture qui doit être faite d'un tel arbre.

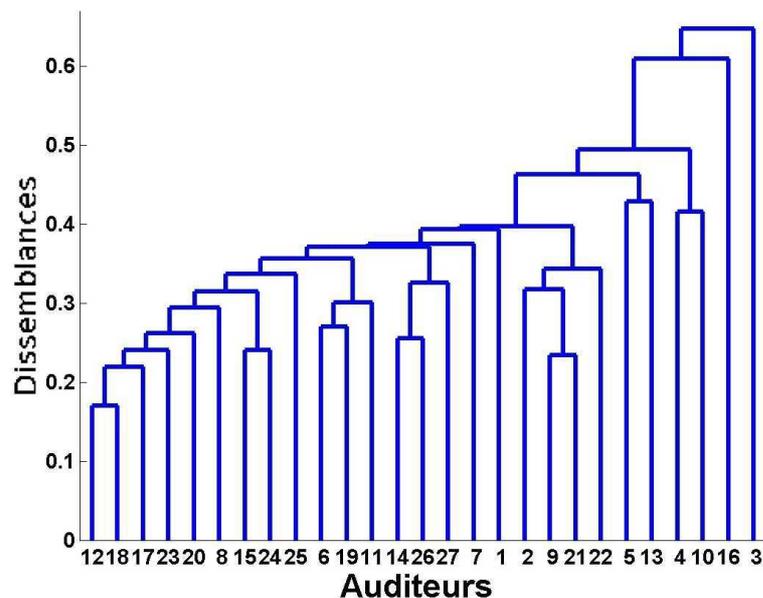


FIGURE B.1 – Exemple de dendrogramme résultant de l'analyse en arbre hiérarchique des dissemblances entre auditeurs (première série d'expériences, Kan'nida, couple AB ORTF)

La construction du dendrogramme est basée sur un processus itératif qui vise à regrouper les éléments les uns après les autres. Les deux éléments les plus similaires, 12 et 18 sur la Figure B.1, sont d'abord regroupés. Leurs "branches" forment un noeud dont une hauteur correspond à leur dissemblance. Ils sont remplacés dans la matrice de dissemblances par un seul élément. La

matrice de dissemblances est formée des vecteurs contenant les dissemblances de chaque élément avec tous les autres. Le vecteur de dissemblances associé au nouvel élément est calculé en prenant la moyenne arithmétique des vecteurs de dissemblances des deux éléments ayant été regroupés. Les deux éléments alors les plus similaires sont à leur tour regroupés, il s'agit de l'agglomérat de 12-18 et de 17 dans notre exemple. Le dendrogramme est ainsi construit de manière itérative, jusqu'à ce que tous les éléments de l'arbre soient reliés par une branche. La dissemblance entre deux éléments est donc représentée par la hauteur du noeud qui les relie. Deux éléments seront d'autant plus similaires que la hauteur de ce noeud sera faible. La distance entre les éléments sur la dimension horizontale n'a aucune signification.

Ce type d'analyse permet de rechercher la présence de classes parmi les éléments. Lors de nos analyses, nous avons considéré que des éléments formaient une classe à part quand ils étaient à la fois similaires entre eux et différents des autres éléments. Aucune classe n'apparaît sur la Figure B.1. Même si les évaluations des auditeurs 3 et 16 sont moins corrélées à celles des autres auditeurs, elles ne sont pas corrélées entre elles. Les auditeurs 3 et 16 ne forment donc pas une classe à part parmi les auditeurs.

B.4 Analyse de variance (ANOVA)

Le principe de l'analyse de variance est décrit en détail par Wonnacott et Wonnacott [102], et des exemples d'utilisation peuvent être trouvés dans [51][77]. Le programme que nous avons utilisé est tiré de [104].

Ce type d'analyse a pour but de quantifier l'influence des différents facteurs d'une expérience sur une variable dépendant de ces facteurs, à partir d'un nombre donné d'observations. L'analyse de variance permet de savoir si l'influence d'un facteur est significative, c'est-à-dire si les changements de valeur de la variable dépendante, observés lors de modifications apportées à ce facteur, sont liés à ces modifications ou au bruit de mesure. L'analyse de variance donne également accès au pourcentage de variance induite par chaque facteur au sein la variance totale des données de l'expérience.

Pour illustrer ces propos, nous pouvons considérer l'exemple de l'utilisation de cette analyse dans le cadre de notre deuxième série d'expériences. Nous souhaitons quantifier l'influence du mode de reproduction des enceintes sur l'évaluation des dissemblances perceptives entre nos enregistrements. Il s'agit alors d'étudier l'effet du facteur "mode de reproduction (R)" des enceintes sur la variable "dissemblance de la paire" d'enregistrements. Le deuxième facteur de l'expérience est la "paire d'enceintes (P)" considérée. L'analyse est effectuée sur un total de 20 observations, correspondant aux vingt auditeurs de l'expérience. Le bruit de mesure est lié à la variabilité des évaluations provenant des différents auditeurs. Le facteur "mode de reproduction (R)" possède 2 niveaux ou modalités, car deux modes de reproduction sont impliqués dans chaque test. Le

facteur "paire d'enceintes (P)" a 55 niveaux, correspondant au nombre de paires qui ont été évaluées, à partir des onze enceintes impliquées dans les enregistrements.

"Le principe de l'analyse de la variance est de comparer la *variance dans les groupes* (*within subject*) qui correspond à la variabilité des jugements à l'intérieur d'une moyenne pour une modalité, à la *variance entre les groupes* (*between subject*) qui correspond à la variabilité entre deux moyennes (groupes) que l'on cherche à comparer. Le résultat de cette comparaison prend la forme d'une probabilité p , qui est la probabilité que la différence observée soit uniquement due à l'erreur expérimentale (*hypothèse nulle*). On fixe alors un critère α en dessous duquel on décide de rejeter l'hypothèse nulle. Il y a traditionnellement deux critères utilisés : $\alpha = 0,05$ et $\alpha = 0,01$. Si l'on obtient une valeur p inférieure au critère que l'on s'est fixé, on dit que l'effet est *significatif*" Lemaitre [77]. La probabilité p est appelée probabilité critique [102], et sera notée p.c. dans ce document.

Les calculs permettant de déterminer cette probabilité critique sont détaillés par Wonnacott et Wonnacott [102]. Ils font intervenir la détermination de grandeurs intermédiaires pour chaque facteur :

- sa variation SC, pour Somme des Carrés, correspondant à la somme des écarts quadratiques des moyennes obtenues pour les différents niveaux du facteur à la moyenne globale des données,
- son nombre de degrés de liberté d.d.l. égal au nombre de ses niveaux moins un (des corrections de type Greehouse-Geisser et Huynh-Feldt pouvant être appliquées lorsque les niveaux du facteur ne sont pas indépendants les uns des autres [16][51]),
- sa variance CM, pour Carré Moyen, égale au rapport de la Somme des Carrés sur le nombre de degrés de liberté,
- le rapport de Fisher expliqué F, correspondant au rapport de la variance expliquée par ce facteur sur la variance inexpliquée.

L'analyse de variance donne également accès au pourcentage de variance induite par chaque facteur au sein de la variance totale des données de l'expérience, due à la fois aux effets des facteurs expérimentaux et au bruit de mesure. Il est généralement noté R^2 . L'ensemble de ces grandeurs est souvent présenté sous forme d'un tableau, appelé tableau ANOVA, similaire à ceux que nous avons présentés dans l'annexe C.3.

L'influence des facteurs expérimentaux est ainsi quantifiée, et c'est également le cas de leurs interactions potentielles. "Une interaction entre deux variables signifie que les effets de l'une d'elles sont influencés par le niveau de l'autre. Appliquée aux évaluations dans un essai d'écoute, une interaction entre les haut-parleurs et les extraits de programme peut signifier que les écarts d'évaluation entre deux (ou plusieurs) haut-parleurs sont différents pour des programmes différents" [30]. Considérons le cas de notre étude portant sur l'influence du mode de reproduction des

enceintes sur l'évaluation des dissemblances perceptives entre les enregistrements. Si le facteur paire d'enceintes (P) a un effet significatif, cela signifie qu'en moyenne sur tous les auditeurs il existe au moins deux paires d'enceintes qui ont des valeurs de dissemblances significativement différentes. Les enceintes utilisées sont alors suffisamment différentes pour que les paires ne soient pas toutes jugées en moyenne aussi dissemblables les unes que les autres. Si le facteur mode de reproduction (R) a une influence significative, alors les moyennes globales des dissemblances pour chaque mode de reproduction, calculées sur tous les auditeurs et toutes les paires d'enceintes, sont significativement différentes. L'interaction paires d'enceintes-mode de reproduction (PxR) est significative lorsque, en moyenne sur tous les auditeurs, il existe au moins deux paires d'enceintes dont les différences de dissemblance ne sont pas identiques pour les deux modes de reproduction.

B.5 Analyse multidimensionnelle (MDS)

L'analyse multidimensionnelle de proximité vise à représenter des dissemblances estimées entre stimuli par des distances entre ces mêmes stimuli symbolisés par des points dans un espace multidimensionnel : ces points seront d'autant plus éloignés au sein de l'espace que les stimuli correspondants auront été estimés plus dissemblables. Cet espace permet d'explorer le nombre et la nature des dimensions sous-jacentes aux estimations de dissemblances [39]. Nous avons utilisé cette analyse sur les dissemblances entre nos enregistrements : sur les évaluations perceptives effectuées par les auditeurs lors des tests psychoacoustiques, et sur les évaluations objectives obtenues à partir des différentes analyses appliquées aux signaux.

B.5.1 Modèle utilisé

Plusieurs modèles d'analyse multidimensionnelle existent, ils sont notamment décrits dans [51][76][77]. Nous avons employé le modèle classique MDSCAL (MultiDimensionnal SCALing). L'algorithme utilisé, SMACOF pour Scaling by Majorizing a Complicated Function, est décrit en détail par Borg et Groenen [39]. C'est un modèle métrique, par opposition aux modèles non-métriques qui conservent uniquement l'information d'ordre contenue dans la matrice de dissemblances. Un modèle non-métrique ordonnera les paires en fonction de leur dissemblances, et ne considérera par la suite que ces informations d'ordre, tandis qu'un modèle métrique conservera les valeurs de dissemblances associées à chaque paire pour la suite de l'analyse. Les modèles non-métriques sont préférables si les propriétés arithmétiques des valeurs de dissemblances assignées aux jugements des auditeurs ne s'appliquent pas forcément à l'échelle utilisée par ces auditeurs pour effectuer leurs jugements [39]. En pratique, Borg et Groenen [39] montrent que les deux types d'analyse fournissent les mêmes espaces lorsque suffisamment de stimuli sont impliqués et que des comparaisons par paires sont utilisées. Gabrielsson *et coll.* [19] trouvent effectivement

des résultats similaires avec les deux modèles, sur des dissemblances faisant intervenir seulement cinq stimuli.

Comme la majorité des analyses multidimensionnelles, l'espace considéré dans nos analyses pour représenter les stimuli est un espace euclidien, aux dimensions orthogonales, indépendantes et bien sûr continues. Comme les dendrogrammes évoqués précédemment dans ce chapitre, l'analyse multidimensionnelle n'est qu'une représentation des données de dissemblances entre les stimuli. Rien ne prouve que ces dimensions correspondent à des dimensions perceptives ou objectives qui permettent de passer de façon continue d'un stimuli à l'autre. C'est en interprétant l'espace que l'expérimentateur fait des hypothèses sur ses données. L'hypothèse de l'existence de dimensions continues, orthogonales et indépendantes n'est en rien justifiée par l'analyse multidimensionnelle elle-même. L'analyse multidimensionnelle permet d'explorer la nature des dissemblances, mais ce sera à l'expérimentateur de justifier ses interprétations.

