



HAL
open science

Distorsions des systèmes de reproduction musicale : Protocole de caractérisation perceptive

Pierre-Yohan Michaud

► **To cite this version:**

Pierre-Yohan Michaud. Distorsions des systèmes de reproduction musicale: Protocole de caractérisation perceptive. Acoustique [physics.class-ph]. Aix-Marseille Université, 2012. Français. NNT : . tel-00806288v2

HAL Id: tel-00806288

<https://theses.hal.science/tel-00806288v2>

Submitted on 12 Dec 2013

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

AIX-MARSEILLE UNIVERSITÉ

ÉCOLE DOCTORALE SCIENCES POUR L'INGÉNIEUR :
MÉCANIQUE, PHYSIQUE, MICRO ET NANOÉLECTRONIQUE (ED353)

THÈSE

Pour obtenir le grade de

DOCTEUR DE L'UNIVERSITÉ AIX-MARSEILLE

Discipline : **ACOUSTIQUE**

Présentée et soutenue publiquement par

PIERRE-YOHAN MICHAUD

le 30 Mars 2012

**DISTORSIONS DES SYSTÈMES DE
REPRODUCTION MUSICALE :
PROTOCOLE DE CARACTÉRISATION
PERCEPTIVE**

Directeurs de thèse :

Philippe HERZOG – Sabine MEUNIER

JURY

Pr. Hugo Fastl	MMK - Technische Universität München	Rapporteur
Dr. Patrick Susini	IRCAM	Rapporteur
Pr. Wolfgang Ellermeier	Technische Universität Darmstadt	Examineur
Dr. Rozenn Nicol	Orange Labs	Examineur
Pr. Laurent Simon	LAUM - Université du Maine	Examineur
Pr. Bruno Torrèsani	LATP - Université de Provence	Examineur
Pr. Gérard d'Aubigny	MS3 LJK - Université Pierre-Mendès-France	Invité
Dr. Philippe Herzog	LMA	Co-directeur de thèse
Dr. Sabine Meunier	LMA	Co-directrice de thèse

Résumé :

Ces travaux concernent l'évaluation perceptive de la reproduction sonore. Le but de notre étude est de proposer un protocole permettant la caractérisation perceptive des non-linéarités et des phénomènes de distorsion modifiant le signal reproduit par un système tel qu'une enceinte acoustique. L'élaboration d'un tel protocole nécessite de rassembler de nombreuses enceintes et d'utiliser une méthode d'évaluation adaptée. Nous ne cherchons pas à évaluer la distorsion en terme de qualité globale mais plutôt à estimer les dissemblances perçues entre différentes enceintes distordues afin de révéler, grâce à une analyse multidimensionnelle MDS, les critères sur lesquelles les auditeurs se basent pour les différencier. Dans un premier temps, nous avons utilisé des signaux de synthèse afin de générer des échantillons dont la distorsion est contrôlable. Nous avons ainsi créé un corpus composé de nombreux échantillons distordus simulant ou se rapprochant du fonctionnement d'une enceinte acoustique avec différents types de non-linéarités. Ensuite nous proposons une méthode adaptée à l'évaluation de la dissemblance sur des corpus étendus que nous avons validée à partir de simulations et de tests d'écoute. Enfin, nous avons appliqué cette méthode d'évaluation sur le corpus d'enceintes que nous avons élaboré afin de tester l'utilisation du protocole proposé et de donner quelques résultats préliminaires concernant les dimensions perceptives liées à la distorsion non linéaire.

Abstract :

This work deals with the perceptual evaluation of the sound reproduction. Our aim is to elaborate a protocol allowing to assess the influence of nonlinearities and distortions of a loudspeaker. Such a protocol requires gathering a large set of loudspeakers together with an appropriate method for its evaluation. This study deals with the perceptual evaluation of dissimilarities between different loudspeakers rather than their absolute quality judgements. Thanks to a multidimensional scaling technique, this approach allows to uncover the criteria used by the listeners to differentiate various distorting loudspeakers. First, we used synthesized signals to generate samples of sounds resulting from controlled distortion. Numerous distorted stimuli have been generated that simulate sounds radiated by a loudspeaker. Then, we propose a method suitable for the evaluation of a large set of stimuli. In order to verify its application to the estimation of dissimilarity, simulations and listening tests have been conducted. Finally, this listening test method has been applied to the created set of distorting loudspeakers in order to verify the use of our protocol and provide some preliminaries results concerning the perception of nonlinear distortion.

Table des matières

Contexte : Caractériser la perception de la distorsion

Chapitre 1

Étude de la distorsion

1.1	Non-linéarité et distorsion	7
1.1.1	Généralités	7
1.1.2	Distorsion linéaire	8
1.1.3	Distorsion non linéaire	9
1.2	Évaluations objectives de la distorsion non linéaire	10
1.2.1	Distorsion harmonique spécifique	10
1.2.2	THD : Distorsion harmonique totale	11
1.2.3	IMD : Distorsion d'intermodulation	11
1.2.4	MTND : Distorsion pour un signal multifréquentiel	12
1.2.5	Cohérence	13
1.2.6	Distorsion résiduelle	14
1.3	Limitations de l'évaluation objective de la distorsion non linéaire	14
1.3.1	Complexité du système de restitution	15
1.3.2	Complexité du signal	18
1.3.3	Complexité du système auditif	19
1.4	Évaluation perceptive de la distorsion non linéaire	22
1.4.1	Caractérisation des systèmes de restitution :	23
1.4.2	Caractérisation de la perception de la distorsion non linéaire	24
1.5	Corrélations entre évaluations perceptives et objectives	26
1.6	Bilan de l'étude de la distorsion non linéaire	27

Chapitre 2

Approche, choix et protocole proposé

2.1	Caractérisation multidimensionnelle de signaux musicaux distordus	33
2.1.1	Application à la reproduction sonore	34
2.1.2	Besoins liés à l'étude multidimensionnelle de la distorsion non linéaire . .	35
2.2	Élaboration d'un corpus de stimuli distordus	38
2.2.1	Modélisation de haut-parleurs	39
2.2.2	Contrôle de la non-linéarité	41
2.2.3	Restitution	42
2.3	Recherche d'une méthode d'évaluation de la dissemblance	43
2.3.1	Méthodes usuelles pour l'évaluation de la dissemblance	43
2.3.2	Méthodes alternatives adaptées à l'évaluation de corpus étendus	47
2.3.3	Méthode retenue	50
2.4	Protocole proposé	51

Chapitre 3

Élaboration et contrôle d'une enceinte "virtuelle"

3.1	Enceinte "génératrice"	55
3.2	Séparation des deux voies de l'enceinte génératrice	56
3.3	Modèle de boomer	58
3.3.1	Analyse : Identification des non-linéarités	58
3.3.2	Synthèse : Décomposition sur la base des polynômes de Chebyshev	61
3.3.3	Modélisation du boomer	63
3.3.4	Vérification objective du modèle de boomer	64
3.4	Modèle d'enceintes virtuelles	68
3.4.1	Enregistrement du tweeter	69
3.4.2	Synchronisation des voies	69
3.4.3	Équilibrage des niveaux	71
3.5	Auralisation	71
3.6	Vérification perceptive : enceinte virtuelle vs. enceinte réelle	72
3.6.1	Procédure	73
3.6.2	Enregistrements réels et virtuels	73
3.6.3	Tests ABX	74
3.6.4	Résultats	74

3.7	Contrôle des non-linéarités dans les enceintes virtuelles	76
3.7.1	Modification des non-linéarités identifiées sur le boomer	76
3.7.2	Introduction de non-linéarités artificielles	77

Chapitre 4

Méthode de comparaison avec permutation de références

4.1	Adaptation à l'évaluation de stimuli audio	83
4.2	Particularités de la méthode	84
4.3	Simulations	86
4.3.1	Simulation 1 : Influence du nombre de stimuli	88
4.3.2	Simulation 2 : Influence du nombre d'auditeurs	88
4.4	Résultats	88
4.4.1	Relation entre les dissemblances connues et estimées	88
4.4.2	Nombre de stimuli	89
4.4.3	Nombre d'auditeurs simulés	90
4.4.4	Caractérisation du biais	91

Chapitre 5

Validation de la méthode de comparaison avec permutation de références

5.1	Test 1 : Évaluation d'un corpus de 12 enceintes	97
5.1.1	Stimuli	97
5.1.2	Auditeurs	97
5.1.3	Procédure	97
5.1.4	Résultats	98
5.2	Test 2 : Évaluation d'un corpus de 37 enceintes	102
5.2.1	Stimuli	102
5.2.2	Auditeurs	102
5.2.3	Procédure	102
5.2.4	Résultats	102
5.3	Influence de la méthode d'évaluation	104

Chapitre 6

Application du protocole pour l'évaluation de la distorsion non linéaire

6.1	Test 1 : Évaluation d'un corpus de 31 enceintes distordues	111
6.1.1	Stimuli	111

6.1.2	Auditeurs	113
6.1.3	Procédure	114
6.1.4	Résultats	114
6.1.5	Bilan intermédiaire	122
6.2	Test 2 : Évaluation d'un corpus de 19 enceintes distordues	122
6.2.1	Stimuli	122
6.2.2	Auditeurs	122
6.2.3	Procédure	122
6.2.4	Résultats	123
6.3	Discussion	124

Conclusion et perspectives

Bibliographie	133
----------------------	------------

Annexe A Enceinte Tannoy System 600	141
--	------------

Annexe B Approximation à l'aide des polynômes de Chebyshev	145
---	------------

Annexe C Salle d'écoute et réponse impulsionnelle	147
--	------------

Annexe D Analyse Multidimensionnelle	149
---	------------

Annexe E Mesure de la distorsion du casque utilisé pour les tests d'écoute	153
---	------------

Contexte : Caractériser la perception de la distorsion

Les systèmes de reproduction sonores sont conçus pour reproduire fidèlement un signal sur une grande plage de dynamique, avec une réponse en fréquence généralement bien contrôlée par la majorité des constructeurs. Ces différents critères de restitution sont contraints par la capacité de l'amplificateur, les propriétés géométriques des enceintes, les matériaux qui composent les haut-parleurs et les composants des filtres dans le cas des enceintes multivoies. Tous ces éléments ont une influence sur le son rayonné par le système et peuvent, en fonction de leurs caractéristiques, modifier le signal, introduisant des changements de timbre, des variations de rayonnement spatial. Il est alors intéressant d'évaluer perceptivement l'influence de ces modifications afin de comprendre la manière dont nous percevons ces différents phénomènes.

Cette étude s'inscrit dans un contexte général d'évaluation de qualité sonore. Avant tout, il est nécessaire de préciser que deux définitions du terme qualité peuvent être distinguées. Chaque définition est associée à une approche dont les motivations diffèrent d'un sens à l'autre et une telle distinction est primordiale pour situer nos travaux. D'un côté, la qualité est associée à un critère de "fidélité" de la restitution où l'évaluation se fait au sens de "bon/mauvais" à partir de jugements de préférence. De l'autre, le terme qualité sonore est aussi employé pour désigner l'étude des caractéristiques d'un son indépendamment d'un jugement de valeur. Cette approche plus exploratoire permet de mettre en évidence les caractéristiques des sons importantes du point de vue de la perception et ainsi définir des attributs qui différencient perceptivement les sons au sein d'un corpus.

Nous nous intéressons à la caractérisation perceptive de la reproduction sonore et plus particulièrement à déterminer comment les caractéristiques d'un son reproduit sont perçues. Il faut bien distinguer l'évaluation d'un système de reproduction et l'évaluation du signal reproduit. La première concerne plus particulièrement l'objet et la deuxième est davantage portée sur le signal restitué. Dans notre étude, le système de reproduction est la cause physique des modifications des caractéristiques du signal, mais nous nous intéressons dans ce travail au signal musical reproduit par le système en tant que stimulus ou échantillon sonore, et non au système qui a servi à obtenir

ce signal.

Cette étude constitue la suite des travaux sur la caractérisation de la reproduction du timbre par un système de reproduction, en particulier les enceintes acoustiques, initiés au laboratoire par Lavandier [2005]. Au cours de sa thèse, Lavandier [2005] a mis en place un protocole permettant d'aborder en parallèle l'évaluation objective et perceptuelle des caractéristiques perçues dans la restitution de signaux musicaux. Ce protocole est basé sur la mesure de dissemblances perceptuelles et physiques entre différents enregistrements d'enceintes. L'objectif est de comparer les mesures objectives et perceptuelles afin de trouver la mesure physique la plus pertinente possible vis-à-vis de la perception de la reproduction du timbre. Nous avons prolongé cette thématique de recherche vers une autre composante essentielle de la restitution sonore : la distorsion, dont les effets sont susceptibles de modifier différemment des caractéristiques spécifiques des signaux reproduits.

Le phénomène de distorsion provient du caractère parfois non linéaire du système de restitution, et se traduit entre autres par la génération de composantes qui n'étaient pas présentes dans le signal d'entrée. Les effets de la distorsion sur le signal restitué ne dépendent pas que des caractéristiques du système de reproduction, mais aussi du niveau et du contenu du signal d'entrée. Certains constructeurs de produits audio ne cherchent pas à réduire à tout prix la distorsion et peuvent même chercher à ajouter des harmoniques au signal d'entrée, par exemple pour renforcer la sensation de basses dans le signal perçu. Cependant, l'objectif de la plupart des constructeurs est de diminuer les distorsions. Une valeur correspondant à un indicateur de la distorsion figure alors parfois sur les fiches techniques accompagnant les systèmes de reproduction (amplificateur ou enceintes). Cette valeur correspond généralement à une mesure de distorsion objective normalisée et réalisée avec des signaux stationnaires. L'utilisation de mesures normalisées permet une comparaison des produits lorsque les conditions dans lesquelles les mesures de distorsion ont été effectuées sont précisées. Néanmoins, ces mesures employées par les constructeurs ne prennent en compte ni la réponse du système à un signal musical, ni la manière dont est perçue une telle réponse par un auditeur. A long terme, notre objectif est de fournir une méthode de caractérisation de l'effet de la distorsion qui soit représentative de ce que nous percevons dans des conditions réalistes d'écoute. Pour élaborer une telle méthode de mesure, il est nécessaire de comprendre la manière dont nous percevons les phénomènes de distorsion dans un signal reproduit.

Comme nous l'avons défini précédemment, nous ne cherchons pas à caractériser quels sont les éléments de la chaîne de restitution qui introduisent de la distorsion mais à étudier quels sont les effets perçus de la distorsion dans le signal reproduit. Pour cette étude, nous nous sommes focalisés sur l'enceinte acoustique dont la distorsion domine celle des autres éléments de la chaîne de restitution. Nous avons choisi d'étudier les effets de la distorsion en se rapprochant le plus possible d'une situation d'écoute musicale. Nous avons donc tenu à respecter le caractère écologique d'une telle écoute et les précautions ont été prises afin d'effectuer, autant que possible, des tests dans des conditions quotidiennes et réalistes d'écoute de musique reproduite. De plus, notre approche

est basée sur les jugements d'auditeurs naïfs (non spécialisés et sans entraînement particulier) afin d'étudier ce que perçoit un auditeur "moyen". Nous avons cherché à obtenir un jugement global afin de retrouver quelles sont les informations les plus importantes contenues dans le signal qui permettent à un auditeur moyen de percevoir des différences entre multiples échantillons sonores distordus. Pour pouvoir mettre en évidence plusieurs attributs dans cette démarche de caractérisation, il est nécessaire de faire intervenir un grand nombre d'échantillons sonores. Un résultat important des travaux de Lavandier [2005] est le nombre supérieur d'attributs révélé pour l'étude du timbre sur un corpus de 37 enceintes par rapport à un corpus de 12 enceintes. C'est un point essentiel de l'évaluation de la reproduction sonore et il est impératif d'avoir un nombre suffisant d'échantillons afin d'explorer au maximum l'espace perceptif associé.

L'objectif du travail présenté dans ce document est donc de proposer un protocole permettant la caractérisation perceptive de l'effet de la distorsion sur un signal musical. L'élaboration de ce protocole a nécessité de développer deux aspects : l'élaboration d'un corpus d'échantillons sonores dont les caractéristiques liées à la distorsion sont contrôlées, et la mise au point d'une méthode permettant l'évaluation de dissemblances entre de nombreux stimuli. Dans ce mémoire ces deux aspects sont abordés à plusieurs reprises pour construire le protocole final en intégrant peu à peu les contraintes associées à ces deux points.

Nous présentons dans le chapitre 1 un état de l'art des méthodes d'évaluation objective et perceptive des non-linéarités. Nous insistons sur les limitations des mesures objectives et développons les liens entre les deux types d'évaluations. Le chapitre 2 expose l'approche et les choix que nous avons effectués pour l'élaboration du protocole que nous proposons. Le chapitre 3 traite ensuite du premier point du protocole qui concerne l'élaboration d'échantillons sonores correspondant aux signaux reproduits par une enceinte dont les non-linéarités peuvent être contrôlées. Les chapitres 4 et 5 sont dédiés à la présentation et à la validation d'une méthode de mesure perceptive adaptée à l'évaluation de corpus étendus de stimuli. Enfin, le chapitre 6 illustre un exemple d'application du protocole proposé sur un corpus d'enceintes distordues et donne quelques résultats préliminaires concernant la perception de l'effet de la distorsion non linéaire.

Chapitre 1

Étude de la distorsion

Sommaire

1.1	Non-linéarité et distorsion	7
1.1.1	Généralités	7
1.1.2	Distorsion linéaire	8
1.1.3	Distorsion non linéaire	9
1.2	Évaluations objectives de la distorsion non linéaire	10
1.2.1	Distorsion harmonique spécifique	10
1.2.2	THD : Distorsion harmonique totale	11
1.2.3	IMD : Distorsion d'intermodulation	11
1.2.4	MTND : Distorsion pour un signal multifréquentiel	12
1.2.5	Cohérence	13
1.2.6	Distorsion résiduelle	14
1.3	Limitations de l'évaluation objective de la distorsion non linéaire	14
1.3.1	Complexité du système de restitution	15
1.3.2	Complexité du signal	18
1.3.3	Complexité du système auditif	19
1.4	Évaluation perceptive de la distorsion non linéaire	22
1.4.1	Caractérisation des systèmes de restitution :	23
1.4.2	Caractérisation de la perception de la distorsion non linéaire	24
1.5	Corrélations entre évaluations perceptives et objectives	26
1.6	Bilan de l'étude de la distorsion non linéaire	27

Ce chapitre présente les évaluations objectives et perceptives de la distorsion. Dans un premier temps, nous introduisons de manière générale la notion de non-linéarité d'un système et les phénomènes de distorsion engendrés par des systèmes non linéaires tels que les enceintes acoustiques. Nous exposons ensuite les évaluations objectives conventionnelles de la distorsion et leurs limites afin d'introduire les différentes études ayant évalué la distorsion d'un point de vue perceptif et les corrélations entre les deux approches.

1.1 Non-linéarité et distorsion

1.1.1 Généralités

Un système \mathbf{S} peut être défini par une relation qui lie un signal d'entrée $x(t)$ et une réponse $y(t)$ (ou $S(x)$). Dans ce cas, de tels systèmes sont appelés système SISO (Single Input Single Output). Si ce système est linéaire, il obéit au principe de superposition : additivité et homogénéité. L'additivité est rencontrée lorsque la réponse résultant de plusieurs signaux d'entrée est égale à la somme des réponses individuelles de chaque signal (équation 1.1). L'homogénéité se traduit par l'égalité entre la réponse à un signal atténué d'un facteur a et la réponse, au même signal, atténuée du facteur a (équation 1.2).

$$\mathbf{S}(x_1 + x_2 + \dots + x_n) = \mathbf{S}(x_1) + \mathbf{S}(x_2) + \dots + \mathbf{S}(x_n). \quad (1.1)$$

$$\mathbf{S}(ax) = a\mathbf{S}(x). \quad (1.2)$$

Un système est alors dit non linéaire lorsque le principe de superposition n'est pas respecté [Novák, 2009]. Un système est dit invariant lorsque son comportement n'est pas modifié dans le temps mais la plupart des systèmes physiques ne sont ni invariants, ni linéaires. La linéarité et la modification du système au cours du temps sont deux aspects liés. Un système peut être considéré comme l'association d'une composante linéaire et une autre composante non linéaire, souvent considérée pour simplifier comme invariante et "sans mémoire".

Une manière de représenter le comportement des non-linéarités d'un tel système consiste à tracer le signal de sortie en fonction du signal d'entrée. Des exemples de courbes entrée/sortie sont présentés sur la figure 1.1. Un système linéaire et sans mémoire correspond à la ligne droite et les systèmes non linéaires sont symbolisés par les exemples de courbe. Généralement, un système physique est approximativement linéaire pour une certaine plage de dynamique et s'écarte de ce régime linéaire pour des niveaux d'entrée différents.