L'algorithme employé a pour but de faire correspondre les dissemblances entre stimuli à des distances entre des points dans un espace euclidien de dimension N , N étant fixé par l'expérimentateur. Les distances sont obtenues en minimisant une fonction coût, appelée contrainte ou plus communément stress [39][51][76], qui évalue l'erreur quadratique entre dissemblances et distances.

B.5.2 Détermination du nombre de dimensions

C'est à l'expérimentateur de choisir le nombre de dimensions de l'espace dans lequel il veut représenter ces données. Plusieurs critères s'offrent à lui pour faire ce choix. Il peut dans un premier temps comparer les valeurs minimales de stress obtenues pour différents nombre de dimensions. Le but n'est pas de réduire le stress à zéro. Les données de dissemblances sont bruitées, et il s'agit de décrire les véritables dissemblances, pas le bruit de mesure. Le stress sera d'autant plus élevé que le bruit de mesure sera important [39]. Plutôt que de considérer les valeurs du stress, il est préférable de s'intéresser à l'évolution du stress en fonction du nombre de dimensions de l'espace. Le stress diminue forcément avec le nombre de dimensions ajoutées à l'espace. Les distances décriront d'autant mieux les dissemblances qu'elles auront pour le faire plus de dimensions, c'est-à-dire de degrés de liberté. Mais à partir d'un certain nombre de dimensions, la diminution du stress devient très faible lors de l'ajout d'une nouvelle dimension. Cette dimension pourra alors être associée au bruit de mesure. Borg et Groenen [39] conseillent de tracer l'évolution du stress en fonction du nombre de dimensions, afin de repérer un "coude" sur cette courbe. Ce coude indique que l'ajout de dimensions supplémentaires semble superflu. À titre d'exemple, la Figure B.2 présente les courbes de stress que nous avons obtenu lors de notre première série d'expériences. Les coudes apparaissant entre deux et trois dimensions, la solution à deux dimensions a été retenue.

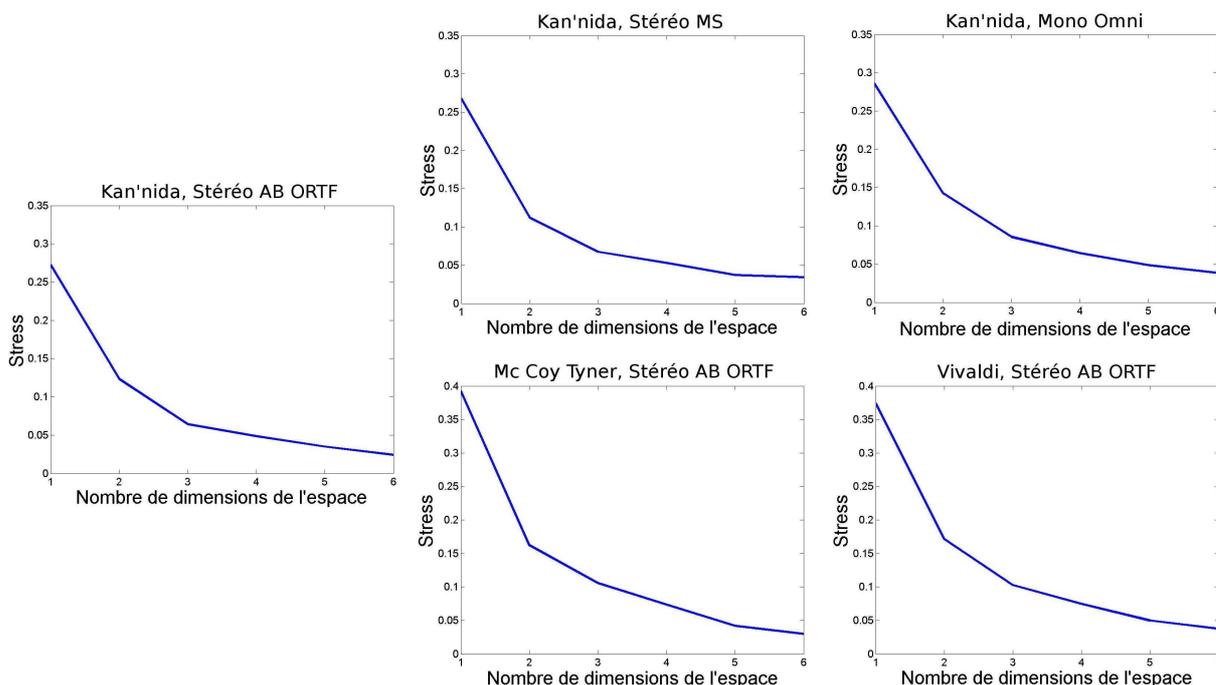


FIGURE B.2 – Détermination du nombre de dimensions lors de notre première série d'expériences : évolution du stress en fonction du nombre de dimensions de l'espace.

Il faut également garder à l'esprit que le nombre de dimensions pouvant être mises en évidence par ce type d'analyse est limité par le nombre de stimuli impliqués dans les évaluations de dissemblances. Kruskal et Wish [50] cités par Marozeau [51] proposent une règle empirique selon laquelle N stimuli ne peuvent conduire qu'à un maximum de $(N-1)/4$ dimensions pour que l'espace obtenu puisse être considéré comme robuste. Cette règle est également évoquée dans [32]. La représentation multidimensionnelle obtenue sera en effet d'autant meilleure que le nombre de stimuli impliqué sera plus élevé. Le bruit de mesure se répartissant aléatoirement sur les estimations de dissemblances, en sous-estimations et en sur-estimations, plus il y aura d'estimations, moins la représentation multidimensionnelle a de chances de décrire ce bruit [39]. Par contre, plus il y a de stimuli, plus le nombre de points sur lequel est calculé le stress est important, et plus la valeur de stress obtenue est grande. Un stress important ne veut donc pas dire que la représentation est mauvaise [39].

Le critère d'interprétabilité des dimensions est aussi à considérer pour déterminer le nombre de dimensions pertinent pour décrire les données de dissemblances [39][51]. Dans notre cas impliquant des sons comme stimuli, l'interprétation des dimensions devra se baser sur l'écoute, les commentaires libres des auditeurs, et les critères objectifs qui pourront être définis pour décrire les dimensions obtenues.

Le programme d'analyse multidimensionnelle que nous avons utilisé ne considère que des

dimensions communes aux différents stimuli, et ne permet pas de prendre en compte une dimension propre à un seul stimulus. Un stimulus ayant une caractéristique particulière non partagée par les autres stimuli pourra générer des valeurs importantes de dissemblance du fait de cette particularité. Le programme d'analyse multidimensionnelle pourra alors conduire à des erreurs en forçant ces dissemblances à être représentées à l'aide des dimensions communes. Par exemple, si des rectangles sont comparés les uns aux autres, et que ces rectangles varient uniquement par leur longueur et leur largeur, sauf pour un rectangle qui est plein alors que tous les autres sont vides, alors le rectangle plein pourrait être jugé très différent des autres. Malgré tout, il se peut que le caractère plein ou vide n'apparaisse pas comme une dimension, car les autres rectangles ne varient pas selon ce critère. Le rectangle plein pourrait alors être positionné dans l'espace bidimensionnel à une position aberrante du point de vue de sa longueur et de sa largeur, du fait de sa particularité qui n'a pas pu être prise en compte. Pour éviter ce problème, certains programmes plus élaborés permettent de tenir compte de cette présence potentielle de spécificités des stimuli [51][77][105].

B.5.3 Orientation des axes

L'algorithme MDSCAL ne fixe pas l'orientation des axes de l'espace multidimensionnel. Cet espace peut être tourné et dilaté, seule la distance relative entre les points est fixée. Par défaut, l'orientation des axes coïncide avec les directions de variance maximale, déterminée à l'aide d'une analyse en composantes principales. L'ordre des dimensions est notamment déterminé par la quantité de variance expliquée par ces dimensions.

Lors de notre étude, c'est la comparaison des espaces résultant de différentes expériences qui nous intéressait. Pour pouvoir les comparer, nous les avons tous orientés et dilatés pour qu'ils coïncident au mieux avec un même espace de référence. "Cette rotation ne modifie pas la solution MDS, elle permet simplement de l'examiner sous un angle propice" Marozeau [51]. Pour les espaces ayant arbitrairement servi de référence, l'orientation par défaut a été conservée.

Lors de la recherche d'attributs objectifs décrivant les dimensions, l'orientation choisie a son importance. Nous avons conservé l'orientation par défaut des espaces de référence, mais pas celles des espaces ayant été orientés lors de la comparaison des expériences. Une recherche plus approfondie d'attributs objectifs nécessiterait de considérer plusieurs orientations, afin de définir les meilleurs attributs possibles.

B.5.4 Stabilité de l'espace perceptif vis-à-vis des évaluations de dissemblance : analyse en bootstrap

Il convient d'être prudent lorsque les positions des stimuli dans les espaces perceptifs sont considérées. En effet, celles-ci sont déterminées avec une résolution plus ou moins élevée. L'éva-

luation psychoacoustique est bruitée, du fait de la variabilité inter- et intra-individuelle des jugements de dissemblances. Plus les auditeurs seront d'accord sur leurs jugements, plus les positions résultant de ces jugements seront définies précisément dans l'espace perceptif. Ainsi, plus un attribut perceptif sera saillant, plus les auditeurs l'identifieront facilement, et plus les positions des stimuli selon la dimension correspondante dans l'espace perceptif seront définies avec précision.

Une manière de se rendre compte de la résolution de nos évaluations psychoacoustiques consiste à tester la stabilité de notre espace perceptif vis-à-vis des différences de jugement de nos auditeurs. Ceci peut être fait par une analyse multidimensionnelle en bootstrap [76][102]. Le bootstrap consiste à réaliser plusieurs analyses multidimensionnelles de suite. Mais, plutôt que de considérer tous les auditeurs à chaque fois, ces derniers sont choisis "avec remise". C'est à dire que s'il y a N auditeurs au total, N auditeurs sont sélectionnés, mais ceux-ci ne sont pas forcément tous différents, un même auditeur peut être sélectionné plusieurs fois. Lors du tirage des auditeurs, nous choisissons aléatoirement parmi tous les auditeurs et pas seulement parmi ceux qui n'ont pas encore été choisis. En fonction des tirages, l'influence de certains auditeurs est plus ou moins grande et l'espace obtenu sera plus ou moins modifié. En représentant le résultat de ces analyses multidimensionnelles successives sur un même espace, à la place d'un point, nous obtenons un nuage de points par stimulus. Plus la taille du nuage est importante, moins la résolution de l'évaluation est bonne.

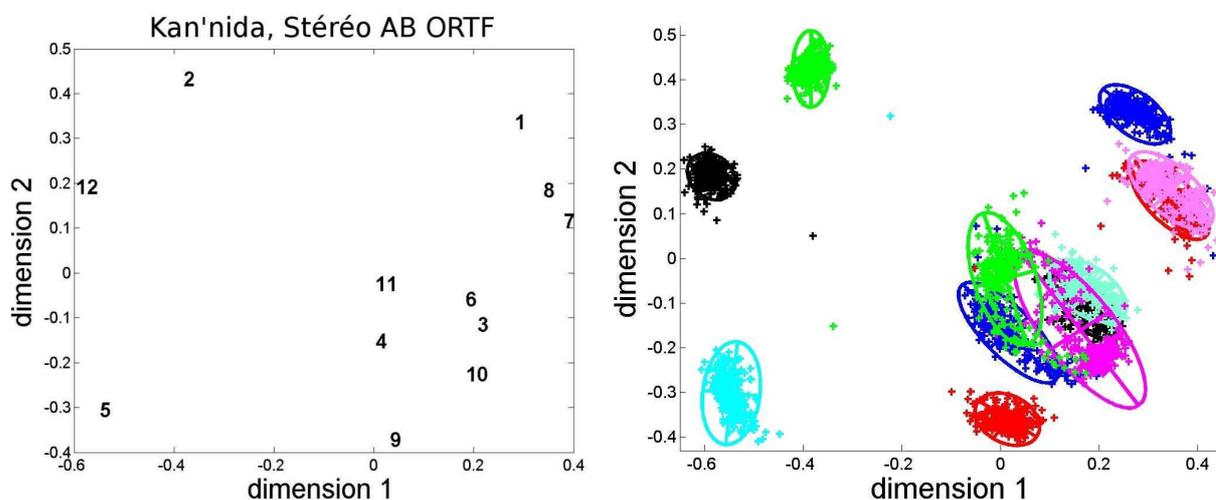


FIGURE B.3 – *Analyse bootstrap : exemple de l'espace perceptif obtenu à partir du test psychoacoustique impliquant l'extrait musical Kan'nida et la prise de son AB ORTF lors de notre première série d'expériences*

A titre d'exemple, la Figure B.3 présente le résultat de l'analyse en bootstrap du test utilisant l'extrait Kan'nida et la prise AB ORTF lors de notre première série d'expériences. L'espace

perceptif obtenu à partir de la matrice de dissemblances moyenne impliquant tous les auditeurs est présentée à nouveau pour comparaison. L'analyse en bootstrap comportait 250 tirages. Les ellipses entourant les nuages de points contiennent 95% de la variance de ces nuages. Nous constatons bien que les stimuli sont davantage caractérisés par des "zones" de l'espace perceptif que par de véritables points. Les nuages étirés selon la dimensions 2 indiquent que cette deuxième dimension est moins bien définie que la première. Comme toute mesure, les évaluations psychoacoustiques ont une résolution forcément limitée. Cette résolution sera d'autant moins bonne que les jugements des auditeurs seront décorrélés et que l'attribut perceptif sera subtil.

ANNEXE C

Données complémentaires

----- oOo -----

Cette annexe contient un certain nombre de données expérimentales venant compléter les résultats évoqués dans le corps principal du document.

----- oOo -----

SOMMAIRE

C.1	Analyses en arbre hiérarchique sur les auditeurs	xxv
C.2	Corrélations entre dissemblances objectives et perceptives	xxviii
C.3	Analyses de variance de la deuxième série d'expériences	xxx
C.4	Analyses en arbre hiérarchique sur les enceintes	xxxii
C.5	Espaces objectifs obtenus à partir des techniques d'analyse "classiques" .	xxxv
C.6	Compléments sur les attributs objectifs préliminaires	xxxviii

C.1 Analyses en arbre hiérarchique sur les auditeurs

Pour chacun de nos test psychoacoustiques, avant de considérer la moyenne des dissemblances sur l'ensemble des auditeurs, nous avons vérifié qu'il n'y avait pas de groupe d'auditeurs ayant des stratégies de réponse différentes. Nous avons pour cela utilisé une analyse en arbre hiérarchique (annexe B.3) sur les matrices de dissemblances entre auditeurs (paragraphe II.4.1 et II.4.2, pages 40 et 41).

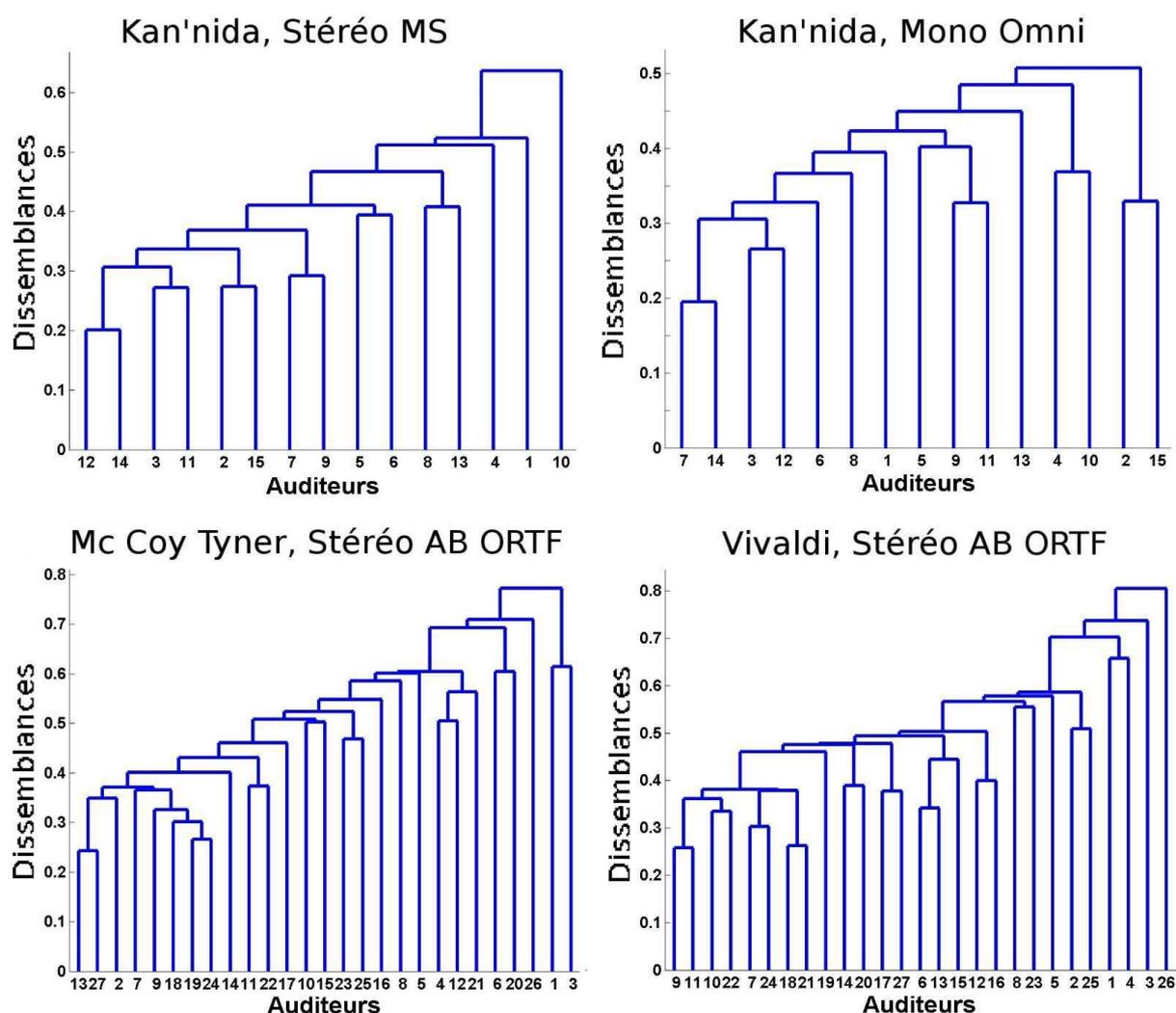


FIGURE C.1 – Dendrogrammes résultant de l'analyse en arbre hiérarchique des dissemblances entre auditeurs pour les quatre autres tests de notre première série d'expériences, en complément de la Figure III.2, page 64.

Les Figures C.1, C.2 et C.3 présentent les dendrogrammes résultant de cette analyse, en complément du dendrogramme déjà présenté à titre d'exemple dans le corps principal du document

(Figure III.2, page 64). Aucune classe d'auditeurs n'apparaît clairement sur ces dendrogrammes. Sur la Figure C.1, les évaluations de l'auditeur 10 sont moins corrélées à celles des autres auditeurs, pour le test impliquant l'extrait musical Kan'nida et la prise de son MS. Pour le test utilisant l'extrait Mc Coy Tyner, ce sont les réponses des auditeurs 1 et 3 qui seraient plus atypiques. Nous constatons qu'au cours de nos expériences les quelques auditeurs "atypiques" n'étaient pas forcément les mêmes d'un test à l'autre. Même si les réponses de certains auditeurs sont parfois moins en accord avec celles des autres pour un test donné, tous les résultats ont été conservés pour le reste de nos analyses. Nous avons préféré avoir recours à un nombre plus important d'auditeurs pour réduire le bruit de mesure, plutôt que d'éliminer les réponses jugées atypiques.

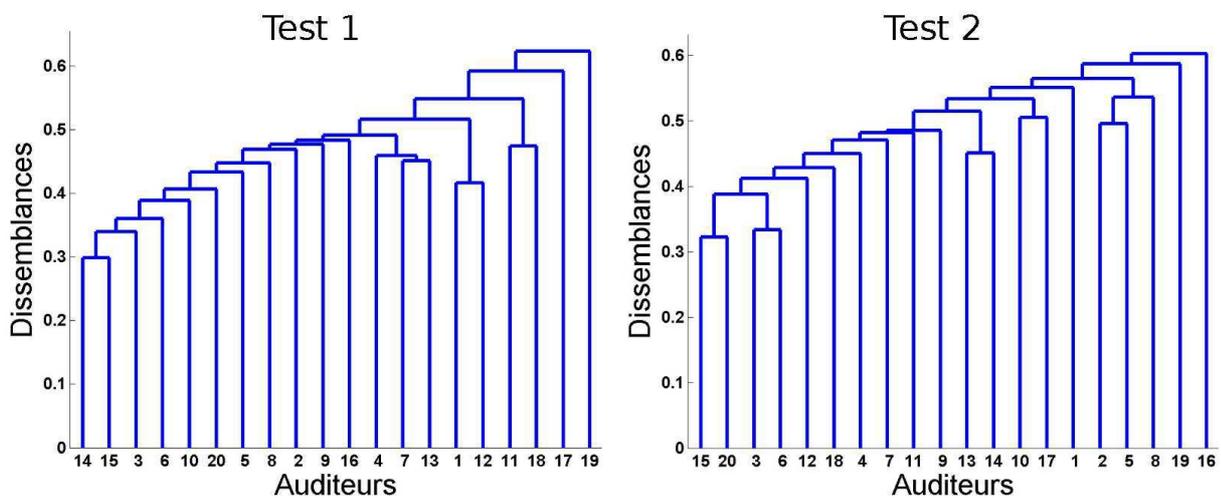


FIGURE C.2 – Dendrogrammes résultant de l'analyse en arbre hiérarchique des dissemblances entre auditeurs pour notre deuxième série d'expériences. Les reproductions monophonique et double monophonique ont été évaluées lors du test 1. Le test 2 impliquait les reproductions double monophonique et stéréophonique.