La majorité des relations entrée/sortie représentant un système physique présente une courbure plus ou moins prononcée. Ce comportement peut être représenté par une somme de poly-

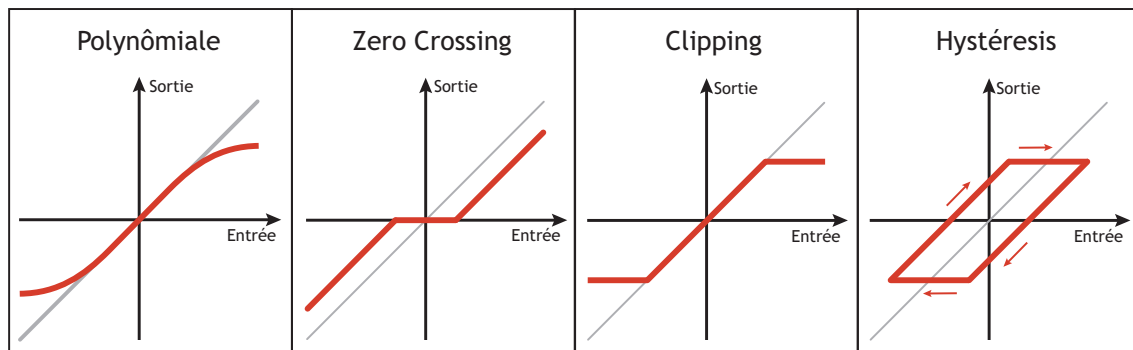


Figure 1.1 – Représentations de courbes entrée/sortie pour les types de non-linéarités polynomiale, zero crossing, clipping et hystérésis.

nômes. La non-linéarité de croisement dite "Zero Crossing" est une forme qui affecte le signal lorsque le niveau d'entrée est proche de zéro. La non-linéarité d'écrêtage dite "Clipping" est la forme de distorsion qui apparaît lorsque le système atteint la limite de son régime linéaire. Lorsque le niveau est trop important, le signal de sortie est alors écrêté. Ces trois exemples de non-linéarités représentent des systèmes sans mémoire. Un exemple de système non linéaire avec un effet de mémoire est une relation entrée/sortie présentant un caractère hystérétique. Dans ce cas particulier, les composantes linéaire et non linéaire sont liées et le système présente un comportement différent lorsque l'amplitude du signal d'entrée augmente ou diminue. En fonction de ses caractéristiques non linéaires, le système introduit alors de la distorsion qui correspond à la modification du signal de sortie. Cependant, la distorsion n'est pas forcément non linéaire. Deux classes de distorsions peuvent être distinguées : la distorsion linéaire et la distorsion non linéaire [Cabot, 1990].

1.1.2 Distorsion linéaire

La distorsion linéaire agit uniquement sur les amplitudes et les phases des composantes présentes dans le signal d'entrée. L'amplitude du spectre est modifiée pour certaines fréquences, et un retard résulte de la modification de la phase. Un système linéaire modifie le spectre et la phase du signal de manière indépendante du signal alors que, dans le cas d'un système non linéaire, la modification de ces deux caractéristiques dépend du contenu du signal d'entrée. La distorsion linéaire provient des caractéristiques électriques ou mécaniques du système considéré. Une certaine quantité de l'énergie du signal peut être absorbée, accentuée, conservée puis libérée plus tard, réfléchi ou propagée à travers certains matériaux à des vitesses différentes [Preis, 1984]. Ces différentes causes de distorsion linéaire sont présentes dans les enceintes acoustiques dont le but est de convertir un signal électrique en signal acoustique le plus rapidement possible. Les effets de la distorsion linéaire influencent le timbre du système de restitution car la réponse en fréquence et la phase sont altérées. De nombreuses études se sont intéressées à caractériser

d'un point de vue perceptif les effets engendrés par la distorsion linéaire de systèmes tels que les enceintes acoustiques. Pour une revue sur le sujet, un état de l'art sur la perception de la reproduction du timbre par les enceintes acoustiques est effectué par [Lavandier, 2005].

1.1.3 Distorsion non linéaire

La distorsion non linéaire est caractérisée par la présence de composantes supplémentaires dans le signal de sortie, composantes qui n'étaient pas présentes dans le signal d'entrée du système. Ce système ne module pas seulement l'amplitude et la phase du signal d'entrée comme c'est le cas pour la distorsion linéaire. Les composantes introduites par le système non linéaire dépendent des caractéristiques du système, de la nature et du niveau du signal d'entrée. La distorsion non linéaire peut notamment être mise en évidence par deux critères usuels : la distorsion harmonique et la distorsion d'intermodulation. Lorsque le signal d'entrée est un signal composé d'une seule sinusoïde, la distorsion harmonique correspond en général à la présence d'harmoniques supérieures dans le signal de sortie. Pour un signal composé de plusieurs fréquences, la distorsion d'intermodulation correspond à l'interaction entre les composantes du signal d'entrée. Par exemple, pour un signal composé de deux fréquences f_1 et f_2 , la distorsion non linéaire se caractérise par les produits d'intermodulation, combinaisons de sommes et de différences de f_1 et f_2 , dans le signal de sortie. L'ordre du produit d'intermodulation dépend de la combinaison entre f_1 et f_2 . Le tableau 1.1 résume les composantes présentes dans le signal de sortie pour des produits de distorsion allant jusqu'à l'ordre 4 pour une fréquence $f_2 > f_1$.

Ordre 1	Ordre 2	Ordre 3	Ordre 4
f_1	$2f_1$	$3f_1$	$4f_1$
f_2	$2f_2$	$3f_2$	$4f_2$
	$f_2 - f_1$	$2f_1 - f_2$	$3f_1 - f_2$
	$f_2 + f_1$	$2f_1 + f_2$	$3f_1 + f_2$
		$2f_2 - f_1$	$3f_2 - f_1$
		$2f_2 + f_1$	$3f_2 + f_1$
			$2f_1 - 2f_2$
			$2f_2 - 2f_1$
			$2f_1 + 2f_2$

Tableau 1.1 – Composantes fréquentielles présentes dans le signal de sortie d'un système non linéaire excité par deux fréquences f_1 et f_2 (avec $f_2 > f_1$). Les éléments présentés dans la première ligne en gras correspondent aux composantes de la distorsion harmonique pour un signal d'entrée f_1 . Les autres éléments du tableau correspondent aux produits d'intermodulation des deux fréquences du signal d'entrée.

Le tableau 1.1 montre que le nombre de produits de distorsion pour la distorsion d'intermodulation est plus important que pour la distorsion harmonique. Cette différence illustre le fait

que l'évaluation de la distorsion non linéaire est délicate car elle dépend à la fois du signal et du système. Il est alors nécessaire d'employer des méthodes d'évaluation adéquates en fonction du signal d'entrée utilisé.

1.2 Évaluations objectives de la distorsion non linéaire

Dans cette section, sont résumées les principales méthodes objectives permettant d'évaluer la distorsion non linéaire liée à un système de reproduction. Nous présentons les méthodes usuelles basées sur l'évaluation des distorsions d'amplitude. Les effets de phase ne sont pas pris en compte dans ces méthodes objectives. Pour chaque méthode, sont présentés le type de signal d'entrée utilisé et la manière de mesurer la distorsion. Une description plus complète des méthodes d'évaluation objective de la distorsion est présentée dans les articles de Czerwinski *et al.* [2001a] et Voishvillo *et al.* [2004].

1.2.1 Distorsion harmonique spécifique

Signal d'entrée : Signal monofréquentiel

Lorsqu'un système non linéaire est excité par un signal comportant une seule composante fréquentielle, sa réponse correspond à une série d'harmoniques du signal d'entrée (Figure 1.2). Cette méthode permet d'évaluer individuellement chaque composante du spectre du signal de sortie. La mesure correspond au rapport du niveau RMS de l'harmonique considérée sur le niveau RMS total du signal.

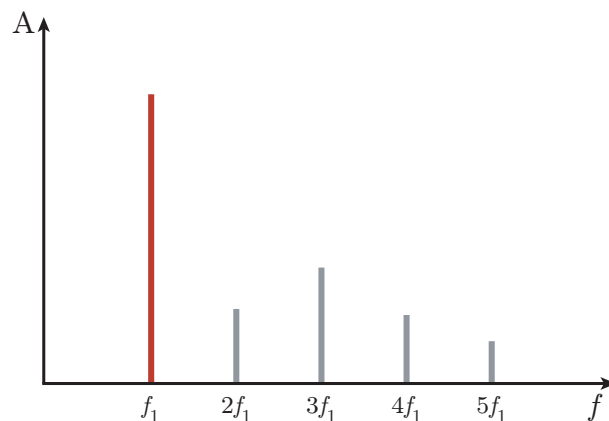


Figure 1.2 – Exemple de réponse d'un système non linéaire à un signal sinusoïdal monofréquentiel de fréquence f_1 .

1.2.2 THD : Distorsion harmonique totale

Signal d'entrée : Signal monofréquentiel

La distorsion harmonique totale (THD) est probablement la mesure de distorsion la plus ancienne. Elle est largement employée pour évaluer les distorsions de différents systèmes de restitution. Le principe de mesure est, de manière identique à celui de la distorsion harmonique spécifique, basé sur l'analyse des harmoniques créées par un signal composé d'une fréquence. La valeur de THD correspond au rapport entre le niveau RMS de l'ensemble des harmoniques sur le niveau RMS total du signal [IEC 60268-5, 1989]. Cette valeur est généralement donnée en pourcentage, 0% indiquant l'absence de distorsion non linéaire.

1.2.3 IMD : Distorsion d'intermodulation

Signal d'entrée : Deux signaux monofréquentiels de fréquences différentes

L'IMD ou distorsion d'intermodulation conduit à quantifier des produits de distorsion qui ne sont pas forcément liés harmoniquement au signal d'entrée. Ces produits de distorsion caractérisent la réponse du système lorsque le signal d'entrée est une somme de deux signaux monofréquentiels ayant des fréquences et des amplitudes différentes. Pour cette méthode, plusieurs combinaisons de signaux peuvent être employées mais deux méthodes sont principalement utilisées pour la mesure de l'IMD.

SMPTE (Society of Motion Picture and Television Engineers) :

Deux fréquences $f_1 = 60Hz$ et $f_2 = 7kHz$ sont régulièrement utilisées avec un rapport d'amplitude de 4 :1 entre les deux, correspondant à une différence de 12 dB. Ce critère cherche à évaluer la modulation des composantes hautes fréquences par la fréquence la plus basse du signal d'entrée (Figure 1.3). La valeur de l'IMD correspond à la somme des amplitudes des produits de distorsion sur l'amplitude de f_2 , la deuxième composante fréquentielle [Czerwinski *et al.*, 2001a].

CCIF (International Telephonic Consultative Committee) autrement appelé DFD (Difference Frequency Distortion) :

Cette mesure fait intervenir en entrée du système deux fréquences assez élevées qui ont la même amplitude mais une différence fréquentielle de 1kHz (Figure 1.4). La mesure d'intermodulation s'exprime par le rapport de la somme des amplitudes des produits de distorsion sur l'amplitude du signal d'entrée. Les produits de distorsions d'ordre pair produisent les composantes fréquentielles les plus basses correspondant à la différence entre les deux signaux d'entrée. Ceux d'ordres impairs produisent des composantes de différence à des fréquences plus élevées. La plupart des applications de ce critère cherchent à évaluer les ordres pairs des produits de distorsion [Cabot, 1999].

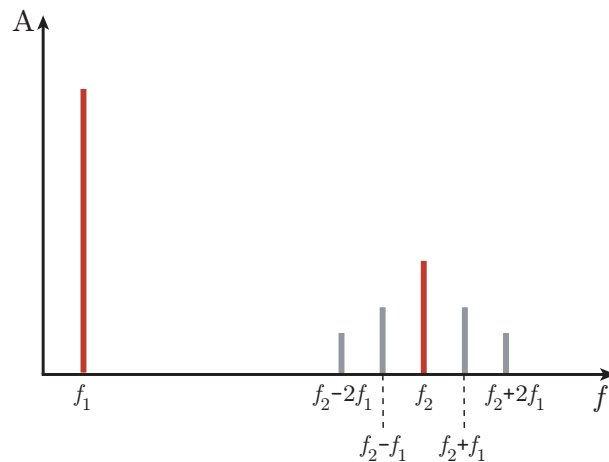


Figure 1.3 – Exemple de réponse d’un système non linéaire à un signal d’entrée composé de deux fréquences suivant le standard SMPTE.

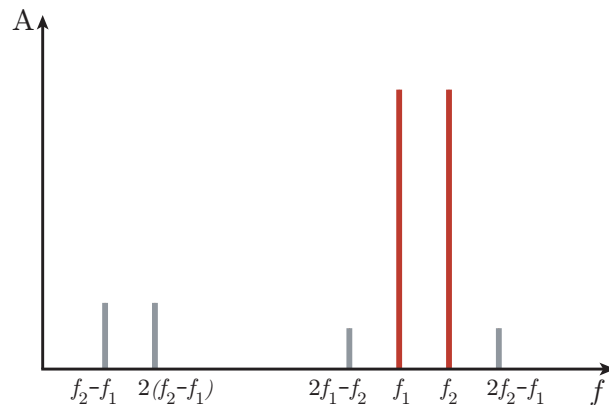


Figure 1.4 – Exemple de réponse d’un système non linéaire à un signal d’entrée composé de deux fréquences suivant le standard CCIF.

1.2.4 MTND : Distorsion pour un signal multifréquentiel

Signal d’entrée : Signal multifréquentiel

La MTND ou Multitone Total Nonlinear Distortion est un critère introduit par Czerwinski *et al.* [2001a,b], basé sur un signal multifréquentiel. Un système non linéaire soumis à un tel signal génère des produits de distorsion variés. La valeur de la MTND provient d’un post-traitement effectué sur la réponse du système lorsque ce dernier est excité par un signal multifréquentiel. La mesure correspond à la somme de l’énergie de toutes les composantes fréquentielles harmoniques et d’intermodulation dans un certain domaine fréquentiel. La MTND est souvent calculée par bandes de 1/3 d’octave. Au sein d’une bande fréquentielle, la valeur moyenne des produits de distorsion est calculée et attribuée à la fréquence centrale de la bande. Ensuite en déplaçant la bande fréquentielle, cette opération est renouvelée et une nouvelle valeur de MTND est attribuée

à cette nouvelle bande. Le processus est répété pour couvrir la totalité du spectre du signal. La figure 1.5 montre la réponse d'un haut-parleur à un signal multifréquentiel accompagnée de la mesure de MTND en fonction de la fréquence [Voishvillo, 2002]. Ainsi, la valeur de cette métrique pour une fréquence donnée dépend de l'amplitude et de la densité des composantes fréquentielles du signal de sortie au voisinage de cette fréquence.

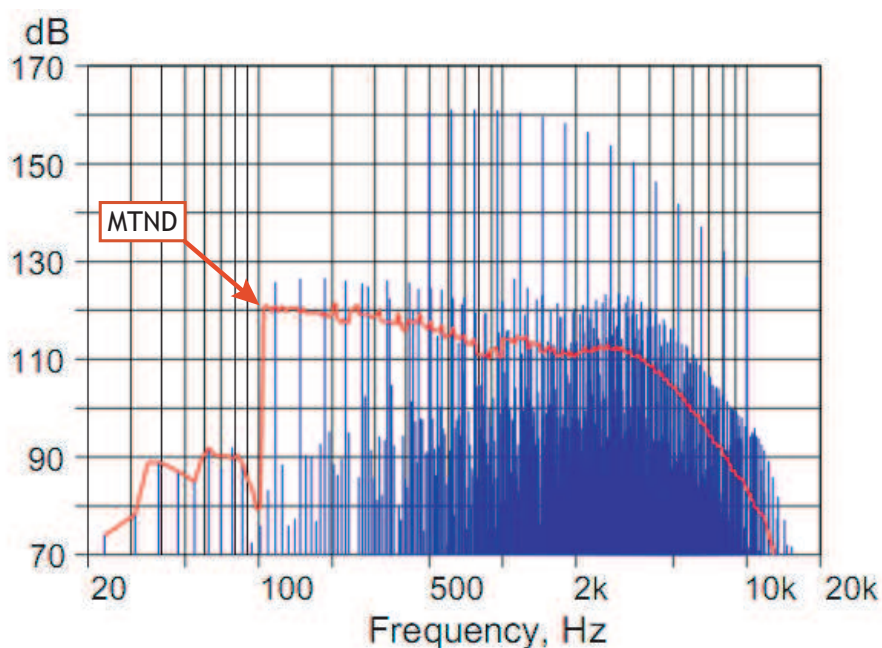


Figure 1.5 – Réponse d'un système non linéaire à un signal multifréquentiel et mesure de MTND effectuée sur cette réponse (d'après Voishvillo [2002]).

1.2.5 Cohérence

Signal d'entrée : Parole, Bruit

Cette méthode est généralement utilisée pour caractériser les appareils auditifs et les tests de distorsion sont effectués sur des signaux de parole. Cependant, du bruit modulé en amplitude et ayant la forme d'un signal de parole peut être utilisé. La mesure à une fréquence donnée correspond au rapport du carré de l'interspectre entre l'entrée et la sortie du système sur le produit de l'autospectre du signal d'entrée et celui du signal de sortie [Kates et Kozma-Spytek, 1994; Kates et Arehart, 2005]. La fonction de cohérence représente la partie du signal de sortie qui est linéairement reliée au signal d'entrée pour chaque fréquence analysée. Pour exprimer une valeur caractérisant la non-linéarité du système, il est plus pratique d'exprimer la fonction d'incohérence. La fonction d'incohérence est exprimée de manière similaire à la mesure de THD : une valeur d'incohérence de 0% indiquant l'absence de distorsion non linéaire. La THD ne permet de prendre en compte que les harmoniques alors que la fonction d'incohérence est sensible plus

généralement à la dégradation générale du signal de sortie, qui peut comporter des composantes non harmoniques et du bruit [Voishvillo *et al.*, 2004].

1.2.6 Distorsion résiduelle

Signal d'entrée : Signal sinusoïdal

La distorsion résiduelle correspond à une mesure effectuée sur le signal de sortie dans lequel la composante fréquentielle linéairement liée au signal d'entrée a été retirée. Ces mesures sont adaptées aux faibles niveaux de distorsion harmonique et lorsque du bruit contamine la mesure de distorsion [Cabot, 1999]. Un critère appelé THD+N (THD + Noise) est utilisé afin de prendre en compte la contribution du bruit dans le signal de sortie. Pour cette méthode, il faut dans un premier temps supprimer la fréquence fondamentale du signal à tester en utilisant un filtre coupe bande (filtre notch). Ainsi, il ne reste qu'un signal "résiduel", comportant les harmoniques et le bruit en sortie. La valeur de THD+N correspond au rapport entre les valeurs RMS du résidu et la valeur RMS de la composante fondamentale.

Afin de détecter la présence de défauts dans la construction des enceintes, Klippel [2003] propose d'autres méthodes d'estimation de la distorsion résiduelle qui correspondent à la différence entre le signal mesuré et le signal d'entrée. Les mesures fournissent une évaluation dans le domaine temporel et fréquentiel du niveau de distorsion résiduelle. Une autre mesure de distorsion résiduelle est proposée par Klippel [2011] et permet d'évaluer la distorsion liée au bruit généré par l'écoulement à travers une fuite de l'enceinte. La mesure est basée sur l'estimation du bruit en supprimant la fréquence fondamentale et les harmoniques avec un filtre à encoches (filtre notch). Le niveau de Modulation absolue (MOD_{Abs}) correspond au niveau du signal résiduel par rapport au seuil absolu d'audition.

1.3 Limitations de l'évaluation objective de la distorsion non linéaire

L'objectif de ces différentes méthodes objectives d'évaluation de la distorsion non linéaire est de fournir une valeur numérique ou une représentation graphique des propriétés non linéaires du système testé [Voishvillo, 2011]. A partir de ces informations, il serait possible de juger de sa performance et de la comparer objectivement à celle d'autres systèmes. Cependant, il faudrait une énorme quantité de données pour obtenir une description précise du comportement non linéaire d'un système. Pour une mesure complète, il faudrait tester une multitude de combinaison de fréquences à des niveaux différents.

Pour estimer la distorsion non linéaire, la mesure de THD s'avère être la moins significative de toutes les méthodes objectives [Czerwinski *et al.*, 2001a; Voishvillo, 2011]. En effet, une

même valeur de THD peut être obtenue pour deux signaux dont les harmoniques sont distribués différemment. Si l'un des signaux contient des harmoniques d'ordre élevé et le deuxième des harmoniques d'ordre plus faible mais avec une distribution des niveaux identique, les distorsions engendrées seront bien différentes mais la valeur de THD peut être la même. La THD peut tout de même fournir des informations importantes sur le système testé si les valeurs de THD sont tracées en fonction de la fréquence et du niveau du signal d'entrée. Ces informations permettent d'avoir des indications rapides sur le comportement non linéaire général du système [Voishvillo *et al.*, 2004]. La mesure d'IMD avec deux fréquences ne permet pas d'obtenir toute l'information de distorsion du système testé [Voishvillo *et al.*, 2004]. Les fréquences pour lesquelles la distorsion est évaluée sont imposées par les fréquences des produits de modulation des deux fréquences du signal d'entrée. Les signaux composés d'un nombre faible de composantes fréquentielles sont finalement trop restrictifs.

Ces deux méthodes normalisées utilisées dans les fiches techniques d'équipement ne semblent donc pas adaptées à l'évaluation des systèmes complexes. Les autres méthodes d'évaluation présentées sont moins employées que la THD ou l'IMD pour l'estimation de la distorsion non linéaire sur les systèmes de restitution. De plus, les signaux mis en jeu dans ces types d'évaluation de la distorsion non linéaire sont des signaux stationnaires, contrairement à un signal musical essentiellement non stationnaire. Ainsi, le fait d'obtenir objectivement une valeur ou une représentation graphique ne permet pas de relier l'évaluation objective de la distorsion non linéaire avec la perception.

Ces différentes limitations révèlent trois aspects de l'évaluation de la distorsion non linéaire à prendre en considération. Tout d'abord, un système de restitution est un système dynamique extrêmement complexe qui regroupe des effets non linéaires électromagnétiques, mécaniques et acoustiques. Le deuxième point concerne le signal transmis au travers d'un système de restitution. Ce signal est, pour des conditions réelles d'écoute, loin des signaux composés de quelques fréquences. Il s'agit pour la plupart du temps d'un signal musical ou de parole caractérisé par des variations instantanées de niveau, d'enveloppe temporelle et de spectre. Le troisième et dernier aspect concerne la complexité du système auditif humain dont les propriétés physiologiques et psychoacoustiques ne sont pas linéaires. Nous allons développer ces trois points afin de comprendre la difficulté d'évaluer objectivement l'effet perceptif de la distorsion non linéaire d'un système de restitution.

1.3.1 Complexité du système de restitution

Dans un chaîne de restitution, la distorsion non linéaire peut provenir de différentes étapes de traitement du signal (CODEC) ou de l'amplificateur mais reste en général moins importante que la distorsion présente dans les enceintes acoustiques associées. De plus, les majorités des

études portant sur l'évaluation de la distorsion non linéaire se contentent d'étudier uniquement l'influence de l'enceinte sur la restitution. Une enceinte acoustique peut être considérée comme un système linéaire lorsque le niveau du signal d'entrée est compris dans une certaine gamme de dynamique. Pour des variations de dynamiques en dehors de cette gamme, le comportement des enceintes devient rapidement non linéaire. Ce caractère non linéaire apparaît majoritairement pour de fortes amplitudes mais il existe des phénomènes d'hystérésis apparaissant aux très bas niveaux. Dans le but de comprendre la distorsion qui intervient dans la restitution par des enceintes acoustiques, plusieurs éléments ont été identifiés comme non linéaires et susceptibles d'introduire de la distorsion. Ces non-linéarités sont dues principalement au déplacement de la bobine des haut-parleurs, aux propriétés géométriques des composants, et au flux d'air dans l'enceinte [Rauhala *et al.*, 2008].

1.3.1.a Le haut-parleur

Les haut-parleurs sont à l'origine de la majorité des non-linéarités présentes dans le comportement des enceintes acoustiques. Pour comprendre l'origine des distorsions, un court rappel est fait sur le fonctionnement d'un haut-parleur en se basant sur la figure 1.6. Le haut-parleur électrodynamique est un transducteur qui théoriquement permet le passage entre les domaines électrique-mécanique et mécanique-acoustique. L'aimant, le noyau et les plaques avant et arrière constituent le circuit magnétique qui imposent un champ magnétique dans l'entrefer. La bobine et la membrane constituent un équipage mobile suspendu dont le déplacement est guidé par les suspensions périphériques et le spider. Le courant qui traverse la bobine située dans l'entrefer génère une force qui entraîne la membrane, générant à son tour des ondes acoustiques.

Dans l'idéal, le haut-parleur devrait convertir linéairement un signal électrique provenant d'un amplificateur en un signal acoustique. Or, l'équipage mobile suspendu et le moteur électromécanique sont deux organes non linéaires agissant en cascade. Les principales non-linéarités rencontrées dans le fonctionnement des haut-parleurs électrodynamiques sont recensées ci-dessous. Pour davantage de détails, la plupart de ces phénomènes sont détaillées dans la revue faite par [Klippel, 2006].

Facteur de force : Le facteur de force Bl est caractérisé par la densité de l'induction magnétique B et la longueur de la bobine l . Ce facteur de force permet le couplage entre les domaines électrique et mécanique dans un haut-parleur. La valeur de Bl peut être considérée comme constante tant que la bobine se déplace légèrement dans l'entrefer autour de son équilibre initial. Pour des amplitudes de déplacement plus importantes la valeur de Bl varie en fonction du déplacement.

Suspension : Le système de suspension qui comprend le spider et la suspension périphérique permet de maintenir la bobine centrée dans l'entrefer et de guider le mouvement de la mem-

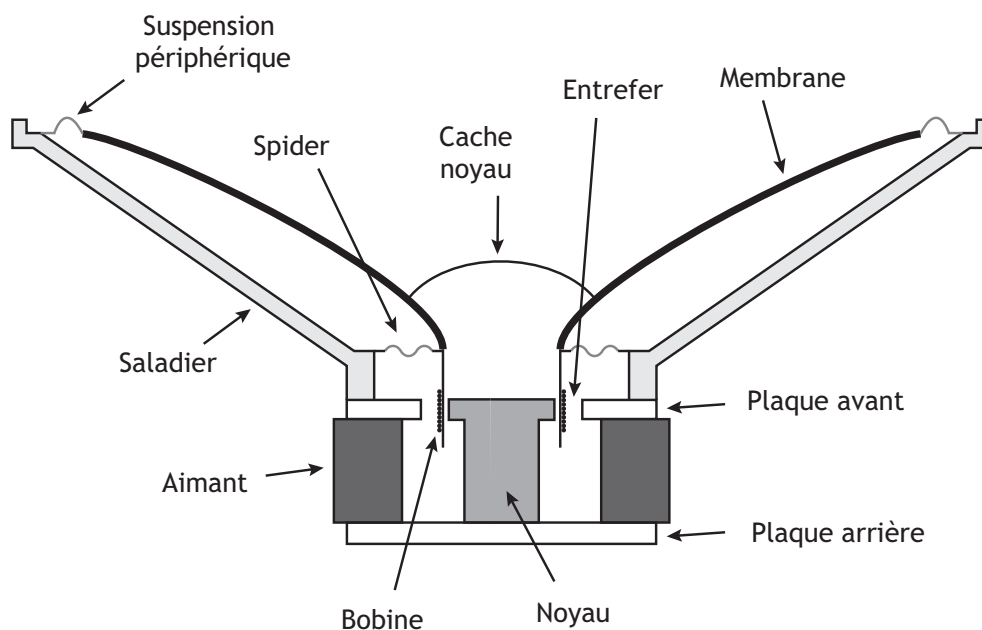


Figure 1.6 – Schéma d'un haut-parleur électrodynamique.

brane dans une seule direction afin d'éviter certains modes de vibration. Ces suspensions génèrent donc une raideur qui se comporte de manière non linéaire pour des amplitudes importantes de déplacement de la membrane.

Inductance : Le courant qui circule dans la bobine génère un champ magnétique. L'inductance L correspond à l'influence de ce champ magnétique sur le courant. Étant donné que ce champ magnétique dépend de la position de la bobine dans l'entrefer mais aussi de la valeur du courant, le comportement non linéaire de l'inductance est aussi lié à ces deux paramètres.

Résistance : La résistance de la bobine dépend de sa température : plus la température de la bobine augmente, plus sa résistance est importante. Cette température variant avec le niveau du signal d'entrée, la résistance présente donc un comportement non linéaire qui dépend essentiellement de l'enveloppe du signal utilisé. Ceci conduit au phénomène de compression thermique [Zuccatti et Bandiera, 2009].

Flexion de la membrane : Les résonances mécaniques de la membrane peuvent, en fonction de la fréquence et des matériaux utilisés, présenter un caractère non linéaire [Quaegebeur, 2007].

Effet Doppler : Une autre source de non-linéarité observée dans un haut-parleur est l'effet Doppler. Ce phénomène s'explique par le changement de distance entre la membrane du haut-parleur et le point d'écoute lorsque la vitesse de vibration est importante (en particulier pour des signaux basses fréquences).

1.3.1.b L'enceinte

Dans une enceinte, le haut-parleur est chargé par une raideur correspondant à l'air présent à l'intérieur de l'enceinte. Pour des pressions acoustiques importantes à l'intérieur des enceintes, la raideur de l'enceinte change avec le déplacement de la membrane et devient une source de non-linéarités [Zoltogorski, 1999]. La présence d'événements sur l'enceinte change aussi le comportement de l'air circulant dans l'enceinte. Pour des basses fréquences et de larges déplacements de la membrane, un écoulement souvent appelé "souffle" peut se produire [Shurer, 2007]. Cet écoulement non linéaire dépend des caractéristiques du signal et peut entraîner des effets audibles sur la reproduction du son.

1.3.1.c Défauts de fabrication

Lors de la production d'enceintes, certains défauts de fabrication peuvent être sources de distorsion audible. L'un des défauts souvent rencontré est le mauvais alignement de la bobine dans l'entrefer du haut-parleur. Ce décentrage peut entraîner un frottement entre la bobine et l'aimant. Une autre source de distorsion intervient lorsque, pour de fortes amplitudes, la bobine percute la plaque arrière du haut-parleur créant ainsi des impulsions à chaque choc. Il existe aussi des sources de distorsion liées à des problèmes d'étanchéité de l'enceinte, de collages défectueux entre plusieurs parties du haut-parleur, ou encore des problèmes d'éléments mobiles à l'intérieur du haut-parleur [Klippel, 2003; Temme *et al.*, 2009].

1.3.2 Complexité du signal

Pour les méthodes normalisées d'évaluation objective de la distorsion non linéaire (THD ou IMD), les signaux utilisés sont des signaux artificiels composés d'un ou deux signaux monofréquentiels. Les produits de distorsion mis en évidence par ces deux mesures sont limités par la nature du signal d'entrée utilisé. La mesure de MTND fait intervenir un signal multifréquentiel. L'avantage de cette méthode est la génération d'un signal d'entrée composé d'une multitude de fréquences permettant d'évaluer un grand nombre de combinaisons de produits d'intermodulation [Czerwinski *et al.*, 2001a]. Contrairement aux signaux utilisés pour la THD ou l'IMD, un signal multifréquentiel est plus proche d'un signal musical en terme de distribution statistique des amplitudes. Pour illustrer cette ressemblance, la figure 1.7 représente les distributions des amplitudes d'un signal monofréquentiel, d'un signal multifréquentiel et d'un signal musical.

La figure 1.7 montre que les amplitudes relatives d'un signal monofréquentiel sont majoritairement égales à -1 ou 1. La distribution du signal multifréquentiel suit une loi se rapprochant de la distribution du signal musical (qui reflète la présence de plusieurs échelles de niveaux). Bien que le signal multifréquentiel semble être une alternative aux méthodes de THD ou IMD, il reste stationnaire. Dans des conditions d'écoute réelle, les enceintes sont utilisées pour reproduire des

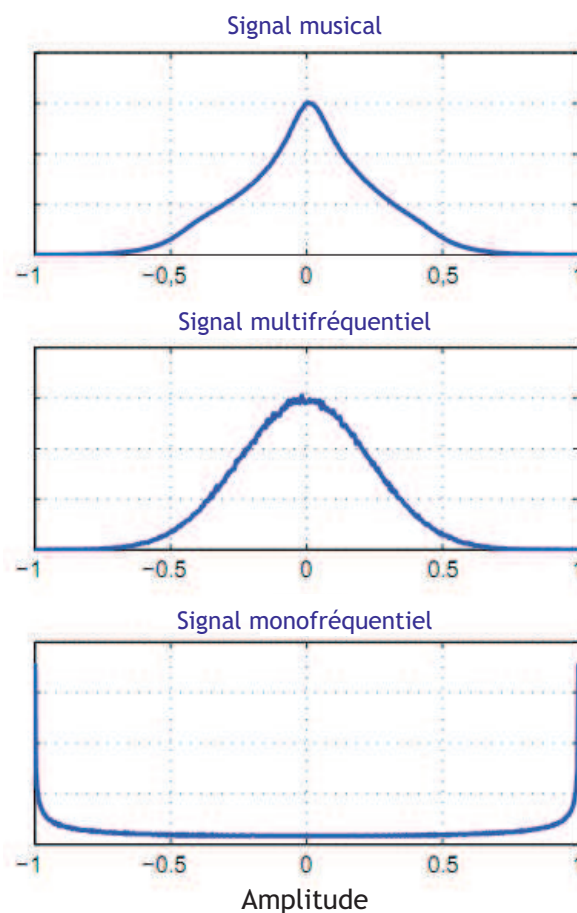


Figure 1.7 – Distributions des amplitudes d'un signal musical, multifréquentiel et d'un signal monofréquentiel (d'après Gunnarsson [2010]).

signaux musicaux qui sont non stationnaires. C'est cette caractéristique de non stationnarité qui rend l'évaluation objective des enceintes avec un signal musical difficile car l'amplitude du signal varie sans cesse d'un instant à l'autre. Or, la plupart des non-linéarités présentées précédemment sont dépendantes du niveau du signal utilisé. Les phénomènes de distorsion produits sont alors dépendants des non-linéarités du système considéré mais aussi du signal d'entrée, et plus particulièrement de son niveau. De ce fait, le contenu temporel et spectral ainsi que la dynamique du signal musical d'entrée ont une influence sur le comportement du système non linéaire considéré et sur les produits de distorsions introduits lors de sa restitution.

1.3.3 Complexité du système auditif

Le système auditif humain est bien loin d'être un simple analyseur de spectre. Pour comprendre comment sont perçus les phénomènes de distorsion, il est intéressant de rappeler brièvement différentes caractéristiques du système auditif. Les principales propriétés liées à la perception de la distorsion sont la non-linéarité de sensibilité de l'oreille et les effets de masquage

temporel et fréquentiel.

La sensibilité du système auditif est différente en fonction de la fréquence et de l'amplitude du signal. Les courbes d'isotonie pour des sons purs permettent de mettre en évidence le traitement non linéaire du système auditif. Comme montré dans la figure 1.8, chaque courbe d'isotonie relie les coordonnées (niveau de pression acoustique et fréquence) des sons purs qui sont perçus comme procurant la même sensation d'intensité [Moore *et al.*, 1997]. Ces courbes montrent que l'oreille humaine est moins sensible aux basses fréquences qu'aux hautes fréquences, en particulier à bas niveau.

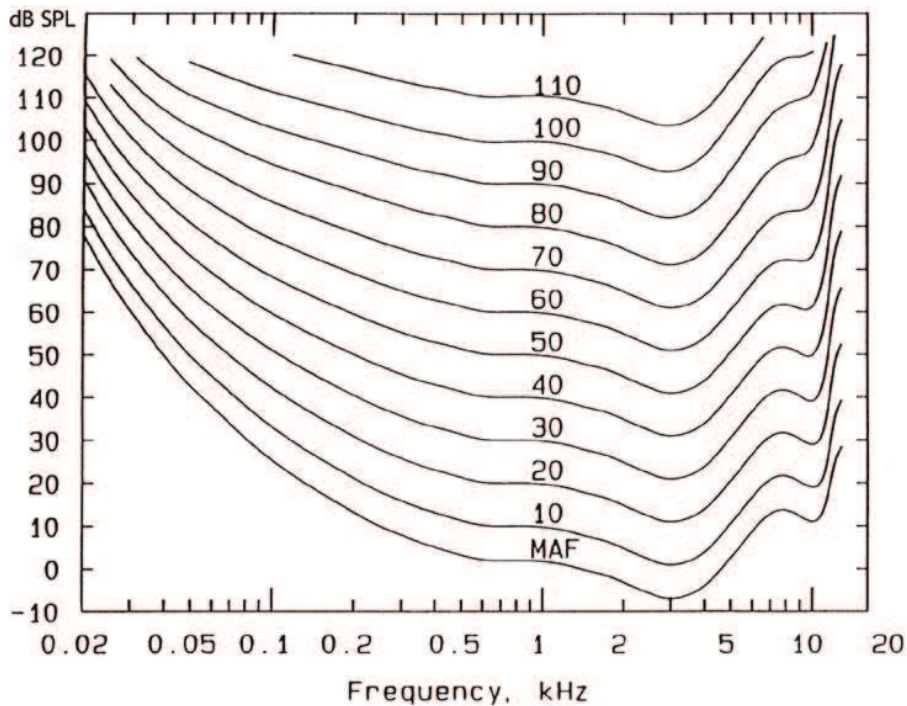


Figure 1.8 – Courbe d'isotonie représentée en fonction de la pression acoustique et de la fréquence (d'après Moore *et al.* [1997]). La courbe MAF (Minimum Audible Field) correspond au seuil d'audition.

Nous avons décrit les phénomènes non linéaires comme introduisant des fréquences supplémentaires dans le signal de sortie d'un système considéré. Le masquage auditif peut alors avoir un effet sur la perception de la distorsion générée par ces phénomènes. Le masquage est caractérisé par l'effet d'un signal (appelé masque) qui augmente les seuils de détection d'un autre signal (appelé signal ou cible). Dans cette partie, les propriétés du masquage sont rappelées brièvement. Pour plus de précisions, on peut se reporter aux ouvrages de Zwicker et Fastl [1999] ou de Moore [2003].

Une des caractéristiques du masquage est mise en évidence lorsque le masque et le signal ne sont pas présentés simultanément mais l'un après l'autre : on parle alors de masquage temporel.

Le masquage peut intervenir dans le domaine temporel sous la forme de masquage antérieur (signal avant le masque) et postérieur (signal après le masque). La figure 1.9 illustre ce phénomène de masquage et montre que la quantité de masquage antérieur est inférieure à la quantité de masquage postérieur. La quantité de masquage correspond à la différence entre le seuil de détection du signal en condition de masquage et son seuil dans le silence.

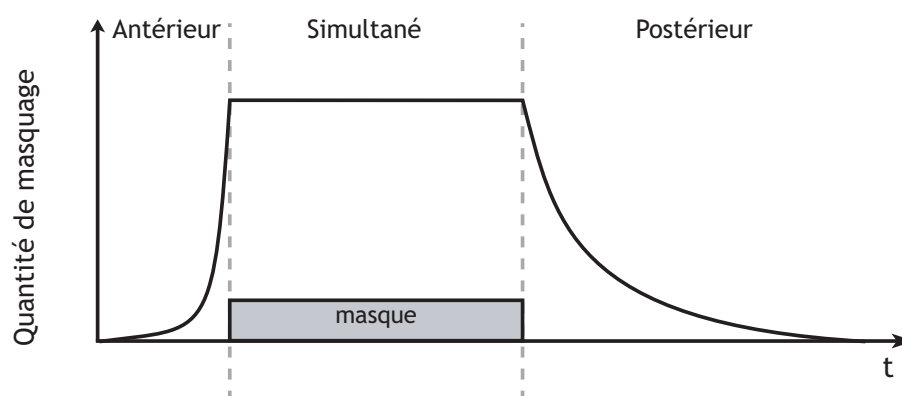


Figure 1.9 – Principe du masquage temporel.

Lorsque le masque et le signal sont présentés simultanément, on parle alors de masquage simultané ou fréquentiel. Le masque crée une sorte de "zone d'ombre" dans laquelle les signaux ayant un spectre tombant dans cette zone et un niveau inférieur sont masqués. Lorsque le masque est un son sinusoïdal, la zone de masquage est triangulaire et symétrique autour de la fréquence du masque. En augmentant le niveau du masque, cette forme devient asymétrique avec une partie plus longue s'étirant vers les hautes fréquences (Figure 1.10). Plus le niveau du masque est fort, plus la zone de masquage est asymétrique et s'étend sur une région fréquentielle importante. Cette propriété entraîne un effet de masquage plus important pour des fréquences supérieures à la fréquence du masque que pour celles qui sont inférieures à la fréquence du masque.

Dans le cas de distorsion harmonique générée par un signal composé d'une seule fréquence, les harmoniques générées apparaissent à des fréquences supérieures à celle du signal. Les harmoniques d'ordre les plus bas auront tendance à être davantage masquées que les harmoniques d'ordres supérieurs. La figure 1.11 montre les effets de masquage sur l'audibilité des produits de distorsion générés par un signal composé d'une fréquence. D'après les courbes de masquage, l'harmonique H_2 est masquée alors que H_3 , plus haute en fréquence, est audible. Cette figure montre que des produits de distorsion d'ordre élevé peuvent être plus audibles que certains d'ordre plus faible.