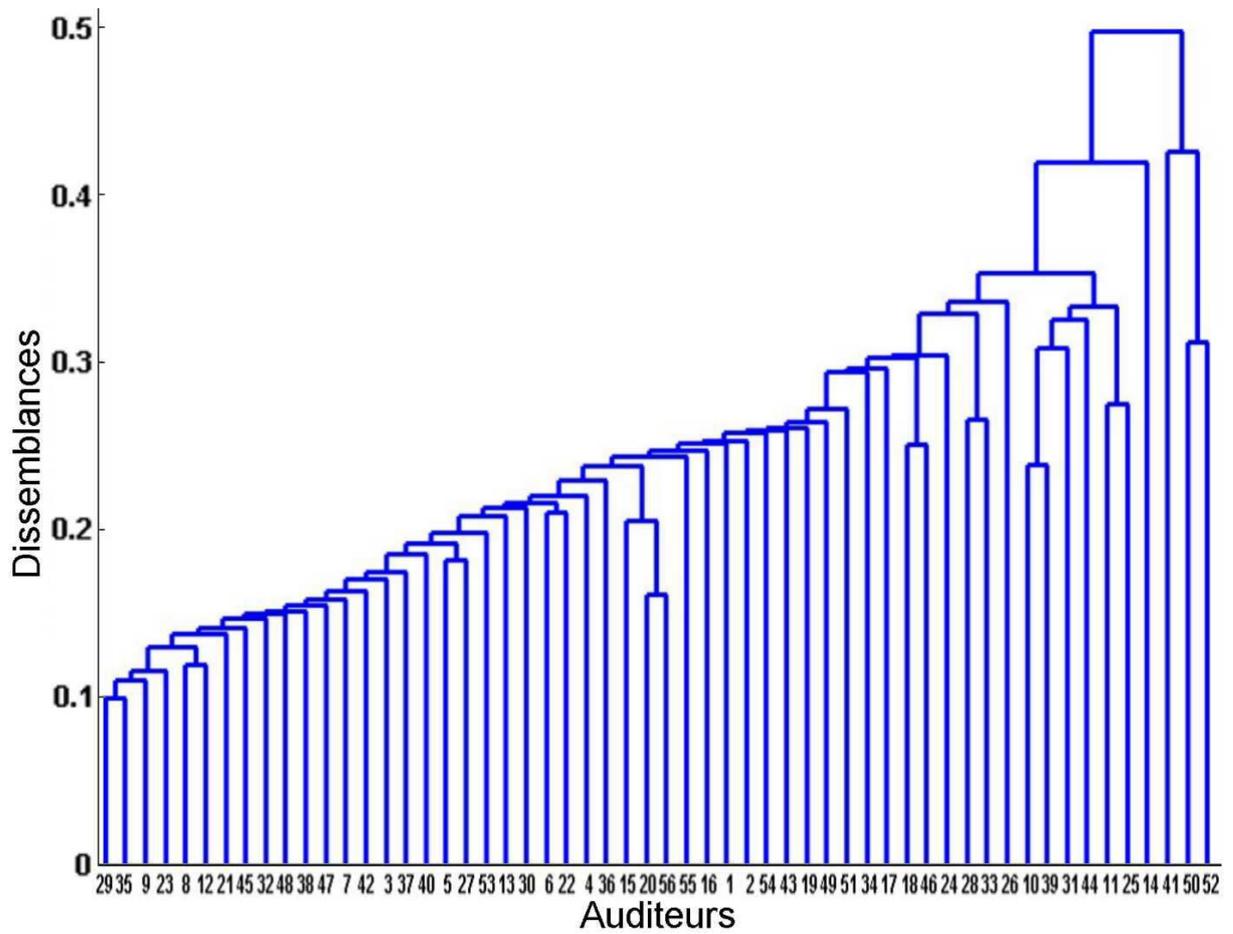


FIGURE C.3 – Dendrogramme résultant de l'analyse en arbre hiérarchique des dissemblances entre auditeurs pour notre troisième expérience.

C.2 Corrélations entre dissemblances objectives et perceptives

Les Tableaux C.1, C.2 et C.3 présentent les corrélations entre les dissemblances perceptives provenant de chacune de nos trois séries d'expériences et les dissemblances objectives calculées sur chacune des voies des enregistrements correspondants.

	Kan AB g	Kan AB d	Kan MS g	Kan MS d	Kan omni	Mc AB g	Mc AB d	Viv AB g	Viv AB d
Domaine temporel	0,35	0,30	0,55	0,62	0,34	<i>0,01</i>	<i>0,04</i>	0,20	0,41
Spectre	0,35	0,30	0,55	0,62	0,34	<i>0,01</i>	<i>0,04</i>	0,20	0,41
Temps-fréquence	0,35	0,30	0,55	0,62	0,34	<i>0,01</i>	<i>0,04</i>	0,20	0,41
Module du spectre	0,44	0,42	0,68	0,51	0,46	0,44	0,57	0,56	0,47
Module du temps-fréquence	0,43	0,42	0,68	0,54	0,45	0,44	0,54	0,57	0,48
Densité spectrale de puissance	0,45	0,44	0,68	0,50	0,47	0,44	0,55	0,57	0,45
Pondération A	0,78	0,71	0,35	0,54	0,76	0,44	0,49	0,61	0,45
Pondération Iso70	0,80	0,75	0,34	0,57	0,77	0,45	0,47	0,62	0,49
Densité de sonie totale	0,75	0,82	0,84	0,80	0,79	0,88	0,80	0,74	0,76
Densité de sonie 1	0,89	0,85	0,85	0,85	0,79	0,85	0,78	0,80	0,78
Densité de sonie 1 moyenne	0,90	0,86	0,83	0,82	0,78	0,88	0,84	0,80	0,79
Densité de sonie 2	0,85	0,84	0,84	0,83	0,82	0,79	0,72	0,79	0,76
Densité de sonie 2 moyenne	0,87	0,86	0,80	0,83	0,81	0,87	0,82	0,81	0,75

TABLEAU C.1 – *Corrélations entre matrices de dissemblances objectives et perceptives pour notre première série d'expériences, impliquant les extraits musicaux Kan'nida (Kan), Mc Coy Tyner (Mc) et Vivaldi (Viv), avec les prises de son AB ORTF (AB), MS et monophonique omnidirectionnelle (omni). Les voies gauche (g) et droite (d) de nos signaux sont analysées séparément en utilisant les différentes techniques d'analyse objective définies au Chapitre 2.*

Lors de notre première série d'expériences, nous n'avons pas su expliquer les valeurs de corrélation particulièrement faibles obtenues avec l'extrait Mc Coy Tyner, à partir du domaine temporel, du spectre et du temps-fréquence, lorsque les informations de phase sont prises en compte. Ces valeurs sont indiquées en italique dans le Tableau C.1. Ce phénomène n'a pas été observé pour les autres signaux de cette expérience, ni pour les autres expériences utilisant

l'extrait Mc Coy Tyner.

	M g	M d	DM g test 1	DM d test 1	DM g test 2	DM d test 2	S g	S d
Domaine temporel	0,46	0,45	0,69	0,64	0,70	0,67	0,69	0,63
Spectre	0,46	0,45	0,69	0,64	0,70	0,67	0,69	0,63
Temps-fréquence	0,46	0,45	0,69	0,64	0,70	0,67	0,69	0,63
Module du spectre	0,60	0,56	0,78	0,71	0,79	0,72	0,79	0,69
Module du temps-fréquence	0,60	0,56	0,79	0,71	0,80	0,72	0,79	0,68
Densité spectrale de puissance	0,57	0,55	0,79	0,70	0,79	0,72	0,79	0,68
Pondération A	0,51	0,46	0,62	0,45	0,64	0,46	0,59	0,40
Pondération Iso70	0,51	0,39	0,70	0,44	0,70	0,46	0,64	0,38
Densité de sonie totale	0,85	0,84	0,87	0,77	0,88	0,78	0,91	0,82
Densité de sonie 1	0,87	0,82	0,87	0,79	0,89	0,80	0,92	0,82
Densité de sonie 1 moyenne	0,87	0,84	0,89	0,87	0,91	0,87	0,93	0,89
Densité de sonie 2	0,85	0,81	0,85	0,78	0,88	0,79	0,90	0,81
Densité de sonie 2 moyenne	0,85	0,84	0,88	0,84	0,89	0,84	0,93	0,88

TABLEAU C.2 – Corrélations entre matrices de dissemblances objectives et perceptives pour notre deuxième série d'expériences, impliquant les modes de reproduction monophonique (*M*), double monophonique (*DM*) et stéréophonique (*S*) pour les paires d'enceintes. Les voies gauche (*g*) et droite (*d*) de nos enregistrements sont analysées séparément en utilisant les différentes techniques d'analyse objective définies au Chapitre 2. Les enregistrements impliquant le mode de reproduction double monophonique ont été utilisés dans les tests 1 et 2 de cette deuxième série d'expériences (paragraphe III.2, page 66).

Lors de notre troisième expérience (Tableau C.3), en plus de l'analyse complète de l'ensemble des dissemblances, nous avons réalisé une analyse partielle en ne conservant que les dissemblances impliquant les onze enceintes qui avaient déjà été utilisées lors de notre première série d'expériences. Notre but était de pouvoir comparer deux expériences faisant intervenir les mêmes enceintes acoustiques utilisées dans deux pièces différentes (paragraphe III.3.3, page 74).

	voie g analyse partielle	voie d analyse partielle	voie g analyse complète	voie d analyse complète
Domaine temporel	0,47	0,48	0,31	0,31
Spectre	0,47	0,48	0,31	0,31
Temps-fréquence	0,47	0,48	0,31	0,31
Module du spectre	0,56	0,61	0,46	0,39
Module du temps-fréquence	0,56	0,61	0,46	0,39
Densité spectrale de puissance	0,53	0,59	0,46	0,40
Pondération A	0,28	0,39	0,32	0,28
Pondération Iso70	0,31	0,44	0,26	0,21
Densité de sonie totale	0,76	0,79	0,63	0,64
Densité de sonie 1	0,82	0,80	0,68	0,68
Densité de sonie 1 moyenne	0,78	0,81	0,68	0,70
Densité de sonie 2	0,79	0,78	0,69	0,67
Densité de sonie 2 moyenne	0,79	0,82	0,70	0,71

TABLEAU C.3 – *Corrélations entre matrices de dissemblances objectives et perceptives pour notre troisième expérience. Les voies gauche (g) et droite (d) de nos signaux sont analysées séparément en utilisant les différentes techniques d'analyse objective définies au Chapitre 2. Une analyse partielle des données a été réalisé en plus de l'analyse complète (paragraphe III.3.3, page 74).*

C.3 Analyses de variance de la deuxième série d'expériences

Dans le cadre de notre deuxième série d'expériences, grâce à une analyse de variance (ANOVA), nous avons pu quantifier l'influence du mode de reproduction des enceintes sur l'évaluation des dissemblances perceptives entre nos enregistrements. Les aspects théoriques de cette analyse sont présentés dans l'annexe B.4. Nous détaillons ici les résultats obtenus dans le cadre de notre expérience. L'impact de l'utilisation de deux enceintes plutôt qu'une seule a été évalué par l'analyse des résultats du test 1 faisant intervenir les enregistrements des reproductions monophonique et double monophonique. L'influence de l'utilisation d'un signal stéréophonique plutôt que mono-

phonique a été quantifiée en appliquant la même analyse sur les résultats du test 2 impliquant les enregistrements des reproductions double monophonique et stéréophonique.

Notre but était d'étudier l'effet du facteur "mode de reproduction (R)" des enceintes sur la variable dépendante "dissemblance de la paire" d'enregistrements. Le deuxième facteur de l'expérience était bien sûr la "paire d'enceintes (P)" considérée. L'analyse de variance permet de révéler les facteurs qui ont une influence significative sur les changements de la variable dépendante. Une probabilité critique p est associée à l'effet de chaque facteur. Elle indique la probabilité pour que l'effet de ce facteur soit en fait dû au hasard. Le seuil de probabilité pour considérer qu'un effet est significatif est fixé à 1% ($p < 0,01$). Le pourcentage de la variance induite par chaque effet, par rapport à la variance totale de la variable dépendante, est également calculé. Il est noté R^2 .

L'analyse est effectuée sur un total de 20 observations, correspondant aux vingt auditeurs de l'expérience. Le facteur "mode de reproduction (R)" possède 2 niveaux, car deux modes de reproduction sont impliqués dans chaque test. Le facteur "paire d'enceintes (P)" a 55 niveaux, correspondant au nombre de paires qui ont été évaluées, à partir des onze enceintes impliquées dans les enregistrements. Dans le cas de la reproduction monophonique, pour les modèles d'enceinte dont nous avons utilisé les enceintes droite et gauche pendant les enregistrements, seules les dissemblances impliquant les enceintes droites ont été conservées lors de l'analyse. L'enceinte droite a été arbitrairement choisie comme référence du modèle en reproduction monophonique, les évaluations impliquant les reproductions monophonique et double monophonique devant porter sur le même nombre de paires pour pouvoir être comparées lors de l'analyse de variance.

Le Tableau C.4 contient les résultats de l'analyse menée sur les dissemblances recueillies lors du test 1. Le mode de reproduction (R) et l'interaction paires d'enceintes-mode de reproduction (PxR) ont une influence statistiquement significative sur les résultats du test psychoacoustique. Cette influence reste malgré tout très faible puisque les deux effets cumulés représentent moins de 4% de la variance des évaluations de dissemblances. Le facteur paires d'enceintes (P) est bien sûr lui aussi significatif : les auditeurs ont perçu des différences entre les paires d'enregistrements pendant les tests d'écoute, et ces différences n'étaient pas les mêmes pour toutes les paires présentées. La variabilité des résultats due aux différences entre auditeur explique également une part de la variance.

Les résultats de l'analyse de variance du test 2 sont présentés dans le Tableau C.5. Le facteur mode de reproduction (R) a encore une influence significative, mais cette influence est très faible. Elle explique moins de 1% de la variance des résultats du test psychoacoustique. L'effet de l'interaction paires d'enceintes-mode de reproduction (PxR) n'est plus significatif.

Pour chacune des deux analyses, des corrections de type Greehouse-Geisser et Huynh-Feldt ont été testées pour tenir compte du fait que nos paires d'enceintes ne sont pas indépendantes les unes des autres [16][51]. Ces corrections n'ont quasiment pas modifié les résultats présentés

	SC	d.d.l.	CM	F	p	R^2 (%)
Auditeur (A)	17,0000	19	0,8947			8,8697
Reproduction (R)	0,6868	1	0,6868	12,9761	0,0019	0,3583
R x A	1,0056	19	0,0529			0,5247
Paire d'enceintes (P)	77,2928	54	1,4313	29,7541	0,0001	40,3272
P x A	49,3567	1026	0,0481			25,7517
P x R	6,4571	54	0,1196	3,0775	0,0001	3,3690
P x R x A	39,8651	1026	0,0389			20,7995
Total	191,6641	2199				100,0000

TABLEAU C.4 – Deuxième série d'expériences : analyse ANOVA du Test 1 impliquant les reproductions monophonique et double monophonique. SC : Somme des Carrés, d.d.l. : degré de liberté (= nombre de niveaux - 1), CM : Carré Moyen, F : rapport de Fisher expliqué, p : probabilité critique, et R^2 : pourcentage de variance induit.

	SC	d.d.l.	CM	F	p	R^2 (%)
Auditeur (A)	15,4657	19	0,8140			8,3555
Reproduction (R)	0,9946	1	0,9946	27,2434	0,0001	0,5374
R x A	0,6937	19	0,0365			0,3748
Paire d'enceintes (P)	78,1697	54	1,4476	30,1539	0,0001	42,2318
P x A	49,2547	1026	0,0480			26,6103
P x R	2,1860	54	0,0405	1,0835	0,3192	1,1810
P x R x A	38,3322	1026	0,0374			20,7093
Total	185,0966	2199				100,0000

TABLEAU C.5 – Deuxième série d'expériences : analyse ANOVA du Test 2 impliquant les reproductions double monophonique et stéréophonique. SC : Somme des Carrés, d.d.l. : degré de liberté (= nombre de niveaux - 1), CM : Carré Moyen, F : rapport de Fisher expliqué, p : probabilité critique, et R^2 : pourcentage de variance induit.

dans les Tableaux C.4 et C.5, et les conclusions restent identiques.

C.4 Analyses en arbre hiérarchique sur les enceintes

Pour chacun de nos tests psychoacoustiques, les dissemblances perceptives entre les enregistrements ont été soumises à une analyse en arbre hiérarchique avant l'analyse multidimensionnelle. Ceci nous a permis de vérifier l'homogénéité de nos stimuli en mettant en évidence la

présence ou non de catégories parmi les enregistrements. Les Figures C.4, C.5 et C.6 présentent les dendrogrammes résultant de cette analyse pour nos différentes expériences, en complément des dendrogrammes déjà présentés dans le corps principal du document (Figures IV.1 et IV.9, pages 81 et 94). Aucun groupement net entre les enregistrements n'est observé, les dimensions continues révélées par l'analyse multidimensionnelle sont donc jugées pertinentes pour décrire la perception des auditeurs lors des tests psychoacoustiques.

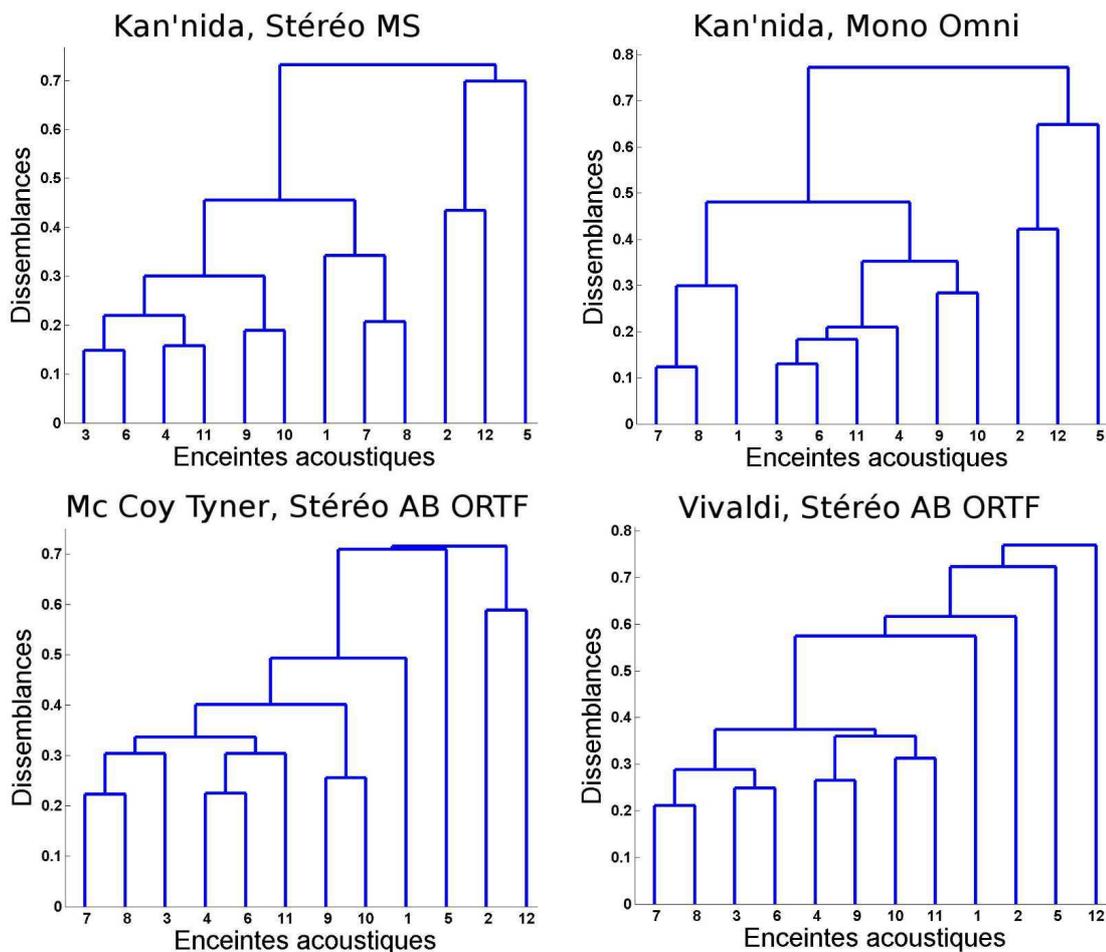


FIGURE C.4 – Dendrogrammes résultant de l'analyse en arbre hiérarchique des dissemblances entre enregistrements pour les quatre autres tests de notre première série d'expériences, en complément de la Figure IV.1, page 81. Les enceintes 7 et 8 sont du même modèle.