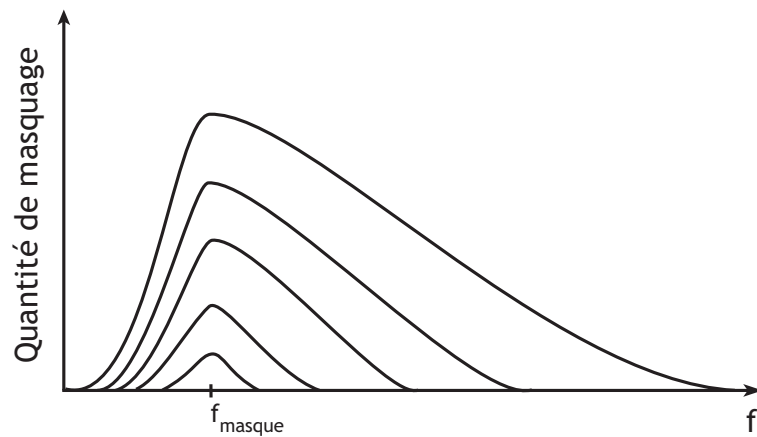


Figure 1.10 – Courbes de masquage fréquentiel schématisées pour des sons purs à plusieurs niveaux.

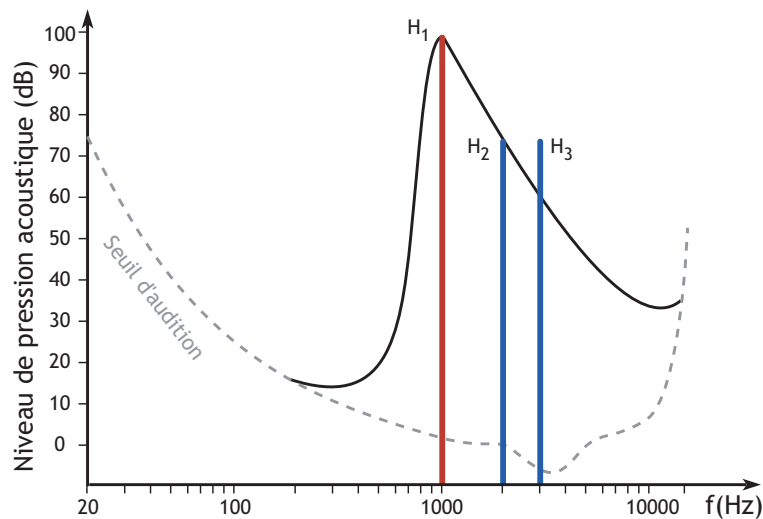


Figure 1.11 – Réponse d'un système non linéaire à un signal composé d'une fréquence H_1 et courbe de masquage associée.

1.4 Évaluation perceptive de la distorsion non linéaire

Les méthodes objectives conventionnelles d'évaluation de la distorsion non linéaire permettent d'obtenir des informations qui caractérisent le système non linéaire considéré. Cette évaluation n'est pourtant pas exhaustive et aucune des méthodes ne permet une évaluation complète de l'effet de la distorsion engendrée par le système. Dans le cas de l'évaluation de l'effet de la distorsion non linéaire des enceintes acoustiques dans des conditions d'utilisation usuelles, cette limitation est due à l'utilisation de signaux majoritairement stationnaires et à la complexité des phénomènes non linéaires présents dans la restitution de l'enceinte.

A partir des différentes limitations de l'évaluation objective, plusieurs études se sont intéressées à évaluer l'effet perceptif de différents mécanismes de distorsion non linéaire. Les différentes

études, présentées dans cette section, ont cherché à relier les mesures perceptives avec les évaluations objectives connues.

L'évaluation perceptive de la distorsion non linéaire rentre dans le cadre de l'évaluation d'un système de reproduction. Nous avons vu qu'il est possible de se concentrer sur le système de reproduction ou de se focaliser plus particulièrement sur le signal reproduit. En effet, une partie des études se sont intéressées au système de reproduction en cherchant à caractériser comment sont perçus les phénomènes non linéaires spécifiques de paramètres connus et mesurés sur des haut-parleurs. D'autres études ont cherché à caractériser les effets de la distorsion en introduisant directement des composantes dans le signal ou en utilisant des modèles de distorsion sans lien direct avec le système de reproduction.

Nous avons ainsi regroupé les différentes études perceptives portant sur l'évaluation de la distorsion suivant ces deux approches. Les principaux résultats de corrélation entre l'évaluation perceptive et objective sont présentés dans la section 1.5.

1.4.1 Caractérisation des systèmes de restitution :

Les travaux de Schmitt [1995] concernent l'étude de différentes non-linéarités spécifiques aux haut-parleurs. Elle a utilisé un modèle de haut-parleur pour pouvoir modifier l'influence de la force électrodynamique, la raideur ou encore l'amortissement. Elle a alors mesuré les seuils de détection de l'effet de différents modification des paramètres dans les cas d'un son pur, d'un accord, et de trois types de musique. Outre les tests où des produits de distorsion harmonique, d'interférence et d'intermodulation étaient ajoutés au signal à tester, Boer *et al.* [1998] ont utilisé un modèle de haut-parleur pour évaluer la distorsion non linéaire. Pour générer de la distorsion, l'influence du facteur de force, de la raideur et de l'inductance pouvait être modifiée. Ils ont ainsi mesuré les seuils de détection pour deux types de musique, jazz et pop. Kristoffersen *et al.* [1999] ont employé un modèle de haut-parleur pour évaluer les distorsions d'ordre 2 et 3 liées au déplacement de la membrane. Ils ont combiné ces distorsions avec différents nombres de haut-parleurs simulés dans deux pièces différentes. Des tests d'écoute ont permis d'évaluer la préférence entre différentes configurations de haut-parleurs en fonction de la pièce. Klippel [2001] a proposé une méthode permettant d'écouter en temps réel l'influence de différents types de non-linéarités qu'il a auparavant mesurées pour des basses fréquences. Les paramètres non linéaires du modèle sont le facteur de force, la raideur de la suspension et l'inductance. La méthode offre la possibilité d'écouter la partie linéaire du signal, le signal distordu par le modèle et aussi la partie distordue seule pour n'importe quel signal d'entrée. Le test d'écoute proposé est une mesure de seuil de détection des différentes non-linéarités pouvant être introduites indépendamment ou simultanément dans le signal test.

1.4.2 Caractérisation de la perception de la distorsion non linéaire

Evaluation de produits de distorsion :

Certaines études se sont intéressées à évaluer indépendamment différents effets de la distorsion non linéaire tels que la distorsion harmonique ou la distorsion d'intermodulation. Généralement, les tests d'écoute consistent à mesurer des seuils de détection. La distorsion qui altère le signal testé est diminuée jusqu'à ce que les auditeurs ne puissent plus détecter la présence de la distorsion.

L'une des toutes premières études effectuée par Bryan et Parbrook [1960] porte sur la mesure des seuils d'audibilité de la distorsion harmonique sur un signal sinusoïdal de fréquence 357Hz présenté à plusieurs niveaux. Lors de cette étude, les harmoniques d'ordre 2 à 8 étaient ajoutées au signal original et les seuils d'audibilité du signal accompagné de ces harmoniques ont ainsi été obtenus pour différentes conditions (niveau / nombre d'harmoniques). Dans une étude similaire, Gabrielson et Sjögren [1972] ont comparé les seuils d'audibilité de la distorsion harmonique pour un son pur à différentes fréquences mais pour des enregistrements de flûte et de clarinette en régime stationnaire. Au cours de cette étude, les seuils ont été mesurés uniquement pour les harmoniques 2 et 3 introduites dans les stimuli à évaluer. De même, Fielder et Benjamin [1988] se sont intéressés à la mesure de seuils de détection de la distorsion harmonique d'ordre 3 introduite à partir de techniques de traitement de signal numérique (DSP) dans un extrait de musique classique.

Fryer [1975] s'est intéressé à la distorsion d'intermodulation indépendamment d'autres types de distorsion. Il contrôlait la quantité de produits de distorsion d'intermodulation du premier ordre ajouté au signal. Il a ainsi mesuré les seuils de détection sur trois types de musique : pop, classique et piano seul. Petri-Larmi *et al.* [1980] ont aussi étudié la distorsion d'intermodulation. Tout comme Fryer [1975], ils ont associé un générateur de distorsion d'intermodulation et un haut-parleur. Les seuils de détection ont ainsi été obtenus pour des extraits de chorale, de piano, de violon, de clavecin et de musique pop.

Enfin, Boer *et al.* [1998] ont évalué la préférence et ont ensuite mesuré les seuils de détection de trois types de distorsion sur des extraits de musique jazz, pop et classique. Ils ont étudié la distorsion harmonique, la distorsion d'intermodulation et la distorsion d'interférence correspondant aux produits d'intermodulation du signal testé avec ce même signal filtré entre 20 et 1000Hz. Pour la distorsion harmonique, ils ont ajouté les harmoniques 2 et 3 de la fréquence 100 Hz au signal de référence. Pour la distorsion d'intermodulation, ils contrôlaient les produits d'intermodulation d'ordre 2 avec un signal situé entre 20 et 100Hz et un deuxième de fréquence plus haute.

Evaluation de distorsions artificielles :

Au lieu de s'intéresser précisément à la perception de la distorsion harmonique ou d'intermodulation, d'autres études ont cherché à évaluer perceptivement les effets d'objets artificiels non linéaires. Les études qui ont utilisé ce moyen d'introduire de la distorsion sont basées sur la simulation du comportement de différents systèmes non linéaires à partir de différentes formes de courbes entrée/sortie.

L'étude menée par Belcher [1978] concernait l'évaluation de la dégradation engendrée par deux types de relation entrée/sortie sur des signaux de voix masculine et de piano seul. La première relation était dite de forme en "S" et la deuxième représentait le phénomène de Clipping. Belcher a ainsi évalué la qualité perçue de ces deux non-linéarités en mesurant des jugements sur une échelle dont les extrémités étaient "imperceptibles" et "inexploitables".

Fryer et Millward [1980] se sont aussi intéressés au phénomène de Clipping et ont obtenu des seuils de détection de la distorsion engendrée sur de la musique disco et classique. En associant un amplificateur synthétisé et un haut-parleur, ils contrôlaient le pourcentage de temps où le Clipping intervenait dans le signal musical. L'étude de Karjalainen [1982] a aussi permis d'évaluer plusieurs phénomènes représentés par différentes courbes d'entrée/sortie. Il a généré des non-linéarités de Zero Crossing et de Clipping, des non-linéarités issues de polynômes de puissance 2 et 3 et une courbe entrée/sortie présentant un angle. Il a obtenu les seuils de détection sur des signaux de parole (voyelles suédoises prononcées par un homme) puis altérées par les différents types de non-linéarités.

Plus récemment, Tan *et al.* [2003] ont publié une étude dans laquelle ils se sont intéressés à différents types de non-linéarité et à leurs influences dans plusieurs régions fréquentielles. Leurs expériences concernent l'évaluation de la distorsion non linéaire introduite sur de la musique jazz mais aussi sur un signal de parole. Dans une première expérience, ils ont produit des non-linéarités de Clipping obtenues à partir de courbes entrée/sortie symétriques et asymétriques, de la non-linéarité de Zero Crossing, et une modification du niveau global du signal. Toutes ces non-linéarités étaient appliquées sur la totalité du spectre des signaux testés. Dans une deuxième expérience, seules les non-linéarités de Clipping et la modification du niveau global étaient appliquées, en particulier dans des bandes fréquentielles spécifiques. Une troisième expérience comprenait des stimuli artificiels, obtenus à partir de courbes entrée/sortie et des stimuli provenant d'enregistrements réels de haut-parleurs. Pour les trois expériences, les jugements de qualité ont été effectués à partir d'une échelle allant de "pas distordu" à "très distordu". L'étude menée par Geddes et Lee [2003a,b] a fait intervenir quatre formes de non-linéarités différentes. Une forme était basée sur la série de Taylor, une autre était la non-linéarité de Zero Crossing et deux autres étaient issues de séries de Fourier faisant intervenir des fonctions sinus et cosinus. Ils ont ainsi mesuré la qualité perçue des distorsions générées par ces 4 courbes entrée/sortie

sur un extrait de musique composé d'un orchestre et d'un chanteur. La méthode d'évaluation consistait à comparer les signaux modifiés avec les non-linéarités au signal original de référence en mesurant les jugements sur une échelle allant de "meilleur que la référence" à "intolérable".

1.5 Corrélations entre évaluations perceptives et objectives

Les mesures perceptives de l'effet de la distorsion non linéaire ont permis dans un premier temps d'obtenir des données concernant la détection ou la préférence de différents types de distorsion. A partir de données perceptives sur la distorsion, les études se sont intéressées à chercher les mesures objectives les plus pertinentes d'un point de vue psychoacoustique. Ces études ont d'abord permis de montrer les limites des mesures de distorsion conventionnelles quant à la perception des effets de la distorsion non linéaire. Dans un deuxième temps, des études plus récentes ont permis d'améliorer les mesures objectives conventionnelles en prenant en compte la complexité du signal musical et en intégrant des propriétés du système auditif humain.

La mesure de THD est la mesure qui reflète le moins la perception de la distorsion non linéaire [Belcher, 1978]. La mesure d'IMD permet d'obtenir de meilleures corrélations avec des données perceptives que celle obtenues avec la THD [Belcher, 1978]. Quoiqu'il en soit, la distorsion engendrée par une ou deux fréquences est bien différente de celle engendrée par un signal musical et rend donc les mesures de THD et d'IMD peu pertinentes d'un point de vue perceptif [Tan *et al.*, 2003]. Le signal multifréquentiel semble être une mesure alternative au THD et IMD mais, aucune étude ne compare des données perceptives avec des valeurs de MTND obtenues à partir d'un signal multifréquentiel.

Voishvillo *et al.* [2004] distinguent deux types de métriques objectives qui prennent en compte les propriétés du système auditif. Les métriques dites "semi-perceptives" sont basées sur des principes psychoacoustiques admis. Shorter [1950] a pondéré des harmoniques supérieures en $n/2$ et $n^2/4$ avec n l'ordre des harmoniques. Ces pondérations permettaient de prendre en compte l'audibilité plus importante des harmoniques d'ordres élevés. Geddes et Lee [2003b] ont proposé la métrique GedLee qui est basée sur deux principes psychoacoustiques. Le premier principe, identique à celui proposé par Shorter [1950], prend en compte le fait que les harmoniques d'ordres élevés sont plus audibles que les harmoniques d'ordre plus faible qui ont tendance à être masquées. Le deuxième principe s'appuie sur le fait que les produits de distorsion sont plus audibles pour des signaux de faibles niveaux étant donné que les seuils de masquage sont plus faibles à bas niveaux. La comparaison de la métrique GedLee avec la THD et l'IMD montre qu'elle est davantage corrélée à la qualité perçue pour des échantillons de musique.

Les métriques objectives dites "perceptives" sont basées sur des données psychoacoustiques telles que le masquage ou sur les modèles physiologiques du système auditif. Les méthodes d'éva-

luation de la qualité d'un signal musical (PEAQ [ITU BS.1387-1, 2001]) ou d'un signal de parole (PESQ [ITU-T P.862, 2001]) prennent en compte les propriétés du système auditif. Pour l'évaluation de la distorsion, Tan *et al.* [2003, 2004] ont développé deux métriques basées sur des principes psychoacoustiques. La première, la métrique DS pour Distortion Score [Tan *et al.*, 2003] est obtenue en appliquant au système non linéaire un signal multifréquentiel. Les signaux sont analysés sur des intervalles de temps de 30 ms et la valeur de la métrique DS est obtenue en additionnant les niveaux des produits de distorsion au sein de chaque filtre auditif de l'oreille interne (ERB). Cette métrique a permis d'obtenir des bonnes corrélations avec la qualité perçue sur des signaux musicaux distordus à partir de non-linéarités artificielles (courbes entrée/sortie). Lors d'une expérience menée sur des non-linéarités artificielles et les non-linéarités engendrées par de vrais transducteurs, la métrique DS était moins efficace. Ils ont alors proposé une deuxième métrique basée sur un modèle perceptif plus sophistiqué qui utilise l'analyse des signaux musicaux et de parole à la place du signal multifréquentiel [Tan *et al.*, 2004]. Cette métrique, R_{nonlin} , est aussi basée sur les mêmes intervalles de temps et analyse en ERB mais inclut un filtrage supplémentaire qui simule la réponse fréquentielle de l'oreille externe et moyenne. L'utilisation de cette métrique permet d'obtenir des corrélations élevées avec la qualité perçue des signaux distordus de manière artificielle et ceux obtenus à partir de vrais transducteurs.

1.6 Bilan de l'étude de la distorsion non linéaire

Les différentes études présentées dans ce chapitre se sont intéressées à relier des évaluations perceptives aux mesures objectives conventionnelles ou améliorées. La tâche est complexe et demande de prendre en compte les contraintes liées à la restitution d'un signal à travers un système non linéaire et les propriétés de l'oreille. Pour faire un bilan, nous avons retenu trois points qui mettent en avant les efforts faits pour évaluer d'un point de vue perceptif la distorsion non linéaire : l'objectif de l'étude (système de reproduction ou signal reproduit), les signaux utilisés aux cours des tests, les méthodes d'évaluation perceptive. Ces trois points rejoignent les trois points qui constituaient les limitations des méthodes objectives (section 1.3).

Nature des signaux utilisés :

A part quelques études [Bryan et Parbrook, 1960; Gabrielson et Sjögren, 1972], toutes les études sur l'évaluation perceptive de l'effet de la distorsion non linéaire font intervenir des signaux non stationnaires. Certaines études se sont intéressées à des signaux de parole et une majorité d'entre elles portent sur l'évaluation de signaux musicaux. L'utilisation de signaux normalement retransmis par les systèmes de restitution permet de se rapprocher des conditions réelles de fonctionnement.

Méthode d'évaluation perceptive :

Les différents tests perceptifs effectués sur l'effet de la distorsion non linéaire font intervenir principalement deux méthodes. D'un côté, des mesures de seuils ont souvent été effectuées pour estimer l'audibilité de l'effet de la distorsion engendrée par différents systèmes non linéaires. Pour ces mesures, le niveau de la distorsion est diminué jusqu'à ce que l'auditeur ne puisse plus percevoir l'effet de la distorsion. D'autres études font intervenir des mesures de préférence ou des jugements de qualité en évaluant chaque échantillon distordu sur une échelle graduée. Les qualificatifs utilisés aux extrémités de l'échelle sont déterminés par les expérimentateurs en fonction de la tâche demandée à l'auditeur (évaluation absolue ou comparaison à un son de référence).

Système de reproduction ou signal reproduit :

Avec l'amélioration des modèles, il est possible de reproduire le comportement du haut-parleur. Les études de la perception des effets de distorsion non linéaire ont utilisé de tels modèles pour introduire des phénomènes non linéaires identifiés et mesurés. L'utilisation d'un modèle paramétrique de haut-parleur permet d'introduire et de modifier l'influence de plusieurs paramètres simultanément ou indépendamment. Les résultats obtenus dans ces études permettent d'estimer comment sont perçus les phénomènes identifiés et d'obtenir des informations quant à leur audibilité et leur gêne perçue pour des signaux de musiques. Cette démarche permet la compréhension de l'effet des mécanismes physiques dans le but de contrôler et ainsi d'améliorer les systèmes de restitution.

Une deuxième partie des études s'est intéressée à la perception du signal reproduit. Certaines ont étudié directement le résultat de la distorsion en ajoutant des harmoniques ou des produits d'intermodulations. Les produits de distorsion sont introduits arbitrairement dans le signal à évaluer, et ce sont la plupart du temps les ordres deux ou trois qui sont évalués. D'autres études ont caractérisé le signal modifié par les systèmes non linéaires obtenus à partir de courbe de comportement entrée/sortie. Cette approche permet de simuler des comportements basés sur des lois mathématiques faciles à paramétrer, modélisant approximativement des mécanismes physiques connus. Elle permet aussi de s'écarter de ce que serait le fonctionnement normal d'un système de restitution et ainsi de générer une plus grande variété d'échantillons sonores distordus. L'étude du signal reproduit est plus orientée vers l'étude de la perception. En cherchant à caractériser comment sont perçus différents signaux musicaux modifiés de diverses manières, ces études ont permis d'apporter des connaissances sur la perception des phénomènes de distorsion et ainsi d'améliorer les métriques objectives pour quantifier l'audibilité et la qualité perçue au sens "bon/mauvais" (PEAQ, GedLee, DS, R_{nonlin}).