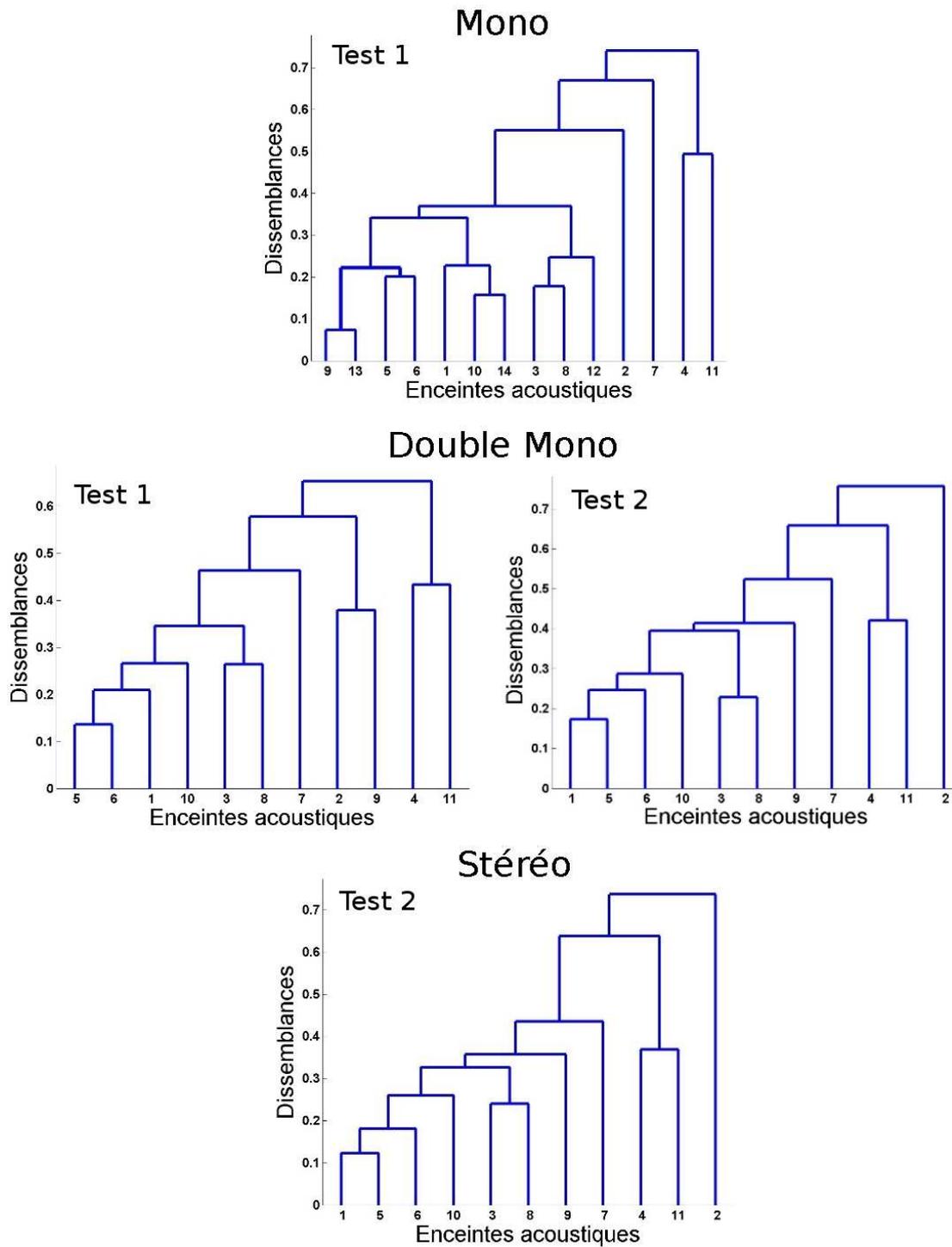


FIGURE C.5 – Dendrogrammes résultant de l'analyse en arbre hiérarchique des dissemblances entre enregistrements pour notre deuxième série d'expériences. Les enceintes 12, 13 et 14 sont respectivement du même modèle que les enceintes 3, 9 et 10.

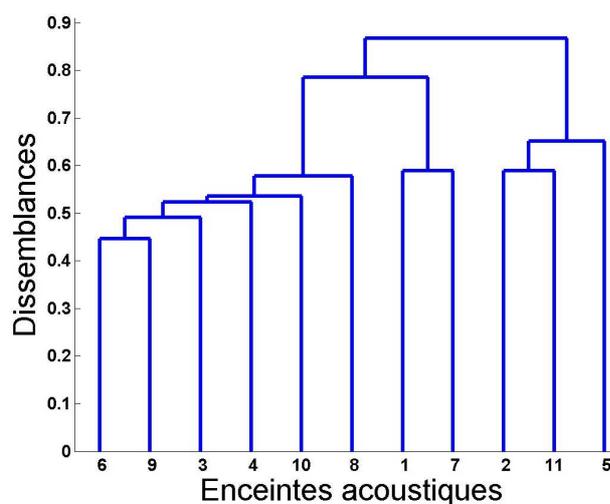


FIGURE C.6 – Dendrogramme résultant de l'analyse en arbre hiérarchique des dissemblances entre enregistrements lors de l'analyse partielle de notre troisième expérience, réalisée en conservant seulement les dissemblances impliquant les onze enceintes qui avaient déjà été utilisées lors de notre première série d'expériences.

C.5 Espaces objectifs obtenus à partir des techniques d'analyse "classiques"

Nous présentons ici des espaces objectifs obtenus à partir des techniques d'analyse classiques, lors de notre première série d'expériences. Ils sont fournis à titre d'exemple pour illustrer les tendances observées sur l'ensemble des enregistrements de nos expériences. Ces exemples sont à comparer aux espaces perceptifs correspondants (Figure IV.2, page 82).

Les domaines temporel, fréquentiel et temps-fréquence, en considérant les informations de phase, ont conduit à des résultats équivalents. Nous avons donc seulement considéré l'espace objectif obtenu à partir des dissemblances calculées dans le domaine temporel. Deux tendances ont été observées sur les résultats. Deux exemples d'espace illustrant ces tendances sont présentés sur la Figure C.7. Soit l'espace est plutôt unidimensionnel mais coïncide difficilement avec la première dimension de l'espace perceptif (Kan'nida, AB ORTF, voie gauche), avec notamment une position aberrante de l'enceinte 1. Soit l'espace est bidimensionnel mais très différent de l'espace perceptif (Kan'nida, MS, voie droite). Nous constatons également une très grande disparité entre les différents espaces objectifs bidimensionnels obtenus. Aucune des dimensions objectives mises en évidence ne peut être reliée à une dimension perceptive.

La densité spectrale de puissance fournissant des dissemblances objectives très semblables à celles obtenues à partir des modules du spectre ou du temps-fréquence, seuls les espaces ob-

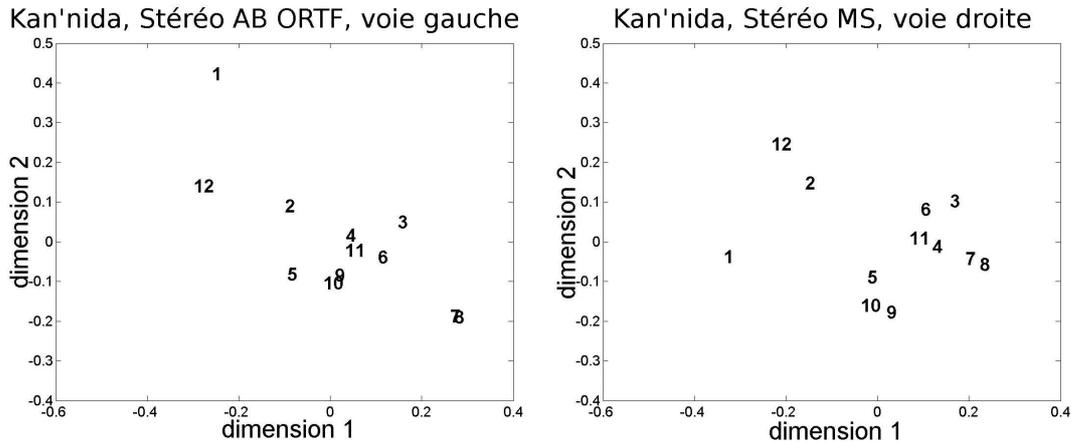


FIGURE C.7 – Exemples d’espaces objectifs déterminés à partir de la forme temporelle de nos signaux. Ils sont à comparer aux espaces perceptifs de la Figure IV.2, page 82.

jectifs obtenus à partir de la densité spectrale de puissance ont été déterminés. De la même façon que dans le domaine temporel, les résultats fournissent deux tendances, avec des espaces unidimensionnels ou bidimensionnels. La Figure C.8 fournit un exemple d’espace unidimensionnel (Kan’nida, AB ORTF, voie gauche), et un exemple d’espace bidimensionnel (Kan’nida, MS, voie gauche). L’enceinte 1 dans ces espaces retrouve une position plus en accord avec les résultats des tests perceptifs, par rapport aux espaces de la Figure C.7. Les espaces objectifs unidimensionnels permettent de se rapprocher un peu de la première dimension de l’espace perceptif (Figure IV.2, page 82). L’espace bidimensionnel présenté commence à ressembler à l’espace perceptif correspondant, mais c’est le cas le plus favorable observé, et d’autres espaces bidimensionnels obtenus avec cette technique d’analyse donnent des résultats beaucoup plus médiocres. Lors de la comparaison directe des dissemblances objectives et perceptives du chapitre III, nous avons pu constater qu’éliminer les informations de phase permettait d’améliorer la corrélation objectif/perceptif. Nous retrouvons ici que cette élimination autorise une description un peu plus fidèle de la première dimension perceptive, mais les résultats restent très hétérogènes.

A l’instar de ce qui a été observé lors du calcul des corrélations entre dissemblances objectives et perceptives, les deux pondérations fréquentielles testés donnent des résultats équivalents, et ces résultats sont très dépendants des signaux considérés. Les espaces objectifs obtenus sont tous bidimensionnels. Deux exemples d’espace reposant sur l’utilisation de la pondération basée sur la courbe isosonique à 70 phons sont présentés sur la Figure C.9. La pondération peut apporter une réelle amélioration pour retrouver l’espace perceptif. C’est le cas pour les espaces objectifs correspondant à l’extrait musical Kan’nida et aux prises de son AB ORTF et monophonique omnidirectionnelle, l’exemple obtenu à partir de la voie gauche de la prise AB ORTF étant présenté sur la Figure C.9. Sur ces espaces, les deux dimensions perceptives apparaissent déjà.

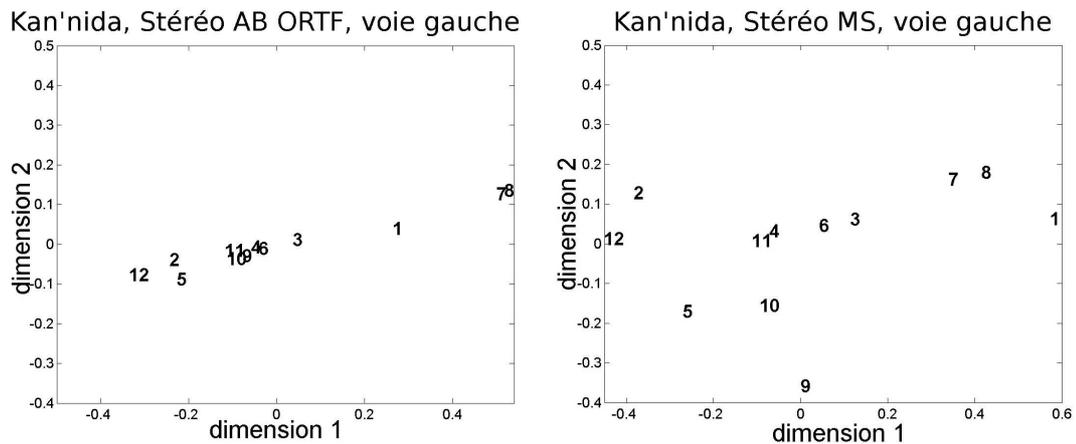


FIGURE C.8 – Exemples d'espaces objectifs déterminés à partir de la densité spectrale de puissance de nos signaux. Ils sont à comparer aux espaces perceptifs de la Figure IV.2, page 82.

Mais la pondération peut également n'avoir aucun effet ou même détériorer les résultats obtenus sans pondération. C'est par exemple le cas pour les enregistrements basés sur la prise de son MS. Cette forte dépendance vis-à-vis des signaux analysés rend l'utilisation de ces pondérations peu pertinente.

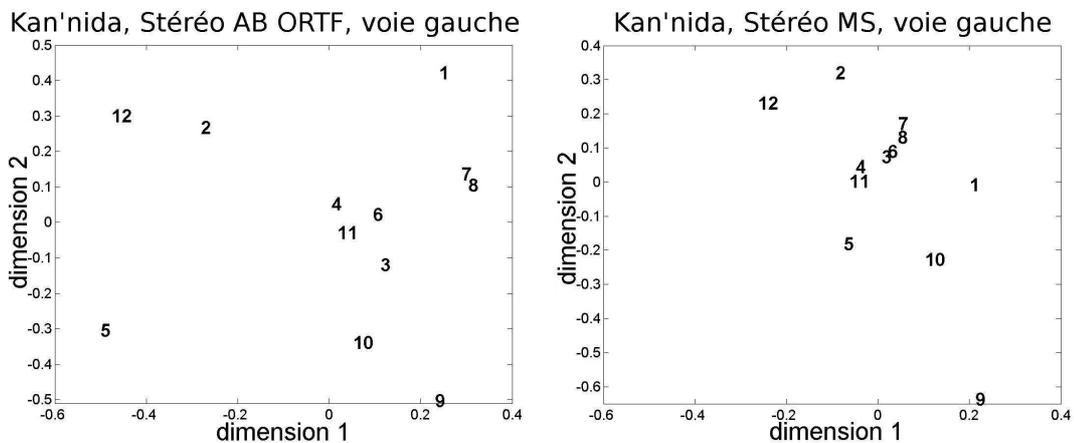


FIGURE C.9 – Exemples d'espaces objectifs déterminés à partir de la densité spectrale de puissance de nos signaux, pondérée selon la courbe isosonique à 70 phones. Ils sont à comparer aux espaces perceptifs de la Figure IV.2, page 82.

C.6 Compléments sur les attributs objectifs préliminaires

Le Tableau C.6 contient les valeurs du coefficient de corrélation entre la première dimension perceptive de notre première série d'expériences et l'attribut objectif "balance grave-aigu" tel que nous l'avons défini (paragraphe IV.6.1, page 100).

Signal	Corrélation balance grave-aigu/dimension 1
Kan AB gauche	0,96
Kan AB droite	0,95
Kan MS gauche	0,94
Kan MS droite	0,94
Kan omni	0,97
Mc AB gauche	0,92
Mc AB droite	0,83
Viv AB gauche	0,93
Viv AB droite	0,78
Moyenne	0,91

TABLEAU C.6 – Corrélation entre la balance grave-aigu et la dimension 1 de l'espace objectif (densité de sonie 2 moyenne), pour notre première série d'expériences, impliquant les extraits musicaux *Kan'nida* (*Kan*), *Mc Coy Tyner* (*Mc*) et *Vivaldi* (*Viv*), avec les prises de son AB ORTF (AB), MS et monophonique omnidirectionnelle (*omni*). Les voies gauche et droite de nos signaux sont considérées séparément.

Les valeurs du coefficient de corrélation entre la deuxième dimension perceptive de notre première série d'expériences et l'attribut objectif "émergence du bas-médium 1" (paragraphe IV.6.2, page 103) se trouvent dans le Tableau C.7.

Dans le cadre de notre troisième expérience, nous avons pu définir trois attributs objectifs pour décrire la "balance médium-aigu" de nos enregistrements à partir de leur densité de sonie 2 moyenne. Pour le premier attribut nous avons considéré le niveau dans les bandes de Bark 13 à 21, c'est-à-dire sur la gamme de fréquence 2240-15500 Hz, et celui dans les bandes 7 à 11, c'est-à-dire sur la gamme de fréquence 560-1800 Hz. Le niveau est calculé en intégrant la densité de sonie moyenne sur chacune des deux gamme de tonie, et nous prenons ensuite le rapport des deux niveaux obtenus pour évaluer l'équilibre médium-aigu. Les deux autres attributs sont définis de la même façon, mais les gammes de fréquence considérées sont différentes. La deuxième attribut utilise les bandes 13 à 21 pour le niveau des aigus et les bandes 7 à 12, c'est-à-dire sur la gamme de fréquence 560-2240 Hz, pour le niveau du médium. Le troisième attribut est calculé à partir du niveau dans les bandes 12 à 21, c'est-à-dire sur la gamme de fréquence 1800-15500

Signal	Corrélation émergence du bas-médium 1/dimension 2
Kan AB gauche	0,65
Kan AB droite	0,87
Kan MS gauche	0,88
Kan MS droite	0,92
Kan omni	0,89
Mc AB gauche	0,88
Mc AB droite	0,96
Viv AB gauche	0,93
Viv AB droite	0,95
Moyenne	0,88

TABLEAU C.7 – Corrélation entre l'émergence du bas-médium 1 et la dimension 2 de l'espace objectif (densité de sonie 2 moyenne), pour notre première série d'expériences, impliquant les extraits musicaux Kan'nida (Kan), Mc Coy Tyner (Mc) et Vivaldi (Viv), avec les prises de son AB ORTF (AB), MS et monophonique omnidirectionnelle (omni). Les voies gauche et droite de nos signaux sont considérées séparément.

Hz, et celui dans les bandes 4 à 11, c'est-à-dire sur la gamme de fréquence 355-1800 Hz. Ces trois attributs décrivent tous aussi bien la seconde dimension de l'espace objectif déterminé à partir de la moyenne droite-gauche de la densité de sonie 2 moyenne de nos enregistrements (Figure C.10). Dans les trois cas, le coefficient de corrélation entre ces trois évaluations de la balance médium-aigu et la dimension objective vaut 0,97.

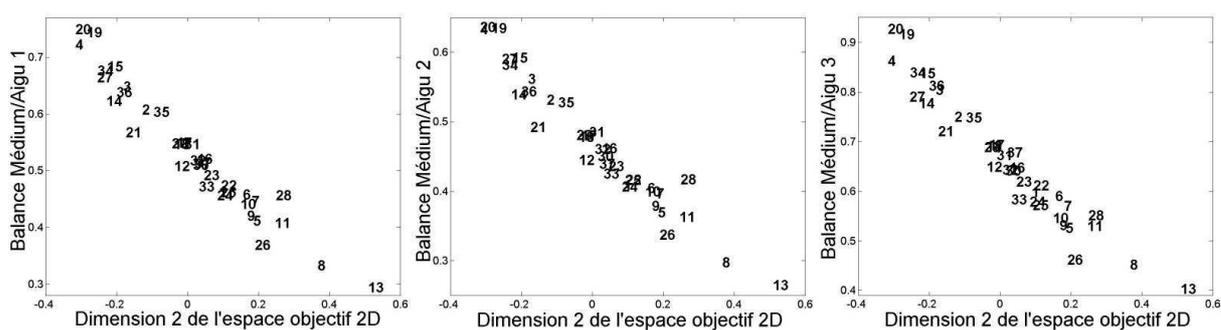


FIGURE C.10 – Troisième expérience : évaluation de la monotonie des relations entre les trois attributs objectifs évaluant la balance médium-aigu de nos enregistrements et la deuxième dimension de l'espace objectif (moyenne droite-gauche de la densité de sonie 2 moyenne). Chaque numéro correspond à une enceinte acoustique.

ANNEXE D

Synchronisation de signaux audio

————— oOo —————

Trois méthodes permettant d'évaluer le décalage temporel entre nos signaux ont été testées. Notre but était de synchroniser les enregistrements devant être comparés ensuite par les analyses objectives. Cette détermination est sensible à la gamme de fréquence considérée, elle semble donc dépendante des enceintes impliquées dans les enregistrements à synchroniser. La réverbération et les premières réflexions dans la pièce rendent la tâche encore plus délicate.

Aucune des méthodes testées n'a donné de résultats parfaits, nous avons donc choisi celle qui nous paraissait la plus adaptée à notre problème. Pour les trois méthodes, les décalages temporels ont été évalué sur les transitoires, en considérant en fait la dérivée temporelle de nos signaux.

————— oOo —————

SOMMAIRE

D.1	Méthode basée sur le cepstre complexe	xliii
D.2	Méthode basée sur l'intercorrélation	xliii
D.3	Minimisation de la distance quadratique dans le domaine temporel	xliv
D.4	Solution choisie	xliv

D.1 Méthode basée sur le cepstre complexe

Une première méthode pour déterminer le décalage temporel de deux signaux consiste à évaluer la part de retard pur contenue dans la phase de leur fonction de transfert [106]. La fonction de transfert des deux signaux est alors considérée comme le produit d'un filtre à phase minimum, traduisant les différences d'amplitude de cette fonction de transfert, et d'un retard pur. La composante à phase minimum est extraite en considérant le cepstre complexe [107] de la fonction de transfert. Le retard entre les signaux est ensuite évalué par régression linéaire sur l'excès de phase. La gamme fréquentielle considérée pour effectuer cette régression est restreinte à la gamme contenant le signal utile. Elle est déterminée en évaluant la cohérence des deux signaux [106].

Pour nos signaux, même lorsque ce critère de cohérence est assuré, le retard obtenu dépend beaucoup de la gamme fréquentielle sur laquelle il est calculé. Nous avons donc rapidement abandonné cette méthode de détermination.

D.2 Méthode basée sur l'intercorrélation

La fonction d'intercorrélation de deux signaux peut également donner accès à leur décalage temporel [106]. Cette fonction appliquée à deux signaux identiques simplement décalés dans le temps sera maximale à l'instant correspondant à leur décalage.

L'évaluation de la fonction d'intercorrélation de nos signaux ne nous a pas permis de déterminer parfaitement leurs décalages temporels. Nous avons rencontrés des incohérences du type : le décalage entre les signaux A et C n'est pas égal au décalage entre A et B plus le décalage entre B et C. Nos enregistrements ne sont pas des signaux identiques simplement décalés dans le temps. Ils ont tous été filtrés par des enceintes différentes, puis par la pièce, avant d'être captés par les microphones. L'échantillonnage de la fonction d'intercorrélation de signaux numériques de durée finie limite également la détermination précise du maximum de cette fonction.

Afin d'améliorer la synchronisation des enregistrements lors de nos deuxième et troisième séries d'expériences, nous avons ajouté un sinus glissant avant chaque extrait musical. Ce signal de synchronisation avait pour but d'éviter les incohérences rencontrées précédemment, en déterminant les décalages temporels sur un signal optimisé pour cela. La gamme fréquentielle du sinus glissant choisi, restreinte entre 300 et 9000 Hz, devait être compatible avec la bande passante de toutes les enceintes. Malheureusement, les décalages obtenus sur ces signaux ne sont toujours pas parfaits.

D.3 Minimisation de la distance quadratique dans le domaine temporel

La troisième méthode que nous avons testée consiste à minimiser la distance entre nos signaux dans le domaine temporel. Cette distance est définie comme la dissemblance objective que nous avons utilisée dans le domaine temporel. Elle est calculée en intégrant sur la variable temporelle le carré de la différence entre les deux signaux. Nous avons en fait conservé la valeur minimum des résultats obtenus à partir du carré de la différence et du carré de la somme des signaux, dans l'éventualité de deux signaux en opposition de phase.

$$Disstemp\{x(t),y(t)\} = Min\{\langle [x(t) - y(t)]^2 \rangle_t, \langle [x(t) + y(t)]^2 \rangle_t\}$$

Les décalages ainsi obtenus conduisent encore à des incohérences du type précédent, de l'ordre de quelques échantillons.

D.4 Solution choisie

Aucune des méthodes testées ne fonctionnant parfaitement, nous n'avons pas réussi à synchroniser parfaitement tous nos signaux les uns par rapport aux autres. Nous n'avons donc plus cherché à réaliser cette synchronisation globale sur l'ensemble des signaux. Les enregistrements ont été considérés par paires. Le décalage entre les deux signaux de la paire a été calculé en minimisant la distance quadratique entre leurs deux formes temporelles. La synchronisation a ainsi été effectuée au sein de chaque paire, indépendamment des synchronisations faites pour les autres paires. Les dissemblances objectives sont ensuite évaluées au sein de chaque paire, sur les signaux dont le décalage temporel a été minimisé.