Les études présentées dans ce chapitre se sont intéressées soit à caractériser l'objet non linéaire soit à comprendre comment est perçu l'effet de la distorsion non linéaire mais avec une approche basée sur des jugements de préférence ou des mesures de seuils. Pour l'évaluation de la distorsion non linéaire, l'approche que nous avons choisie est proche des études sur la perception du signal reproduit et complète les études sur la caractérisation de la perception de la distorsion. Notre approche et les choix que nous avons effectués sont présentés dans le chapitre suivant.

Chapitre 2

Approche, choix et protocole proposé

Sommaire

2.1	Caractérisation multidimensionnelle de signaux musicaux distordus	33
2.1.1	Application à la reproduction sonore	34
2.1.2	Besoins liés à l'étude multidimensionnelle de la distorsion non linéaire .	35
2.2	Élaboration d'un corpus de stimuli distordus	38
2.2.1	Modélisation de haut-parleurs	39
2.2.2	Contrôle de la non-linéarité	41
2.2.3	Restitution	42
2.3	Recherche d'une méthode d'évaluation de la dissemblance	43
2.3.1	Méthodes usuelles pour l'évaluation de la dissemblance	43
2.3.2	Méthodes alternatives adaptées à l'évaluation de corpus étendus	47
2.3.3	Méthode retenue	50
2.4	Protocole proposé	51

Dans ce chapitre nous développons l'approche que nous avons choisi pour caractériser perceptivement des effets de la distorsion non linéaire. Ce choix implique certaines contraintes par rapport au corpus d'échantillons que nous allons tester mais aussi par rapport à la méthode de caractérisation perceptive que nous allons employer. Au cours de ce chapitre nous expliquons les besoins et les choix effectués pour mettre en place notre protocole. Le bilan présenté à la fin de ce chapitre résume le protocole que nous proposons pour la caractérisation des non-linéarités des systèmes de reproduction. Au cours de ce chapitre, lorsque nous utilisons le terme "caractérisation", il se réfère à la caractérisation perceptive des signaux.

2.1 Caractérisation multidimensionnelle de signaux musicaux distordus

Pour la caractérisation perceptive de la distorsion, au cours du chapitre 1, nous avons présenté les deux méthodes principalement employées. Certaines études emploient des mesures de seuils permettant d'obtenir une information sur l'audibilité de différents types de distorsion non linéaire. D'autres études présentées se sont intéressées à évaluer la qualité perçue en termes de préférence. Ces deux types de jugements spécifiques de la distorsion non linéaire ne donnent qu'une information limitée de la distorsion non linéaire. Ils ne fournissent aucune information quant aux critères sur lesquels les auditeurs ont basé leurs jugements [McDermott, 1969]. Deux extraits distordus de manière différente pourraient être jugés de qualité identique.

Lors de l'étude sur l'audibilité de la distorsion non linéaire dans la reproduction des haut-parleurs, Schmitt [1995] a dans un premier temps mesuré des seuils d'audibilité pour différents extraits musicaux modifiés à partir d'un modèle de haut-parleur contrôlable. A la suite de ce test, un entretien avec les auditeurs a permis de révéler quelles étaient les modifications qu'ils entendaient au cours du test. Schmitt [1995] a ainsi élaboré une liste de ces modifications :

- Coloration du son : La distorsion non linéaire peut être perçue comme une coloration du son. La sensation de coloration d'un son est liée à la modification de la réponse fréquentielle, comme la distorsion linéaire.
- Modification de hauteur : Ces changements de hauteurs étaient perçus pour les passages de musique avec des instruments à percussion.
- Modification de la dynamique : La génération de distorsion non linéaire entraîne une modification de la dynamique et plus particulièrement de la compression dynamique.
- Modification de la structure temporelle : Cet effet apparaît pour les coups d'instruments à percussion aux basses fréquences et se caractérise par la modification de leur durée.
- Modification de la sensation d'espace : Cet effet était relié aux changements de hauteur.
- Bruit additionnel : Certains artefacts n'étaient pas perçus comme faisant partie du signal à évaluer et étaient ainsi considérés comme du bruit additionnel.

Ces éléments de modification montrent le caractère multidimensionnel associé à l'évaluation de l'effet de la distorsion non linéaire. Les systèmes non linéaires entraînent des modifications temporelles et spectrales du signal d'entrée. Il est possible d'identifier les critères utilisés par les auditeurs pour différencier des stimuli sonores en utilisant une technique de MDS (MultiDimensional Scaling) appliquée à des jugements de dissemblances. Pour les dissemblances perceptives, l'analyse MDS fournit un espace perceptif dont les dimensions sont les critères principaux utilisés par les auditeurs pour effectuer leurs jugements de dissemblances. C'est dans cette optique qu'Atsushi et Martens [2002] se sont intéressés à trouver les dimensions sous-jacentes aux processus d'élaboration d'effet de distorsion pour guitare électrique. Leur étude consiste à trouver une sorte d'étalonnage perceptif des effets de distorsion pour permettre de relier au mieux les algorithmes de génération de distorsion à la perception. A partir de mesures de dissemblances, ils ont trouvé deux dimensions sur lesquelles les auditeurs se basaient pour différencier différents processus de distorsion. Cette étude illustre ainsi l'intérêt de mesurer des dissemblances perçues pour pouvoir caractériser les phénomènes de distorsion.

2.1.1 Application à la reproduction sonore

Dans le but d'étudier les effets de la distorsion non linéaire, nous avons choisi d'appliquer cette approche multidimensionnelle à la caractérisation de système de reproduction particulier : les enceintes acoustiques. Nos travaux sont dans la lignée de ceux effectués par Lavandier [2005] sur la caractérisation du timbre restitué par les enceintes acoustiques. Lavandier [2005] a choisi l'approche multidimensionnelle pour évaluer les différences de timbre entre des enceintes acoustiques. A partir de l'enregistrement de plusieurs enceintes, le protocole qu'il a proposé est basé sur deux démarches, physique et perceptive, menées en parallèle. La figure 2.1 présente les deux démarches du protocole menées à partir de l'enregistrement d'une enceinte.

Le point important du protocole est que les signaux enregistrés sont identiques pour les analyses physiques et les tests d'écoute. Les mesures perceptives consistent en l'évaluation des dissemblances perçues par des auditeurs entre les différentes enceintes. Les mesures physiques permettent d'obtenir des critères objectifs de dissemblance entre les signaux d'enceintes. Plusieurs métriques objectives ont été employées permettant des mesures dans les domaines temporels et fréquentiels et d'autres faisant intervenir des propriétés du système auditif. Le but du protocole proposé est de fournir une méthode de mesure objective qui différencie les enregistrements de la même manière que les auditeurs. Ainsi, ce protocole vise à proposer pour la reproduction du timbre une métrique objective pertinente d'un point de vue perceptif. La validation de cette pertinence est effectuée en deux temps. La corrélation entre les deux données de dissemblances (objectives et perceptives) est calculée dans un premier temps. Cette corrélation montre que la métrique objective fournit des valeurs de dissemblances reliées linéairement à celles obtenues

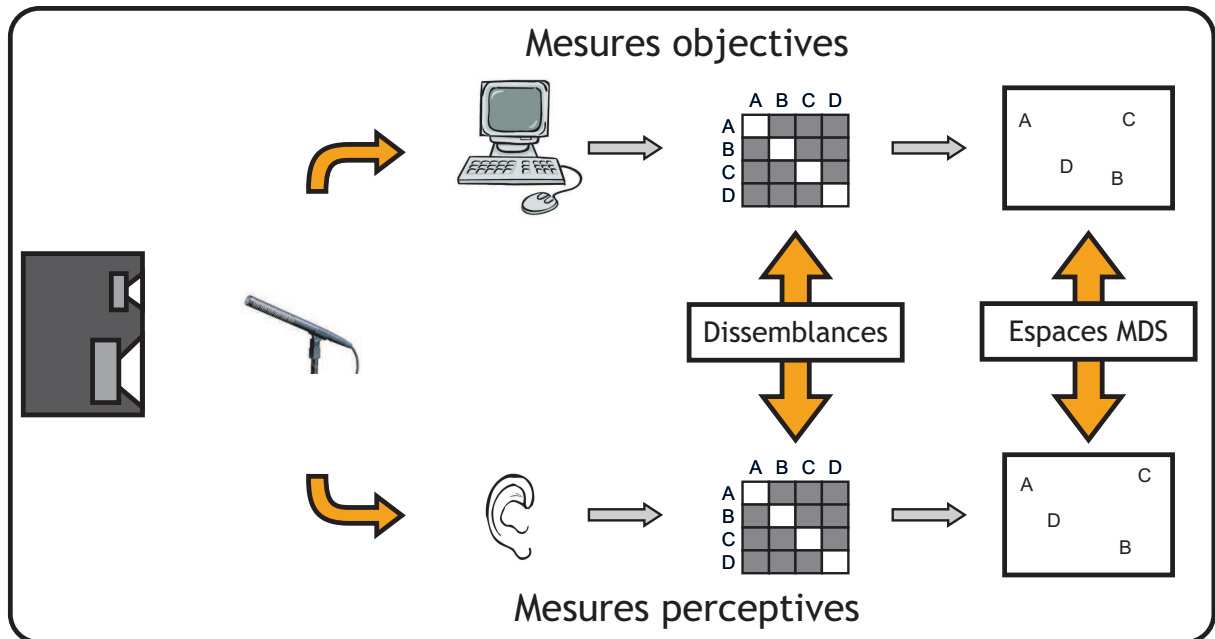


Figure 2.1 – Protocole d’évaluation physique et perceptive des enceintes acoustiques proposé par Lavandier [2005].

au cours des tests d’écoute [Lavandier *et al.*, 2008a]. Une deuxième étape consiste à analyser les dissemblances objectives et perceptives avec une technique d’analyse multidimensionnelle (MDS). L’analyse MDS sur les données objectives permet de comparer l’espace objectif avec l’espace perceptif toujours dans le but de vérifier que la méthode objective est pertinente vis-à-vis de la perception [Lavandier *et al.*, 2008b].

La manière dont Lavandier [2005] a caractérisé la restitution du timbre est adaptée à notre problématique de caractérisation des effets de la distorsion non linéaire. Nous étendons ses travaux à la prise en compte des modifications liées à la distorsion non linéaire dans la reproduction d’un signal musical.

2.1.2 Besoins liés à l’étude multidimensionnelle de la distorsion non linéaire

L’approche multidimensionnelle répondant bien à notre but d’évaluer les différentes caractéristiques de la distorsion, nous avons décidé d’adapter le protocole proposé par Lavandier [2005]. Néanmoins, l’étude de la reproduction sonore telle qu’elle a été menée par Lavandier [2005] nécessite d’être adaptée à notre problématique d’évaluation de la distorsion non linéaire dans la reproduction d’un signal musical. Dans l’étude que nous proposons, nous nous intéressons donc aux aspects méthodologiques afin de proposer un protocole permettant de caractériser perceptivement la distorsion non linéaire. Nous cherchons à caractériser la distorsion non linéaire dans une situation d’écoute réaliste avec une méthode d’évaluation adaptée. Nous développons l’aspect

d'écoute réaliste puis celui de la méthode de tests d'écoute pour mettre en évidence les besoins liés à l'élaboration de notre protocole.

2.1.2.a Réalisme des échantillons distordus

Pour caractériser les effets de la distorsion non linéaire dans un contexte réaliste, nous cherchons à mener des tests d'écoute pendant lesquels l'auditeur se trouve dans une situation d'écoute de musique reproduite. Pour cela, il est nécessaire de faire intervenir des mécanismes réels correspondant à un contexte d'écoute d'extraits musicaux reproduits. Pour respecter ce contexte écologique d'écoute et pour pouvoir contrôler la distorsion, l'utilisation d'enceintes acoustiques apparaît comme un besoin pour l'élaboration d'échantillons sonores.

Le fait d'utiliser des objets comme les enceintes pour les tests d'écoute conduit à des contraintes très spécifiques que nous allons tenter de gérer. De manière générale, les tests d'écoute pour l'étude de système de reproduction nécessitent d'être contrôlés attentivement. Toole [1982] répertorie les différents paramètres expérimentaux qui peuvent influencer les jugements des auditeurs, également présentés dans la norme [IEC 60268-13, 1998] : la salle d'écoute, la position des enceintes et de l'auditeur, le mode de reproduction des enceintes (monophonique ou stéréophonique), le niveau sonore, l'expérience de l'auditeur, les extraits musicaux et enfin la procédure expérimentale. Dans l'élaboration des échantillons distordus, il est nécessaire de respecter ces contraintes pour s'assurer que les jugements des auditeurs ne soient pas biaisés.

Nous cherchons à caractériser les effets de la distorsion dans des conditions d'écoute naturelle. Bose [1968] dit "We have no satisfactory correlation between any of these loudspeakers distortion measurements and their subjective effects on music reproduction and we realize the need for a method of detecting distortion in a way that is meaningful to the listener in the environment for which the speaker is intended ". Bose [1968] illustre le fait qu'il est préférable pour des tests d'écoute sur la distorsion d'évaluer les enceintes dans des conditions qui aient du sens pour l'auditeur. Même si des enregistrements dans un environnement anéchoïque s'avèreraient plus précis pour évaluer la distorsion, nous avons besoin d'évaluer les effets de la distorsion perçus dans un environnement d'écoute "normal" ou au moins réaliste.

2.1.2.b Analyse multidimensionnelle

Nous cherchons à mettre en évidence le caractère multidimensionnel de la distorsion introduite dans la restitution sonore à partir de jugements de dissemblances. Pour évaluer certaines caractéristiques perceptives des systèmes de reproduction en s'affranchissant d'évaluations absolues ou de jugements de qualité, Lipshitz et Vanderkooy [1981] recommandent d'évaluer les différences relatives entre enceintes. Il convient alors d'effectuer des tests permettant d'évaluer la dissemblance perçue entre les enceintes. C'est l'approche que Lavandier [2005] a choisi pour

évaluer les différences entre enceintes acoustiques sur la restitution du timbre. Lavandier [2005] précise "En évitant les jugements de qualité, de fidélité ou de préférence, nous espérons être moins dépendants des goûts et a priori des auditeurs". Le fait de demander aux auditeurs d'évaluer la différence entre enceintes semble être une tâche plus simple que l'évaluation absolue et surtout une tâche indépendante de leur expérience de la reproduction sonore. L'évaluation de l'effet de la distorsion d'un extrait musical telle que nous l'envisageons est similaire à celle de Lavandier [2005] et une tâche d'évaluation de dissemblance semble donc pertinente. La mesure de dissemblances perçues nous permettra de mettre en évidence à partir de l'analyse multidimensionnelle les dimensions sous-jacentes aux phénomènes de distorsions présents dans la restitution d'un signal musical.

L'analyse multidimensionnelle est une technique statistique permettant de représenter un corpus de stimuli dans un espace de faible dimension [Borg et Groenen, 1997]. Cette représentation est obtenue à partir d'une matrice de données de dissemblances estimées entre chaque paire de stimuli. L'analyse MDS permet de caractériser un corpus de sons sans savoir a priori en quoi les stimuli diffèrent [Borg et Groenen, 1997]. Autrement dit, les techniques de MDS sont appliquées dans des situations où l'expérimentateur ne comprend pas entièrement les attributs spécifiques selon lesquels les stimuli se distinguent les uns des autres ou ne veut pas les présupposer. La MDS nécessite donc seulement une estimation de la dissemblance globale entre les stimuli d'un corpus et non une mesure suivant un attribut spécifique. Pour s'assurer de la validité de l'analyse multidimensionnelle, il est nécessaire de tester des corpus homogènes d'échantillons. Cette contrainte permet d'éviter l'apparition de classes de stimuli dans l'espace perceptif et ainsi d'obtenir des dimensions continues. Un moyen de vérifier l'existence de groupement de stimuli avant l'interprétation des dimensions, est de faire intervenir l'analyse en arbre hiérarchique avant chaque analyse MDS. Les dendrogrammes nous renseignent sur de possibles classes au sein des stimuli. En l'absence de groupements nets, les dimensions continues révélées par la MDS peuvent être jugées pertinentes pour décrire la perception des auditeurs [Susini *et al.*, 1999]. L'interprétation des dimensions est l'étape finale de l'analyse MDS, et il est important de vérifier la continuité des dimensions.

Afin de pouvoir mettre en évidence un nombre suffisant de dimensions pour décrire les phénomènes testés, il est nécessaire de faire intervenir un grand nombre de stimuli [Kruskal et Wish, 1978]. De plus, un grand nombre de stimuli peut permettre de décrire plus précisément les dimensions. Le nombre de stimuli testé est directement lié à la méthode d'évaluation. Étant donné que nous cherchons à évaluer un grand nombre de stimuli, il est ainsi essentiel de faire intervenir une méthode de tests permettant l'évaluation de grands corpus.

2.1.2.c Résumé

Pour mettre en place un protocole permettant de caractériser les effets de la distorsion en suivant l'approche multidimensionnelle répondant aux besoins présentés précédemment, les deux points à considérer sont résumés sur la figure 2.2 : le corpus de stimuli que nous voulons évaluer et la méthode qui permettra d'évaluer ce corpus.

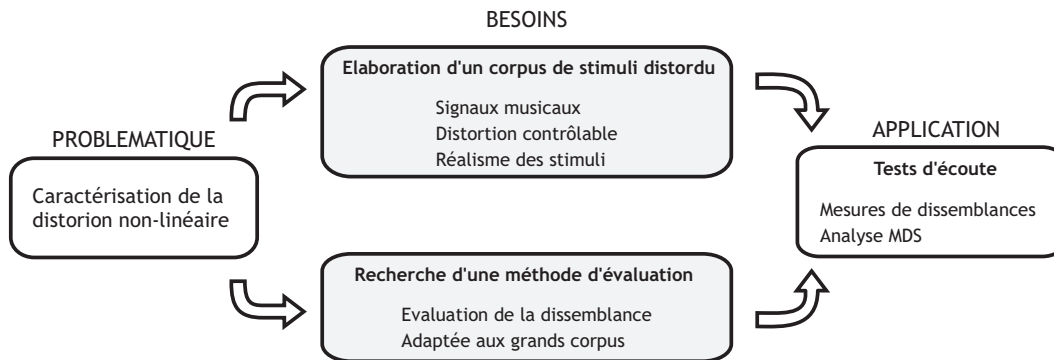


Figure 2.2 – Problématique de l'étude, besoin pour la conception du protocole et application aux tests d'écoute.

Le premier point concerne l'élaboration d'un corpus d'échantillons pour lesquels il est possible d'introduire et de contrôler différents types de distorsion. Dans le reste du document nous consacrons le terme "enceintes" à ces échantillons sonores car nous avons employé et modifié le son rayonné par des enceintes acoustiques comme moyen pour générer de la distorsion dans le signal musical. Le deuxième besoin correspond à la recherche d'une méthode d'évaluation perceptive susceptible d'estimer sans biais la dissemblance entre de nombreux stimuli. Ces deux points doivent être traités pour pouvoir obtenir des mesures de dissemblances entre les enceintes "distordues" susceptibles de révéler, via une analyse MDS, les dimensions sur lesquelles les auditeurs se sont basées pour effectuer leurs jugements de dissemblances. Dans la suite du chapitre nous revenons sur ces deux points en reprenant les besoins pour l'élaboration du protocole et les solutions que nous avons choisies.

2.2 Élaboration d'un corpus de stimuli distordus

Pour respecter le contexte d'écoute de musique reproduite, il est nécessaire de faire intervenir la reproduction d'un système de restitution réaliste. Pour suivre le protocole proposé par Lavandier [2005], il est nécessaire d'enregistrer différentes enceintes pour constituer un corpus. Une telle démarche est adaptée à la caractérisation du timbre mais pas à la caractérisation de la distorsion non linéaire. Si nous cherchons à nous focaliser uniquement sur les effets de la distorsion, il est indispensable de s'affranchir autant que possible de tout mécanisme autre que la

distorsion. Dans un corpus composé de différentes enceintes, le timbre propre à chaque système pourrait prédominer sur les caractéristiques de distorsion.

Utiliser de multiples enregistrements d'enceintes n'est donc pas compatible avec notre problématique visant à évaluer uniquement l'effet des non-linéarités sur les jugements de dissemblances perçues entre signaux distordus. Étant donné la difficulté de contrôler uniquement la distorsion non linéaire des enceintes réelles, il nous est apparu plus adapté d'avoir recours à des échantillons sonores de synthèse. Pour la caractérisation de la distorsion, les signaux de synthèse permettent plus facilement de s'assurer que les échantillons sonores sont comparables. De plus, ayant décidé d'évaluer les enceintes dans une même salle d'écoute, l'utilisation de signaux de synthèse nécessitera de pouvoir introduire l'influence d'une salle unique. Nous envisageons de combiner des signaux de synthèse avec la réponse de la salle dans laquelle Lavandier [2005] a enregistré son corpus de 37 enceintes, afin de pouvoir au besoin comparer au sein d'un même test des enregistrements réels et nos signaux de synthèse.

2.2.1 Modélisation de haut-parleurs

Nous avons cherché à simuler le comportement d'une enceinte en élaborant un modèle construit à partir de l'enregistrement d'une enceinte réelle. Pour caractériser le rayonnement d'une enceinte indépendamment d'une salle, il est nécessaire d'effectuer un enregistrement à une distance relativement proche. Pour caractériser uniquement son comportement, l'enregistrement d'une enceinte doit être effectué dans un environnement anéchoïque. Se pose alors le problème d'ajouter à cet enregistrement la signature d'une salle de manière réaliste. Nous développons cette étape appelée auralisation par la suite.

Un enregistrement en champ proche dans l'axe du boomer caractérise le comportement du haut-parleur assimilé à un piston plan. La pression mesurée en champ proche est alors directement liée au débit de la membrane [Keele Jr., 1974]. Par contre, le champ rayonné par un haut-parleur dédié aux hautes fréquences (tweeter) est bien plus complexe que celui du boomer, et il n'existe pas encore de méthode rigoureuse permettant de modéliser correctement le signal issu du tweeter à partir d'un enregistrement. Pour se rapprocher du rayonnement d'une enceinte acoustique, nous avons donc décidé d'associer le modèle d'un boomer et l'enregistrement du tweeter d'une même enceinte deux voies que nous appelons "génératrice". Ce choix nécessite de séparer le champ émis par le boomer et celui émis par le tweeter pour pouvoir isoler chaque champ contribuant au rayonnement complet de l'enceinte. Nous appelons dans la suite "enceinte virtuelle" un signal synthétisé selon ce procédé.

Il faut garder à l'esprit que l'enregistrement du tweeter doit être réalisé à une distance correspondant au champ lointain, tout en garantissant que le champ direct soit dominant. La position de mesure optimale pour le boomer et le tweeter n'est donc pas la même. Or il est préférable pour

le réalisme des échantillons que ces positions de mesures soient proches et qu'elles soient chacune dans l'axe du haut-parleur concerné. Cette configuration n'est facilement réalisable que dans le cas d'un haut-parleur coaxial, dont le tweeter se trouve au centre du boomer. Ainsi, les centres acoustiques des deux transducteurs sont très proches. Inversement, pour une enceinte avec des transducteurs séparés, les haut-parleurs ont des centres acoustiques éloignés ce qui conduit à une directivité dépendant fortement de la fréquence notamment autour de la fréquence de coupure du filtre de l'enceinte, ce qui limiterait le réalisme de l'étape d'auralisation.

Les enceintes virtuelles ainsi créées doivent alors permettre d'introduire puis de contrôler la distorsion non linéaire. Dans un premier temps, nous nous sommes bornés à simuler et modifier le comportement non linéaire du boomer. Les enceintes virtuelles ainsi obtenues sont donc composées d'un modèle de boomer modifié et de l'enregistrement du tweeter, permettant de générer des variantes de l'enceinte "génératrice". Pour la modélisation du boomer, deux approches peuvent alors être suivies : une approche basée sur l'identification de phénomènes physiques spécifiques au comportement d'un haut parleur [Klippel, 2006], ou une approche globale modélisant le haut parleur à partir d'une identification en "boite noire" [Novák, 2009].

2.2.1.a Approche physique

Dans l'approche "physique", les phénomènes non linéaires sont supposés connus et un modèle paramétrique permet de les contrôler. Klippel [2006] résume les différents phénomènes identifiés comme le comportement non linéaire du facteur de force, de la suspension ou de l'inductance. Le comportement non linéaire de ces phénomènes est exprimé en fonction du déplacement de la bobine et de l'intensité qui la parcourt. Ces différentes relations sont intégrées dans un modèle à constantes localisées de haut-parleur sous la forme d'un circuit équivalent où chaque composant représente un phénomène. Ce modèle, valable uniquement aux basses fréquences, est implémenté dans le domaine numérique afin de reproduire le son qui serait produit par le haut-parleur. Ensuite, l'influence de chaque élément non linéaire identifié peut être amplifié ou diminué afin de mettre en avant l'influence d'une non-linéarité spécifique ou la combinaison de plusieurs phénomènes. Ce modèle de haut-parleur est utilisé dans plusieurs études sur la perception de la distorsion [Schmitt, 1995; Klippel, 2001; Boer *et al.*, 1998; Kristoffersen *et al.*, 1999]. Par contre, une telle approche peut rapidement amener à des formulations complexes dues aux nombreux phénomènes non linéaires présents simultanément dans la restitution du haut-parleur, et ne peut décrire que les phénomènes inclus a priori dans le modèle.

2.2.1.b Approche globale

L'approche "globale" permet de modéliser le haut-parleur comme une "boite noire". Pour ce type de modélisation, les distorsions introduites par le système non linéaire testé sont constatées

sans se préoccuper de leur origine [Novák, 2009]. Autrement dit, aucun a priori sur l'origine des non-linéarités spécifiques du haut-parleur n'est nécessaire pour obtenir le modèle.

Pour cette méthode, le système non linéaire (ici le haut-parleur), peut être représenté sous la forme de plusieurs branches parallèles, chacune basée sur le comportement d'un terme non linéaire d'ordre spécifique. Sur le diagramme présenté sur la figure 2.3, les branches du modèle associent un polynôme d'ordre N croissant agissant sur le signal d'entrée, et une réponse fréquentielle spécifique à cet ordre. Le modèle résultant est assez général, mais il repose sur l'hypothèse que le haut-parleur peut être représenté par de telles branches qui traitent séparément une non-linéarité instantanée et un filtre temporel.

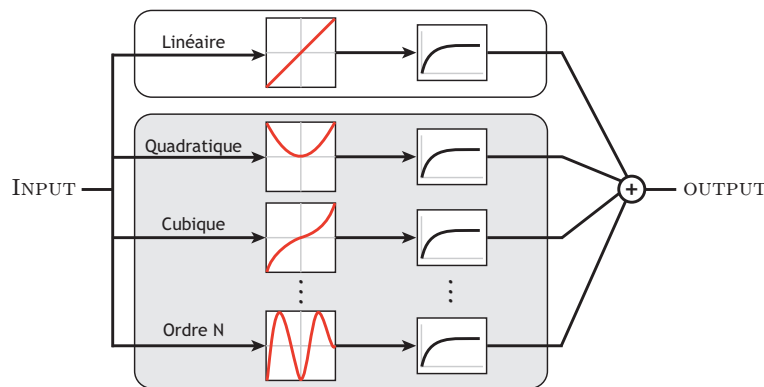


Figure 2.3 – Diagramme d'un système non linéaire. Chaque branche associe une composante à un ordre N et une réponse fréquentielle spécifique.

Nous avons choisi cette méthode pour modéliser le haut parleur car elle associe une méthode d'identification rapide à une décomposition efficace permettant la re-synthèse de signaux. Un avantage considérable de cette approche est qu'il existe une méthode d'identification proposée par Novák *et al.* [2010] qui permet de décomposer le comportement du boomer sur une base de polynômes orthogonaux. Nous détaillerons ce point essentiel de la méthode par la suite dans la section 3.3.2.

2.2.2 Contrôle de la non-linéarité

A partir du modèle de boomer nous pouvons paramétrer la non-linéarité afin de modifier la distorsion dans le signal de boomer synthétisé. Ce boomer modifié est ensuite combiné à l'enregistrement du tweeter pour recréer une enceinte virtuelle présentant des caractéristiques de distorsion différentes de l'enceinte originale.

Une manière efficace d'introduire de la distorsion dans un signal consiste à utiliser une courbe entrée/sortie décrivant le comportement non linéaire d'un système donné. Dans la littérature des formes de non-linéarités artificielles (polynomiale, de zero crossing ou de clipping) sont souvent employées pour l'étude de la distorsion non linéaire. Pour reprendre quelques exemples récents,

les études réalisées par Tan *et al.* [2003] ou Geddes et Lee [2003a] font intervenir différentes formes de non-linéarités artificielles appliquées à des extraits de musique ou de parole. L'utilisation de ces non-linéarités artificielles - instantanées ou "sans mémoire" - permet de simuler une grande variété de phénomènes non linéaires mais sont relativement loin du comportement réel d'un haut parleur. Or nous souhaitons aussi contrôler des signaux de synthèse se rapprochant du son émis par les enceintes. Nous avons donc utilisé le modèle présenté précédemment basé sur l'identification d'un système non linéaire appliquée à un haut-parleur de graves (boomer), soit pour re-synthétiser le haut-parleur original, soit pour modifier son comportement non linéaire. Nous avons de plus adapté ce même modèle afin de représenter de la même manière les non-linéarités artificielles évoquées précédemment. Ce modèle permet ainsi de générer une plus grande variété de phénomènes non linéaires pouvant rester proches du comportement du haut-parleur original, ou s'approcher de non linéarités artificielles, en variant continuellement des paramètres. Nous pouvons donc élaborer un corpus d'enceintes virtuelles dont les caractéristiques de distorsion sont variées et maîtrisées. Notre protocole permet de plus d'obtenir une résolution aussi fine que possible de la modification de la distorsion. Cette spécificité permet de maîtriser l'homogénéité du corpus d'échantillons ainsi créés.

2.2.3 Restitution

Étant donné que les signaux d'enceinte virtuelle sont issus d'un procédé de synthèse, il est nécessaire de les faire écouter au travers d'un système de restitution. De plus, nous cherchons à évaluer des différences relatives entre les échantillons et une commutation rapide d'un signal à un autre est alors impérative. Nous avons également choisi de faire évaluer les "enceintes virtuelles" combinées à la signature d'une salle d'écoute. Les caractéristiques de la pièce peuvent avoir des effets de masquage sur certains produits de distorsion ou en mettre d'autres en valeur [Cabot, 1990]. En étant conscients de l'influence de la pièce, Boer *et al.* [1998] ont néanmoins étudié l'audibilité des distorsions non linéaires des enceintes dans une pièce d'écoute et Kristoffersen *et al.* [1999] ont évalué les préférences d'enceintes qui distordaient dans un "salon de taille typique". Afin de pouvoir comparer éventuellement les signaux d'enceintes virtuelles obtenues à partir du modèle et celle enregistrée par Lavandier [2005], nous devons garantir les mêmes caractéristiques de réverbération. Nous avons pour cela utilisé la technique d'auralisation pour introduire dans nos échantillons d'enceintes synthétisées les caractéristiques de la même salle d'écoute que celle utilisée par Lavandier *et al.* [2005]. Étant donné que nous cherchons à évaluer les différences relatives entre signaux distordus, il est nécessaire d'effectuer une comparaison rapide compte tenu de notre mémoire auditive limitée. Une solution souvent adoptée dans le cadre de la reproduction sonore consiste à reproduire au casque les signaux à évaluer. Les signaux peuvent soit être les enregistrements d'enceintes acoustiques dans une salle, soit, dans notre cas, des signaux

synthétisés auxquels une réponse de salle a été appliquée.

Un dernier point à gérer rigoureusement est le niveau d'écoute des stimuli [Bech et Zacharov, 2006]. Le niveau de restitution agit à deux étapes dans l'étude que nous menons. Le niveau lors de l'identification de l'enceinte joue un rôle sur la quantité de distorsion générée : des niveaux plus élevés auront tendance à plus solliciter les non-linéarités des enceintes acoustiques. D'un autre côté, le niveau des stimuli reproduits au casque lors des tests d'écoute doit être contrôlé. Comme précisé dans le chapitre précédent, l'oreille se comporte de manière non linéaire avec l'amplitude du signal. Ainsi, la comparaison de deux stimuli peut être influencée par cette éventuelle différence de niveau et donc perturber le jugement des auditeurs. Pour l'évaluation des dissemblances, le contrôle du niveau est donc indispensable afin de fixer l'attention des auditeurs sur les phénomènes de distorsion que nous avons introduits dans les différentes enceintes synthétisées. Il convient alors d'égaliser tous les échantillons sonores d'un même test à un même niveau de restitution perçu, bien qu'ils correspondent à des phénomènes physiques qui dépendent normalement du niveau.

2.3 Recherche d'une méthode d'évaluation de la dissemblance

2.3.1 Méthodes usuelles pour l'évaluation de la dissemblance

Comme mentionné plus haut, il est nécessaire de mesurer des dissemblances adaptées à l'analyse multidimensionnelle sur un grand nombre de stimuli. Dans cette section nous exposons les deux méthodes usuelles pour l'évaluation de la dissemblance qui ont par ailleurs été utilisées par Lavandier [2005] pour la caractérisation du timbre. Les principes de ces deux méthodes et leurs limitations sont présentés.

2.3.1.a Comparaison par paires

La méthode de comparaison par paires est la méthode recommandée pour évaluer directement des dissemblances entre des stimuli audio [IEC 60268-13, 1998; AES20, 1996; Lipshitz et Vanderkooy, 1981], ou pour tout autre type de stimuli [Borg et Groenen, 1997; Takane *et al.*, 2009]. Dans cette méthode, les auditeurs doivent évaluer directement la dissemblance entre chaque stimulus d'une paire en donnant une estimation sur une échelle variant de "très semblables" à "très dissemblables". Ces estimations permettent d'obtenir une matrice de dissemblance individuelle pour chaque auditeur. En général, chaque paire de sons est présentée une seule fois au cours du test (c'est à dire que si la paire (i,j) est évaluée, la paire (j,i) ne l'est pas), et un stimulus n'est jamais comparé à lui même. La moyenne des matrices individuelles fournit une matrice contenant les jugements moyens de dissemblances des auditeurs.

Cette méthode est largement utilisée pour estimer la dissemblance afin d'évaluer différentes

caractéristiques sonores. Pour citer quelques exemples, Grey [1977] a employé la méthode de comparaison par paires pour évaluer la dissemblance perçue entre différents instruments de musique afin d'extraire les paramètres physiques du timbre instrumental. De même, Caclin *et al.* [2005] ont mené une expérience dans laquelle ils génèrent plusieurs stimuli en modifiant des caractéristiques physiques du timbre. Susini *et al.* [1999] ont analysé l'espace perceptif résultant de l'évaluation de la dissemblance de sons d'intérieurs de voitures dans une étude concernant la qualité sonore des habitacles. D'autres études sur la caractérisation perceptive du son rayonné par des barres ou des plaques de différents matériaux ont aussi fait intervenir une évaluation par paires de la dissemblance perçue [McAdams *et al.*, 2004, 2010; Canevet *et al.*, 2004]. La méthode de comparaison par paires se retrouve aussi employée dans des études menées sur la caractérisation la composante spatiale du son. Blauert et Lindemann [1986] ont travaillé sur la caractérisation de la sensation d'espace dans des salles de concerts, ou encore Zahorik [2009] a étudié la dissemblance entre des sons de paroles dans différentes pièces simulées. Des études sur des applications encore plus éloignées utilisaient l'estimation de dissemblance par paires. C'est le cas de la caractérisation de sons "de tous les jours" par Cermak [1979]; Bergman *et al.* [2009] ou la caractérisation des implants cochléaires par McKay *et al.* [1996]. Pour la caractérisation du rayonnement des enceintes, la comparaison par paires est aussi employée [Lipshitz et Vanderkooy, 1981; Gabrielson et Lindstrom, 1985].

Malheureusement, la méthode de comparaison par paires implique des tests assez contraignants pour évaluer un grand nombre de stimuli [Borg et Groenen, 1997]. En effet, le nombre de présentations (paires) de l'expérience augmente rapidement avec le nombre de stimuli. Par exemple, pour un test mené sur un corpus de 30 stimuli, l'évaluation de 435 paires est nécessaire pour obtenir une matrice de dissemblance complète par auditeur. Dans ce cas, le test entier peut prendre beaucoup de temps et la fatigue ou le manque d'attention de l'auditeur peuvent devenir significatifs [Scavone *et al.*, 2002]. Dans la plupart des évaluations perceptives, il est conseillé d'effectuer des pauses entre différentes séries de jugements afin d'éviter les désagréments d'un test d'une durée trop importante. Une durée maximale de 30 minutes par session de tests est ainsi recommandée par la norme AES20 [1996]. Compte tenu de la durée maximale de l'ordre de 30 minutes et d'une durée de stimuli de quelques secondes, le nombre de paires pouvant être évaluées sans biais au cours d'une même session est donc a priori limité.

La majorité des études qui font intervenir la méthode de comparaison par paires impliquent ainsi au maximum 20 stimuli. Cependant l'évaluation d'un plus grand nombre de stimuli pourrait permettre notamment une répartition plus homogène des stimuli suivant les dimensions perceptives et donc une meilleure description de l'espace perceptif. De plus, le nombre de dimensions révélées par la MDS est limité par le nombre de stimuli, donc évaluer un plus grand corpus est nécessaire pour révéler un nombre éventuellement supérieur de dimensions. Le nombre de dimensions discernables est relié au nombre de stimuli par une relation empirique qui précise que

n stimuli permettent de mettre en évidence un maximum de $(n - 1)/4$ dimensions [Kruskal et Wish, 1978].

Pour évaluer un grand corpus, il pourrait être envisagé de subdiviser un test de comparaison par paires. A notre connaissance, aucune étude n'a comparé les résultats d'un test sur un corpus étendu selon que le test ait été effectué en une seule session ou en plusieurs sessions. La plupart des études évoquant des corpus étendus de stimuli indiquent seulement que l'évaluation de tels corpus avec la comparaison par paires serait "irréalisable" [Tsogo *et al.*, 2000], "indésirable" [Spence et Domoney, 1974], "peu pratique" Rao et Katz [1971] ou encore "limitée" par le nombre de stimuli [Bonebright, 1996]. Par ailleurs, Poulton [1979] souligne l'existence d'un biais séquentiel qui est présent entre le jugement effectué lors d'un essai et ceux effectués pour les essais précédents. Dans le cas de la comparaison par paires, pour éviter ce biais, tous les essais devraient donc être évalués au sein d'une même session car le jugement effectué lors d'un essai peut être influencé par le jugement des essais précédents. Poulton [1982] précise que le biais séquentiel concerne n'importe quelle série de jugements, à moins que chaque stimulus ne soit comparé intentionnellement à un stimulus de référence.

A notre connaissance, aucune étude ne permet de savoir si le fractionnement d'un test en plusieurs sessions tend à réduire ce biais séquentiel ou non, ou s'il peut introduire un autre artefact comme une modification de l'échelle utilisée par les auditeurs. La comparaison par paire est ainsi la méthode la plus reconnue pour l'évaluation de la dissemblance mais il n'est pas possible de savoir rigoureusement si elle est adaptée à l'étude d'un corpus étendu telle que nous l'envisageons.

2.3.1.b Classification libre

La méthode de classification libre est une méthode adaptée à l'évaluation de plus grands corpus de stimuli [Borg et Groenen, 1997]. Elle est en général utilisée pour des tâches de catégorisation, mais elle offre aussi la possibilité d'obtenir des jugements de dissemblance de manière indirecte entre de nombreux stimuli [Thibaut, 1997].

Le principe de la méthode consiste à présenter la totalité des stimuli à l'auditeur. Des symboles correspondant aux stimuli sont répartis aléatoirement sur un écran et il est demandé à l'auditeur de les regrouper par similitude. L'auditeur a la possibilité d'écouter les stimuli autant de fois que nécessaire et de les déplacer pour les regrouper en fonction de leur ressemblance. Des stimuli jugés semblables sont placés dans un même groupe et les stimuli jugés différents apparaissent dans des groupes différents.

La matrice individuelle d'un auditeur est remplie binairesment de la manière suivante : la valeur 1 est attribuée lorsque deux stimuli sont placés dans des groupes différents et la valeur 0 lorsque deux stimuli apparaissent dans un même groupe. Pour cette méthode indirecte, la

matrice de dissemblance résulte de la moyenne des matrices individuelles remplies avec ces valeurs binaires. Ainsi, deux stimuli qui auraient été classés par tous les auditeurs dans le même groupe auront une valeur de dissemblance moyenne de 0. A l'inverse, deux stimuli jamais classés ensemble auront une dissemblance égale à 1. On peut ainsi mesurer une forte ressemblance entre les deux premiers stimuli et une forte dissemblance entre les deux derniers. Par contre, deux stimuli plus ou moins semblables seront classés ensemble par certains auditeurs et séparés par d'autres, induisant une valeur de dissemblance intermédiaire. Cette méthode indirecte nécessite de ce fait un grand nombre d'auditeurs afin d'estimer des valeurs de dissemblances sans trop d'effet de quantification.

Bien que moins utilisée que la comparaison par paires, cette méthode est ainsi employée dans des études cherchant à évaluer les dissemblances entre de nombreux sons. A titre d'exemples, Lemaitre *et al.* [2007] ont mené un test sur l'évaluation du timbre de 43 sons d'avertisseurs de voiture avec la méthode de classification libre. De même, Bonebright [2001] a mené une expérience avec cette méthode sur 74 sons d'objets de la vie quotidienne dans le but d'évaluer, à l'aide de la MDS, leurs structures perceptives.