La fonction d'intercorrélation nous a permis d'effectuer une première synchronisation sur les sinus glissant enregistrés avant chaque extrait musical. Elle a ensuite été affinée au sein de chaque paire par minimisation de la distance temporelle.

ANNEXE E

Méthodes d'analyse objective utilisées

————— oOo —————

Cette annexe présente les paramètres que nous avons employés lors des différentes analyses objectives appliquées à nos signaux, après que leur décalage temporel ait été minimisé.

————— oOo —————

SOMMAIRE

E.1	Déterminations du spectre et des pondérations fréquentielles	xlvi
E.2	Détermination de la représentation temps-fréquence	xlvi
E.3	Détermination des densités de sonie	xlvi

E.1 Déterminations du spectre et des pondérations fréquentielles

En plus du spectre de nos signaux évalué par l'intermédiaire de leur transformée de Fourier discrète, nous considérons leur densité spectrale de puissance. Celle-ci est estimée par la méthode du périodogramme moyenné modifié de Welch [108] cité par [106]. Une fenêtre de Hanning est utilisée, sans recouvrement entre les portions du signal sur lesquelles a lieu l'estimation. Le seul paramètre que nous avons fait varier est la taille de ces portions de signal, qui conditionne la résolution fréquentielle de la densité spectrale résultante. Nous voulions déterminer l'influence de cette résolution fréquentielle sur le calcul des dissemblances objectives entre les densités spectrales de puissance de nos enregistrements (paragraphe II.5.2, page II.5.2). Nous avons testé les estimations utilisant des portions de signal de 2048, 8192 et 32768 échantillons. Le Tableau E.1 présente les coefficients de corrélation entre les dissemblances objectives obtenues à partir de ces trois évaluations, menées sur les signaux de notre première série d'expériences.

Portions de signal utilisées (en nombre d'échantillons)	8192 / 2048	8192 / 32768	2048 / 32768
Kan AB g	0,99	0,99	0,99
Kan AB d	0,99	0,99	0,99
Kan MS g	0,99	0,98	0,99
Kan MS d	0,99	0,99	0,98
Kan omni	0,99	0,99	0,99
Mc AB g	0,99	0,99	0,99
Mc AB d	0,99	0,98	0,99
Viv AB g	0,99	0,99	0,99
Viv AB d	0,99	0,99	0,99
Moyenne	0,99	0,99	0,99

TABLEAU E.1 – *Corrélations entre les dissemblances objectives obtenues à partir de trois évaluations de densité spectrale de puissance, utilisant des portions de signal de taille différente. Les signaux considérés sont ceux de notre première série d'expériences, impliquant les extraits musicaux Kan'nida (Kan), Mc Coy Tyner (Mc) et Vivaldi (Viv), et les prises de son AB ORTF (AB), MS et monophonique omnidirectionnelle (omni). Les voies gauche (g) et droite (d) des signaux stéréophoniques sont analysées séparément.*

Pour les trois cas testés, les dissemblances objectives évaluées à partir de la densité spectrale de puissance des signaux ne sont pas dépendantes de la résolution fréquentielle utilisée. Nous avons donc choisi de présenter dans ce document uniquement les résultats impliquant la densité spectrale de puissance estimée sur 8192 échantillons. C'est également sur cette densité

spectrale que nous appliquons ensuite les pondérations fréquentielles devant rendre compte de la sensibilité de l'oreille. Ces pondérations sont basées sur les lignes isosoniques [82] à 40 phones pour la pondération "A", et à 70 phones pour la deuxième pondération testée, 70 phones correspondant approximativement au niveau de reproduction des enregistrements lors de nos tests psychoacoustiques.

E.2 Détermination de la représentation temps-fréquence

Une analyse temps-fréquence de nos signaux est réalisée en considérant leur transformée de Fourier à court terme (TFCT), également appelée transformée de Fourier glissante. Elle est calculée à partir de la transformée de Fourier discrète (TFD) du signal fenêtré, en faisant glisser cette fenêtre selon la variable temporelle, les portions successives ainsi analysées pouvant se recouvrir partiellement [109]. Cette transformée estime le spectre à court terme, localisé dans le temps, du signal. Nous avons utilisé une fenêtre de Hanning de 1024 échantillons, avec un taux de recouvrement de 50% entre les portions de signal analysées successivement.

E.3 Détermination des densités de sonie

Les modèles auditifs utilisés pour déterminer les différentes évaluations de densité de sonie considérées dans ce document sont issus des travaux de Zwicker *et coll.*. La densité de sonie totale de nos signaux est calculée à l'aide d'un programme normalement destiné à l'étude des signaux stationnaires [61][62]. La sensibilité de l'oreille et l'étalement spectral lié au fonctionnement du système auditif sont modélisés. Ils sont appliqués de façon globale sur le spectre des signaux. La densité de sonie totale estime un spectre perçu de ces signaux, selon l'échelle des fréquences perçues ici modélisée par l'échelle des Barks. Les densités de sonie 1 et 2 sont elles calculées en fonction du temps. Elles permettent de considérer les signaux dans un plan temps-fréquence perçu. La densité de sonie 1 est obtenue simplement en appliquant le programme du calcul de la densité de sonie totale sur des portions successives de 100 ms du signal, ces portions étant alors considérées comme stationnaires. La densité de sonie 2 est calculée à l'aide d'un deuxième programme, dédié à l'étude des signaux non-stationnaires, qui ajoute la prise en compte de l'étalement temporel du signal dû au fonctionnement du système auditif [66][83]. Cette densité de sonie 2 est déterminée toutes les 10 ms. Les moyennes temporelles des densités de sonie 1 et 2 ont également été considérées. Comme la densité de sonie totale, ces densités de sonie moyennes estiment un spectre perçu des signaux, mais celui-ci est calculé différemment, en tenant notamment compte du caractère non-stationnaire des signaux étudiés.

Boulet [25] rappelle le principe de ces programmes. Des pondérations fréquentielles sont dans un premier temps appliquées aux signaux pour modéliser les filtrages par l'oreille externe

et l'oreille moyenne. Il s'agit ensuite de déterminer le niveau du signal par bande critique afin d'obtenir le diagramme d'excitation de la membrane basilaire [65][66][67]. "La sonie est liée à la répartition de l'excitation, provoquée par un stimulus, le long de la membrane basilaire. Par conséquent, on représente le fonctionnement de l'oreille ou plus précisément de la membrane basilaire, comme celui d'un banc de filtres. Ces filtres "physiologiques" sont modélisés selon Zwicker par des bandes critiques ; ce sont des bandes spectrales adjacentes dont l'unité est le Bark. La largeur des bandes critiques est constante (100 Hz) en dessous de 500 Hz. Au dessus de 500 Hz, cette largeur augmente avec la fréquence." Boulet [25] (Tableau E.2). La sonie dans chaque bande est ensuite calculée par une fonction puissance du niveau. La densité de sonie est finalement obtenue en modélisant l'étalement spectral lié au fonctionnement du système auditif, à partir des diagrammes de masquage d'un son pur masqué par un bruit à spectre étroit. Le programme calculant la densité de sonie 2 modélise les phénomènes d'étalement temporel au sein du système auditif par un quadripôle [66][83]. Ces phénomènes sont notamment responsable du masquage auditif postérieur, un son pouvant être masqué par un autre son le précédant [65][66][67].

Les phénomènes d'étalement spectral liés au système auditif et modélisés lors du calcul des différentes densités de sonie dépendent du contenu spectral des signaux considérés, ce contenu pouvant évoluer au cours du temps. Pour la densité de sonie totale, le modèle auditif est appliqué de façon globale sur l'ensemble du signal, alors que le calcul des densités de sonie 1 et 2, et donc celui de leur moyenne temporelle, est effectué en fonction du temps. Par rapport à la densité de sonie totale, ces dernières analyses tiennent compte du fait que les phénomènes d'étalement spectral liés au système auditif dépendent de l'évolution temporelle des signaux non-stationnaires considérés.

Tiers d'octave	Bande critique approchée	Fréquence "centrale" (en Hz)	Limite de bande	
			(en Hz)	(en Bark)
1 à 6	1	63	90	0,9
7 à 9	2	125	180	1,8
10 et 11	3	224	280	2,8
12	4	315	355	3,5
13	5	400	450	4,4
14	6	500	560	5,4
15	7	630	710	6,6
16	8	800	900	7,9
17	9	1000	1120	9,2
18	10	1250	1400	10,6
19	11	1600	1800	12,3
20	12	2000	2240	13,8
21	13	2500	2800	15,2
22	14	3150	3550	16,7
23	15	4000	4500	18,1
24	16	5000	5600	19,3
25	17	6300	7100	20,6
26	18	8000	9000	21,8
27	19	10000	11200	22,7
28	20	12500	14000	23,6
29	21	-	15500	24

TABLEAU E.2 – Correspondance entre les tiers d'octave, les bandes critiques approchées considérées dans les modèles auditifs utilisés, et les gammes fréquentielles correspondantes [66]

Résumé

Cette étude cherche à relier les évaluations physiques et perceptives de différences de restitution du timbre par des enceintes acoustiques. Un protocole expérimental compatible avec les deux approches a été mis au point. Il consiste dans un premier temps à enregistrer le champ acoustique rayonné par des enceintes dans une pièce. Ces enregistrements sont ensuite soumis, d'une part à des tests d'écoute au casque, et d'autre part à des analyses de signal. Différentes méthodes d'analyse ont été testées, et pour chacune de ces méthodes, une "métrique" permettant d'évaluer la dissemblance entre deux signaux a été définie. Les dissemblances objectives ainsi obtenues sont alors comparées aux dissemblances perceptives résultant des tests psychoacoustiques. Dans un premier temps, une comparaison directe est réalisée en calculant le coefficient de corrélation entre les dissemblances. Ensuite, un deuxième type de comparaison permet de s'affranchir du critère de linéarité, et prend en compte le caractère multidimensionnel de notre perception de la reproduction sonore : les dissemblances perceptives et objectives sont soumises à une analyse multidimensionnelle, et les espaces ainsi obtenus sont comparés visuellement. Nos premiers résultats montrent l'importance d'utiliser des modèles auditifs afin de différencier les signaux d'une manière analogue à celle des auditeurs. Des attributs objectifs préliminaires sont également proposés. Ils permettent de décrire deux des dimensions principales utilisées par les auditeurs pour différencier nos enregistrements.

Abstract

This study looks for relationships between perceptual and physical evaluations of differences between the restitution of timbre by loudspeakers. An experimental protocol compatible with both approaches was designed. It first consists in recording the sound radiated by loudspeakers in a room, and then submitting the recorded sounds to listening tests under headphones, on the one hand, as well as to signal analysis, on the other hand. Different methods of signal analysis were tested, and for each method, a "metric" evaluating the dissimilarity between two signals was defined. The resulting objective dissimilarities are then compared to the perceptual ones obtained during listening tests. First, a direct comparison is realized by calculating the correlation between dissimilarities. Then, a second comparison abandons the criterion of linearity, and takes into account the multidimensional nature of the perception of sound reproduction : perceptual and objective dissimilarities are submitted to multidimensional scaling analysis, and the resulting spaces are visually compared. Our first results showed the importance of using auditory models to differentiate the signals in the same way as listeners did. Preliminary objective attributes describing two of the main perceptual dimensions used by listeners to differentiate our recordings are also proposed.