Dans le travail de Lavandier [2005], les résultats perceptifs obtenus sur un corpus de 37 enceintes avec la classification libre étaient proches de ceux obtenus avec la comparaison par paires sur un corpus de 12 enceintes. Les espaces perceptifs résultant de l'analyse des dissemblances évaluées selon ces deux approches comprenaient des dimensions communes. Cependant, une des dimensions ne présentait pas le même comportement d'un espace perceptif à l'autre. La première dimension était liée à un équilibre spectral : la comparaison par paire a dévoilé une dimension qui évoluait depuis les sons ayant trop de basses, vers des sons "équilibrés" et enfin des sons manquant de basses. L'analyse des données issues du test de classification libre a révélé un ordre différent suivant la dimension spectrale correspondante : étaient classés d'abord les sons manquant de basses, puis les sons ayant trop de basses et enfin les sons jugés "équilibrés". Lavandier [2005] a émis l'hypothèse que les sons auraient été jugés en partie selon une préférence et non selon leur seule similarité. Ce résultat conduit à douter de la validité de la méthode de classification libre pour l'évaluation de la dissemblance, au moins dans le cas de signaux musicaux.

Aldrich *et al.* [2009] ont également effectué la comparaison entre la méthode de comparaison par paires et la classification libre sur un même corpus de 20 sons. Ils ont démontré que les auditeurs participant à une tâche de classification libre auraient tendance à grouper les stimuli par catégories plutôt que par similarités lorsque l'origine des sons est identifiable. C'est le cas d'un test sur des sons d'instruments de musique ou encore des sons de l'environnement. Finalement, Aldrich *et al.* [2009] concluent que pour un corpus étendu composé de plus de 20 stimuli, la méthode de classification libre est utilisable mais susceptible de fournir des informations catégorielles. Dans une autre étude comparative entre la comparaison par paires et la classification libre, Parizet et Koehl [2012] trouvent aussi des différences entre les deux espaces perceptifs. Ils conseillent alors

d'utiliser la classification libre comme une première étape pour déterminer un sous-ensemble réduit de sons qui ensuite serait évalué avec la comparaison par paires. C'est la solution qu'ont choisi Scavone *et al.* [2001] pour évaluer 150 sons complexes en utilisant le Sonic Mapper [Scavone *et al.*, 2002] qui fournit la combinaison d'une tâche de classification libre et de comparaison par paires. Dans un premier temps, les auditeurs procédaient à la tâche de classification afin de créer des groupes, puis ils effectuaient des jugements de dissemblances entre les stimuli au sein de chaque groupe avec la méthode de comparaison par paires. Cette approche mixte ne nous semble pourtant pas envisageable car la durée cumulée des deux tests pourrait être importante et la tâche de classification libre est souvent perçue comme compliquée à effectuer pour les auditeurs.

A partir des limitations des deux méthodes décrites précédemment, il semble donc nécessaire de rechercher une méthode qui puisse fournir des mesures de dissemblances sur un corpus étendu de stimuli.

2.3.2 Méthodes alternatives adaptées à l'évaluation de corpus étendus

Afin de trouver une méthode alternative à la comparaison par paires et à la classification libre, nous avons examiné des méthodes connues permettant d'obtenir des données de dissemblance et adaptées à des corpus étendus de stimuli. Les études de Rao et Katz [1971], Subkoviak et Roecks [1976], Bijmolt et Wedel [1995] et Tsogo *et al.* [2000] présentent plusieurs méthodes alternatives à la comparaison par paires et à la classification libre. La plupart des méthodes sont adaptées à une évaluation indirecte de la dissemblance qui permet de prendre en compte la difficulté d'évaluer des corpus étendus. Nous avons choisi de présenter les méthodes suivant la tâche qui est demandée pour évaluer la dissemblance. Les premières concernent les tâches de similarité incomplète, les deuxièmes les tâches de classement, et les dernières méthodes font intervenir les tâches de regroupement. Nous nous sommes intéressés aux méthodes utilisées pour l'évaluation de stimuli sonores, mais nous avons étendu nos recherches vers d'autres domaines. C'est pour cette raison que nous parlons ici de participants et non d'auditeurs pour la description des différentes méthodes.

2.3.2.a Tâche de similarité incomplète

La manière la plus pragmatique d'aborder le problème de la comparaison par paires lorsque le nombre de stimuli est élevé est de réduire le nombre de paires. Spence et Domoney [1974] ont proposé une méthode qui consiste à ne pas considérer ou à supprimer une partie de la matrice de dissemblance. Seule une partie des paires à évaluer est alors présentée à chaque participant. Deux proportions de suppression ont été testées dans cette étude : un tiers ou deux tiers de la matrice de dissemblance. Spence et Domoney ont défini deux manières de supprimer les paires dans la matrice. Dans la première, les paires sont supprimées aléatoirement. La deuxième

est dite à "suppression cyclique" car les paires sont effacées dans un ordre tel que tous les stimuli à juger apparaissent un nombre de fois égal dans les paires à évaluer. Ils ont montré que supprimer un tiers des paires à évaluer de façon cyclique ou même aléatoire permettait d'obtenir des données de dissemblances satisfaisantes. Sachant qu'un participant n'évalue qu'une partie des paires, cette méthode demande de faire intervenir un grand nombre de participants afin d'obtenir une matrice de dissemblance moyenne plus précise. Graef et Spence [1979] ont évalué quelles sont les dissemblances qui sont importantes lorsqu'un tiers de la matrice est supprimé. Ils ont ainsi proposé trois nouvelles façons de supprimer un tiers des paires de stimuli. La première consiste à supprimer des plus petites dissemblances, la deuxième les dissemblances moyennes et la dernière les plus grandes dissemblances. Ils montrent que supprimer le tiers correspondant aux dissemblances moyennes permet d'obtenir des résultats satisfaisants.

2.3.2.b Tâche de classement

Classement par similarité :

Une autre méthode définie par Rao et Katz [1971] consiste à demander aux participants d'ordonner des stimuli en fonction de leur similarité croissante. A partir d'un corpus de n stimuli, l'un d'entre eux est désigné comme étant le stimulus de référence. Ensuite, parmi les $n - 1$ stimuli, k stimuli de comparaison sont présentés aux participants qui doivent les comparer au stimulus de référence et les ordonner du plus similaire au moins similaire. La matrice individuelle, dont la diagonale est fixée à 1, est remplie ligne par ligne, chaque ligne correspondant à une référence. L'élément (i,j) correspond à l'ordre attribué au stimulus j lorsque i est présenté en tant que stimulus de référence. Pour chaque jugement par rapport à une référence, est assignée une valeur de 2 à $k + 1$ en fonction de l'ordre des stimuli. La valeur de 2 correspond au stimulus qui aura été jugé le plus similaire et la valeur $k + 1$ au dernier stimulus. Rao et Katz proposent que les $(n - k - 1)$ éléments de la matrice restants soient remplis avec une valeur d'ordre moyen : $(n + k + 2)/2$. Ce procédé est répété $n - 1$ fois, en changeant la référence à chaque fois. La matrice de dissemblance finale correspond alors à la moyenne de toutes les matrices individuelles.

Classement par similarité - tâche interactive :

Young *et al.* [1982] proposent une méthode de classement interactive qui reprend le principe de la méthode présentée ci-dessus mais avec une tâche légèrement différente. Lorsque le stimulus le plus similaire à celui de référence est choisi parmi k stimuli de comparaison, il est alors retiré de la liste des stimuli à évaluer. L'essai suivant, un stimulus de comparaison, pas encore évalué, est ajouté à la liste. Ce procédé est répété jusqu'à ce que les $n - 1$ suivants soient ordonnés. Ce procédé est effectué pour chaque référence afin de remplir chaque ligne de la matrice de dissemblance.

2.3.2.c Tâche de regroupement

Classification hiérarchique :

Rao et Katz [1971] ont proposé la méthode de classification hiérarchique comme variante à la méthode de classification libre. Pour cette méthode, les participants doivent d'abord effectuer une classification libre et regrouper en un certain nombre de groupes plusieurs stimuli. Une fois cette première tâche effectuée, il est demandé aux participants d'associer deux groupes considérés comme étant les plus similaires. Cette deuxième tâche d'association est effectuée jusqu'à ce que le groupe entier de stimuli soit reformé. Rao et Katz ont défini une mesure de dissemblance entre deux stimuli i et j comme le nombre de groupes distincts dans lesquels i et j ont été classés. Ainsi, pour chaque participant une matrice individuelle de dissemblance peut être construite et la moyenne sur la totalité des participants permet d'obtenir une matrice de dissemblance adaptée à l'analyse MDS. Il existe une variante appelée méthode de classification hiérarchique tronquée. Le principe des deux méthodes est identique mais la configuration initiale est différente. Au départ d'un test de classification hiérarchique, le participant doit associer deux stimuli comme les plus proches parmi la totalité du corpus testé. La méthode tronquée consiste à présenter au participant un nombre prédéfini de groupes comme configuration initiale et il doit mener la tâche entre ces groupes.

Comparaison en triades :

Pour la tâche de comparaison en triade, un ensemble de trois stimuli est présenté aux participants qui doivent choisir la paire de stimuli les plus similaires et celle qui présente le plus de dissemblance [Takane, 1982]. Pour remplir la matrice de dissemblance, la paire contenant les sons les plus différents obtient la valeur 2, la paire avec les sons similaires, la valeur 0 et pour la paire restante la valeur 1. Pour un test comprenant n stimuli, le nombre total de triades est de $n.(n - 1).(n - 2)/6$. La conception du test en blocs incomplets ou BIBD (balanced incomplete block design) est une alternative qui permet de réduire le nombre de triades d'un facteur $\lambda/(n - 2)$. Dans ce type de design incomplet chaque paires de stimuli apparaît λ fois dans tout le test.

2.3.2.d Méthode dite de "pick any" (Choisir k parmi p stimuli) :

Les méthodes dite de "pick any" consistent à présenter au participant un stimulus de référence accompagné de p stimuli de comparaison [Rao et Katz, 1971]. Le participant a pour tâche au cours d'un essai de choisir les k parmi p stimuli qui semblent les plus similaires au stimulus de référence. Les expérimentateurs fixent le nombre de stimuli de comparaison entre 2 et $n - 1$ et ont la possibilité de choisir le nombre k ou de laisser le participant choisir. Le processus est répété $n - 1$ fois avec chaque stimulus apparaissant comme le stimulus de référence. Une matrice de

similarité initialisée avec la valeur 0 est remplie en fonction des k stimuli choisis pour chaque référence. Pour une référence i , si le son j est choisi comme le plus similaire, l'élément (i,j) de la matrice de similarité est incrémenté. Cette méthode de choix du stimulus le plus similaire à une référence a été aussi proposée par Torgerson [1952] comme une variante de la méthode de comparaison en triades. Au sein d'une triade un stimulus est désigné comme étant le stimulus de référence et le participant doit choisir entre les deux restants celui qui semble le plus similaire à la référence.

2.3.3 Méthode retenue

A partir des travaux de Rao et Katz [1971], Subkoviak et Roecks [1976], Bijmolt et Wedel [1995] et Tsogo *et al.* [2000] plusieurs méthodes semblent adaptées à notre problématique. La plupart de ces méthodes ont été employées dans diverses expériences et généralement comparées entre elles et aux méthodes standard. Par exemple, Bijmolt et Wedel [1995] ont comparé les méthodes de comparaison par paires, de classification libre, de classement par similarité et de comparaison en triades. Il apparaît que la classification libre est le meilleur compromis entre la durée du test, la quantité d'information manquante et la fatigue ressentie par le participant. Giordano *et al.* [2011] se sont intéressés à la comparaison des méthodes de comparaison par paires, classification libre et classification hiérarchique. Ils concluent que les deux méthodes de classification sont plus efficaces que la méthode de comparaison par paires faisant référence à la durée du test et la quantité d'information de dissemblance par rapport à la validité des résultats perceptifs.

Parmi les méthodes présentées précédemment la méthode dite de "pick any" présentée par Rao et Katz [1971] n'a pas été approfondie et surtout n'a pas été comparée à la méthode standard de comparaison par paires. D'après Rao et Katz [1971], elle semblerait équivalente à la méthode de comparaison en triades et à la méthode de classement par similarité. De plus, la méthode a été appliquée en perception visuelle par Rogowitz *et al.* [1998] afin d'évaluer la similarité sur un corpus étendu composé de 97 images.

L'expérience de Rogowitz *et al.* [1998] nous donne un exemple d'application et nous permettra d'adapter la méthode aux tests d'écoute. Lors d'un essai, 9 images choisies aléatoirement parmi les 97 étaient présentées. Une de ces images était considérée comme l'image de référence et le participant devait juger parmi les 8 restantes, quelle image était la plus similaire à la référence. La figure 2.4 donne un aperçu visuel des 9 images présentées au cours d'un essai lors du test mené par Rogowitz *et al.* [1998].

Pour cette tâche le nombre de stimuli à choisir comme étant similaires à la référence a été fixé à 1 et aucune définition de similarité n'était donnée aux participants avant le test. Chaque référence est comparée aux $n - 1$ autres stimuli soit 96 stimuli pour cette étude, répartis aléatoirement en



Figure 2.4 – Exemple d’un essai du test de Rogowitz *et al.* [1998]. Le stimulus de référence à gauche est comparé aux 8 autres stimuli de comparaison situés à droite.

groupe de 8. Pour ce corpus de 97 stimuli, 12 présentations de 8 images étaient donc effectuées pour chaque référence. Le test entier comprenait alors 1164 essais au total. Rogowitz *et al.* proposent d’initialiser la matrice de similarité individuelle s de chaque participant avec la valeur 0. Lorsque le participant juge le stimulus j comme étant le plus similaire à la référence i , alors l’élément s_{ij} de la matrice individuelle est incrémenté de 1. La valeur de dissemblance associée est obtenue par $d_{ij} = 1 - s_{ij}$. En effectuant le test de comparaison à une référence sur un nombre suffisant de participants, Rogowitz *et al.* ont appliqué l’analyse MDS sur la matrice moyenne de dissemblances afin de trouver les dimensions sur lesquelles les participants se sont appuyés pour effectuer leurs jugements de similarité entre les images.

Cette méthode n’a pas été employée pour l’évaluation de stimuli audio mais la tâche ainsi que la conception du test semblent pouvoir permettre l’étude de corpus étendus de stimuli. Nous avons décidé d’adopter cette méthode, appelée par la suite méthode de **comparaison avec permutation de références** et abrégée **CPR** pour le reste du document. Le nom a été choisi car l’une des particularités de la méthode est de faire intervenir chacun des stimuli comme référence comparée aux autres stimuli du corpus. La méthode est décrite plus en détails au cours du chapitre 4 et comparée aux méthodes de comparaison par paire et de classification libre dans le chapitre 5.

2.4 Protocole proposé

En utilisant la méthode de comparaison avec permutation de références et le principe de synthèse d’enceintes virtuelles, nous disposons donc des deux outils pour l’évaluation perceptive de l’effet de la distorsion non linéaire. La figure 2.5 résume le protocole que nous proposons, depuis l’enceinte génératrice jusqu’à l’espace perceptif obtenu après analyse des données de dissemblances.

Le protocole que nous proposons permet, dans un premier temps, de créer un modèle d’enceinte à partir de l’enregistrement d’une enceinte. Ce modèle possède la particularité d’être

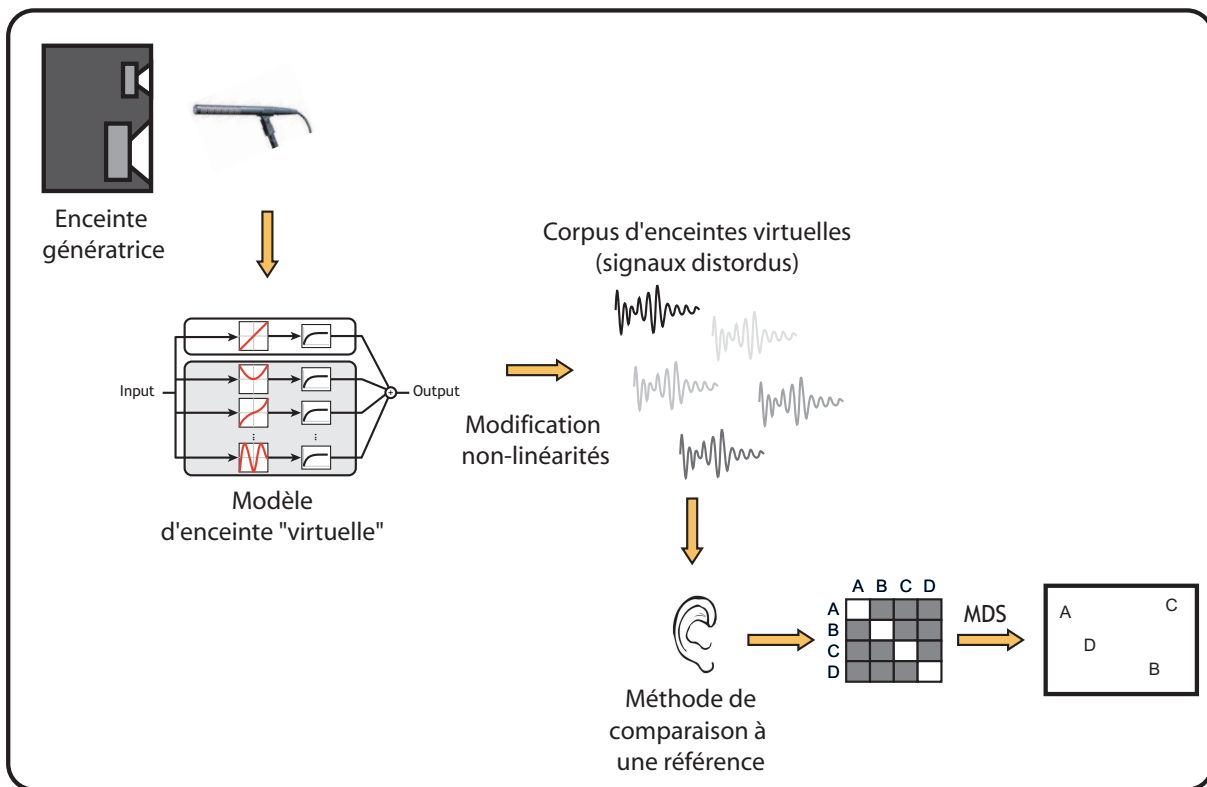


Figure 2.5 – Protocole proposé pour l'évaluation des non-linéarités des enceintes acoustiques.

la combinaison d'un modèle du boomer à non-linéarités contrôlables et de l'enregistrement du tweeter. Nous pouvons ainsi générer une grande variété de signaux, que nous appelons "enceintes virtuelles", présentant des distorsions différentes et obtenir un corpus étendu à volonté. La méthode de comparaison à une référence utilisée pour évaluer les dissemblances entre les enceintes virtuelles du corpus doit alors nous permettre de révéler l'espace perceptif lié à l'évaluation de la distorsion.

Chapitre 3

Élaboration et contrôle d'une enceinte "virtuelle"

Sommaire

3.1	Enceinte "génératrice"	55
3.2	Séparation des deux voies de l'enceinte génératrice	56
3.3	Modèle de boomer	58
3.3.1	Analyse : Identification des non-linéarités	58
3.3.2	Synthèse : Décomposition sur la base des polynômes de Chebyshev	61
3.3.3	Modélisation du boomer	63
3.3.4	Vérification objective du modèle de boomer	64
3.4	Modèle d'enceintes virtuelles	68
3.4.1	Enregistrement du tweeter	69
3.4.2	Synchronisation des voies	69
3.4.3	Équilibrage des niveaux	71
3.5	Auralisation	71
3.6	Vérification perceptive : enceinte virtuelle vs. enceinte réelle	72
3.6.1	Procédure	73
3.6.2	Enregistrements réels et virtuels	73
3.6.3	Tests ABX	74
3.6.4	Résultats	74
3.7	Contrôle des non-linéarités dans les enceintes virtuelles	76
3.7.1	Modification des non-linéarités identifiées sur le boomer	76
3.7.2	Introduction de non-linéarités artificielles	77

Ce chapitre présente dans un premier temps les différentes étapes de l'élaboration d'une "enceinte virtuelle". Le terme "virtuelle" a été adopté par opposition à "réelle" pour caractériser les enceintes dont nous avons simulé le comportement à partir de l'enregistrement d'un tweeter et la synthèse d'un boomer. Par la suite, le terme "enceinte virtuelle" désigne en fait le signal qui est obtenu à partir de cette simulation. La combinaison des deux voies, associée à une étape d'auralisation permet de simuler le son produit par une enceinte dans une salle d'écoute pour une écoute au casque. La figure 3.1 présente les cinq phases de l'élaboration des enceintes "virtuelles".

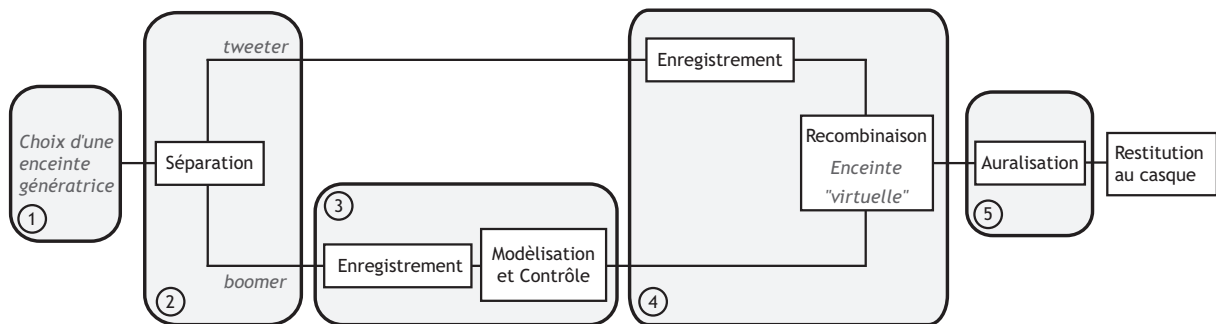


Figure 3.1 – Diagramme présentant les étapes de l'élaboration d'une enceinte "virtuelle" et l'étape d'auralisation permettant d'associer la signature d'une salle pour une écoute au casque.

A partir de l'enceinte virtuelle, nous avons modifié le modèle pour contrôler et introduire de la distorsion non linéaire. Cette partie est présentée après l'élaboration du modèle d'enceinte virtuelle.

3.1 Enceinte "génératrice"

Nous avons mentionné dans la section 2.2.1 l'importance d'utiliser une enceinte coaxiale pour permettre la réduction du champ 3D de l'enceinte à un enregistrement ponctuel indépendant pour chaque voie de l'enceinte. Nous avons ainsi choisi l'enceinte Tannoy System 600. Cette enceinte de milieu de gamme est une enceinte de monitoring bass reflex dont les deux haut-parleurs sont coaxiaux. Nous avons choisi d'obturer les deux événements situés en dessous des haut-parleurs car ils ne sont pas coaxiaux et pourraient perturber la prise de son en champ proche. Bien que modifiant sensiblement la réponse de l'enceinte aux basses fréquences, le fait de boucher les événements ne se traduit pas par une réponse aberrante. Notre enceinte génératrice est donc légèrement différente de l'enceinte commercialisée. L'enceinte est présentée sur la figure 3.2. Le schéma de l'enceinte et les caractéristiques techniques données par le constructeur sont présentés en annexe A.



Figure 3.2 – Photographie de l'enceinte Tannoy Sytem 600 avec ses deux événements bouchés.

3.2 Séparation des deux voies de l'enceinte génératrice

La constitution des "enceintes virtuelles" débute par la séparation du rayonnement du tweeter de celui du boomer de l'enceinte. Enregistrer les deux voies simultanément n'est pas possible car le rayonnement de chaque voie aurait interféré avec l'autre. Afin de mesurer indépendamment les deux voies, plusieurs options ont été envisagées. Nous avons écarté l'usage d'un filtrage actif, n'ayant ni l'équipement, ni les moyens de l'ajuster à l'enceinte. Une première solution passive, consistant à débrancher simplement le haut-parleur qui n'est pas enregistré, n'a pas non plus été retenue car nous voulions conserver les caractéristiques électriques et acoustiques de l'enceinte. Avec les bornes du haut-parleur débranchées, le filtre passif à l'intérieur de l'enceinte n'aurait pas le même comportement. Une deuxième solution passive consiste à remplacer le haut-parleur par une résistance pour simuler son impédance nominale. Les haut-parleurs ont en général une impédance nominale de 4 ou 8 Ω . En plaçant cette résistance aux bornes du filtre passif, son comportement se rapprocherait de la normale. Néanmoins, l'impédance d'un haut-parleur réel variant beaucoup avec la fréquence, cette configuration n'assurerait pas un fonctionnement normal du filtre.

Nous avons donc finalement choisi d'utiliser une paire d'enceintes du même modèle, a priori construites de telle sorte que leurs caractéristiques soient sensiblement identiques. Étant donné que nous cherchons à conserver les propriétés du filtre passif, l'utilisation d'un haut-parleur pour en remplacer un autre permet de conserver le fonctionnement normal du filtre. Pour enregistrer un des deux haut-parleurs de l'enceinte à l'intérieur d'une chambre anéchoïque, le signal de l'autre haut-parleur est envoyé à la deuxième enceinte, placée à l'extérieur. Cette solution nous a semblé être la plus adaptée pour respecter au mieux le comportement électrique et acoustique de l'enceinte tout en utilisant un équipement disponible au moment de ce travail. Pour illustrer cette technique de séparation de voies, la figure 3.3 présente les deux dispositifs utilisés pour

enregistrer indépendamment soit le tweeter, soit le boomer de l'enceinte. Pendant qu'une des voies de l'enceinte 1 est enregistrée, le signal à la sortie du filtre passif est envoyé à l'autre voie de l'enceinte 2. Pour une raison de clarté, l'enceinte schématisée sur la figure 3.3 présente deux haut-parleurs décalés. Cette représentation séparée des haut-parleurs, en fait coaxiaux, est utilisée dans les figures afin de distinguer clairement l'enregistrement du tweeter de celui du boomer.

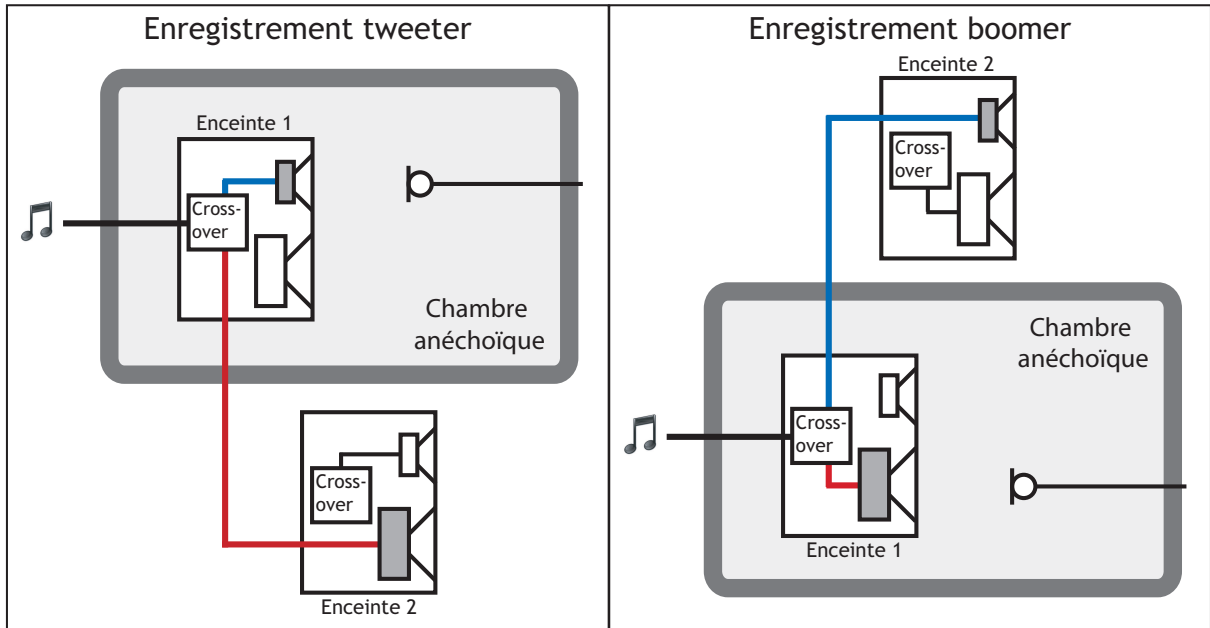


Figure 3.3 – Dispositif expérimental choisi pour enregistrer indépendamment le tweeter et le boomer d'une enceinte.

Avant d'utiliser cette technique, il est nécessaire de s'assurer que les deux enceintes sont à peu près identiques. Nous avons donc mesuré la réponse fréquentielle de chaque enceinte. L'enregistrement a été réalisé en chambre anéchoïque avec un microphone G.R.A.S. 1/2" modèle 40AF positionné à 1 m dans l'axe des haut-parleurs coaxiaux. Le signal utilisé était un sweep dont la fréquence de départ était 10 Hz et la fréquence finale était 22050 Hz. La restitution et l'acquisition ont été effectuées avec un convertisseur audio Fostex VC-8 couplé à une carte son RME DIGI9652, en utilisant une fréquence d'échantillonnage de 44100 Hz. La figure 3.4 représente les réponses fréquentielles obtenues pour les deux enceintes.

Le comportement fréquentiel de ces deux enceintes est sensiblement identique. Sans surprise, les différences les plus importantes apparaissent aux alentours de la fréquence de coupure du filtre passif ($f_c = 1800$ Hz). Ces différences reflètent les tolérances de fabrication des paires d'enceintes de coût raisonnable. Les réponses des deux enceintes pour des hautes fréquences reflètent également des phénomènes de diffraction qui indiquent un manque d'anéchoïcité lors des mesures. Nous pensons que ces artefacts résultent de la structure métallique sur laquelle

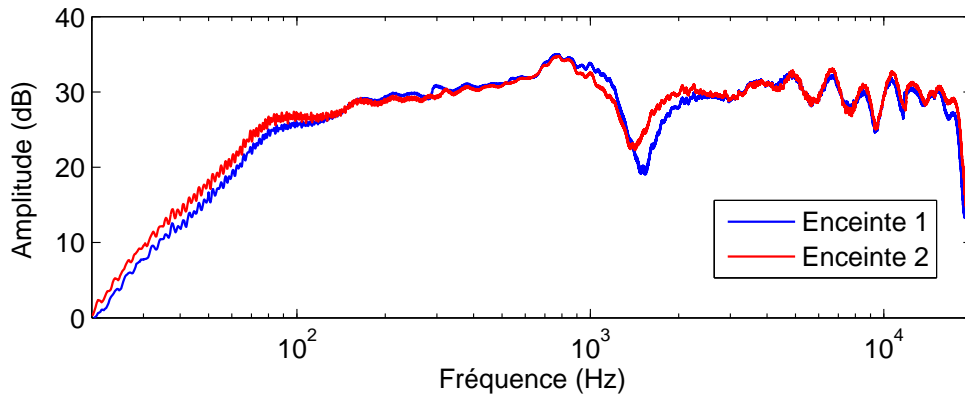


Figure 3.4 – Fonctions de transferts des deux enceintes utilisées pour la séparation des voies.

étaient placés le microphone et les enceintes. Ces mesures n'ont donc pas été effectuées dans des conditions rigoureusement normalisées, comme préconisé dans la norme IEC 60268-5 [1989], mais notre but n'était pas d'évaluer les performances absolues. Ces mesures permettent de valider la similitude des deux enceintes dans un même environnement et pour des conditions de restitution et d'acquisition identiques. Ceci est largement suffisant pour l'approche envisagée.

3.3 Modèle de boomer

Le modèle de haut-parleur substitué au rayonnement du boomer est basé sur la méthode proposée par Novák *et al.* [2010]. A partir de l'enregistrement du haut-parleur, cette méthode permet, dans une première phase d'analyse, d'identifier le comportement linéaire et non linéaire du haut-parleur. La deuxième étape consiste en la synthèse de la réponse de ce haut-parleur, basée sur une décomposition à l'aide des polynômes de Chebyshev. Des filtres précédemment identifiés d'après la mesure du haut-parleur sont associés à la décomposition, afin d'obtenir un modèle synthétisant le haut-parleur. La figure 3.5 présente le diagramme illustrant le modèle que nous utilisons. C'est un modèle composé de plusieurs branches, chacune d'entre elles associant un polynôme de Chebyshev et un filtre linéaire. La première branche du système correspond au comportement linéaire du boomer. Les autres branches de la décomposition associent des polynômes d'ordre supérieurs pour synthétiser la réponse non linéaire du haut-parleur.

3.3.1 Analyse : Identification des non-linéarités

Méthode d'identification : Travaux de Farina

Pour identifier le comportement non linéaire du boomer, la méthode d'analyse employée est basée sur la méthode de convolution présentée par Farina [2000]. Cette méthode consiste à analyser la réponse d'un système non linéaire en utilisant un "sweep exponentiel" en tant que

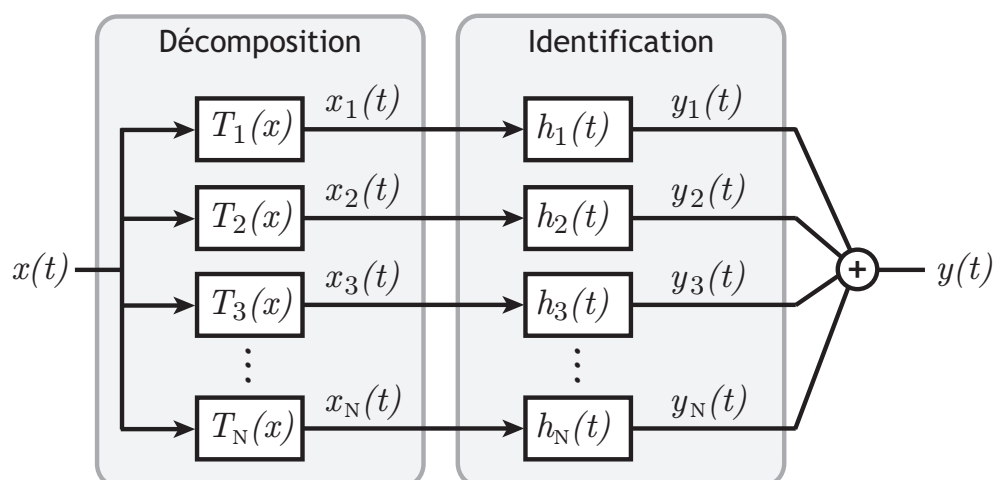


Figure 3.5 – Diagramme présentant le modèle de synthèse permettant d’associer le signal d’entrée et les réponses identifiées du boomer.

signal d’entrée $s(t)$. La fréquence instantanée de ce signal augmente exponentiellement avec le temps. Le signal $y_s(t)$ récupéré à la sortie du système non linéaire est ensuite convolué avec un filtre $\tilde{x}_s(t)$ correspondant approximativement au filtre inverse de $x_s(t)$: ce filtre est généré de telle sorte que la convolution de $x_s(t)$ avec $\tilde{x}_s(t)$ s’approche au mieux de la fonction Dirac $\delta(t)$ [Farina, 2000]. Dans notre cas, ce filtre inverse est obtenu sous la forme de la réplique du sweep d’entrée retourné temporellement, et avec une modulation d’amplitude permettant de prendre en compte la différence d’énergie entre les basses et les hautes fréquences. La figure 3.6 présente les étapes de l’identification du boomer de l’enceinte.

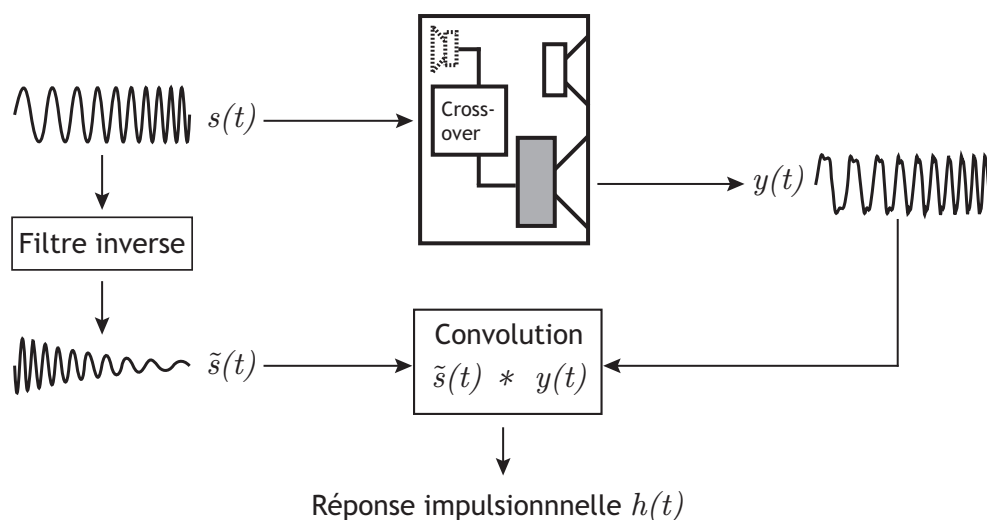


Figure 3.6 – Diagramme présentant la méthode d’identification du boomer. Le haut-parleur en pointillés représente le haut-parleur qui est envoyé à l’extérieur de la chambre anéchoïque avec la technique de séparation de voies. Figure adaptée de Novák [2009].

La convolution du filtre inverse $\tilde{x}_s(t)$ et du signal de sortie $y_s(t)$ fournit un signal $h(t)$ caractéristique du système. Pour un système linéaire, le signal $h(t)$ correspond à sa réponse impulsionnelle. Pour un système non linéaire, le résultat peut être considéré comme la somme d'une série de réponses élémentaires (3.1).

$$h(t) = y(t) * \tilde{x}_s(t) = \sum_N^{\infty} h_N(t + \Delta t_N). \quad (3.1)$$

La réponse $h(t)$ présente alors plusieurs maxima, dont le premier à $t = 0$ correspond à celui de la réponse impulsionnelle de la partie linéaire du système. Les suivants sont des artefacts résultant des produits de distorsion d'ordre 2 à N. La figure 3.7 illustre graphiquement le résultat de la convolution entre $y(t)$ et $\tilde{x}_s(t)$ pour l'identification d'un système non linéaire jusqu'à l'ordre 5.

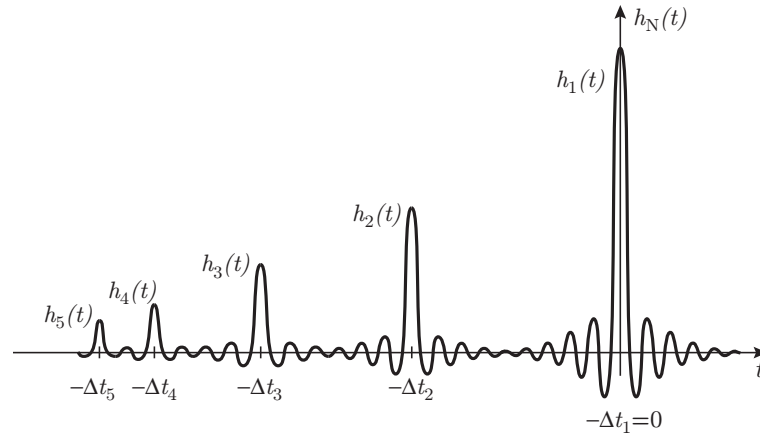


Figure 3.7 – Représentation des réponses impulsionnelles d'ordre supérieur (d'après Novák [2009]).

Chaque réponse est précisément localisée dans le temps et le décalage Δt_N entre la première et la N-ième réponse impulsionnelle dépend des caractéristiques du sweep généré et de l'ordre considéré. La relation 3.2 donne le Δt_N pour un sweep d'une durée T balayant l'intervalle f_1 à f_2 .

$$\Delta t_N = T \frac{\ln(N)}{\ln\left(\frac{f_2}{f_1}\right)} \quad (3.2)$$

Le décalage augmente avec le logarithme de N, donc les réponses des ordres élevés sont plus rapprochées [Farina, 2000]. Pour séparer les réponses, il est possible de les fenêtrer, à condition que chaque $h_N(t)$ décroisse suffisamment rapidement pour ne pas empiéter sur la réponse suivante.

d Améliorations de la méthode : Travaux de Novák *et al.*

La méthode initialement proposée par Farina [2000] permet d'estimer la distorsion d'un système non linéaire mais son identification effective a été proposée par Novák *et al.* [2009]. Cette méthode a ainsi été améliorée afin de synchroniser les phases des réponses fréquentielles d'ordre supérieur en affinant les propriétés du sweep exponentiel. Cette synchronisation est nécessaire pour pouvoir séparer les réponses $h_N(t)$ d'un système d'une manière compatible avec un modèle de décomposition polynomiale [Novák *et al.*, 2009].

La méthode de Novák *et al.* [2009] permet ainsi d'isoler précisément les réponses temporelles qui peuvent aussi être exprimées dans le domaine fréquentiel par la transformée de Fourier. Les réponses fréquentielles résultantes $H_N(f)$ expriment ainsi le comportement fréquentiel du système en terme de produits de distorsion d'ordres supérieurs. La figure 3.8 montre ces réponses fréquentielles jusqu'à l'ordre 5. La réponse $H_1(f)$ correspond à la partie linéaire du système, et pour les ordres plus élevés ($N > 1$), les $H_N(f)$ correspondent à la partie non linéaire.

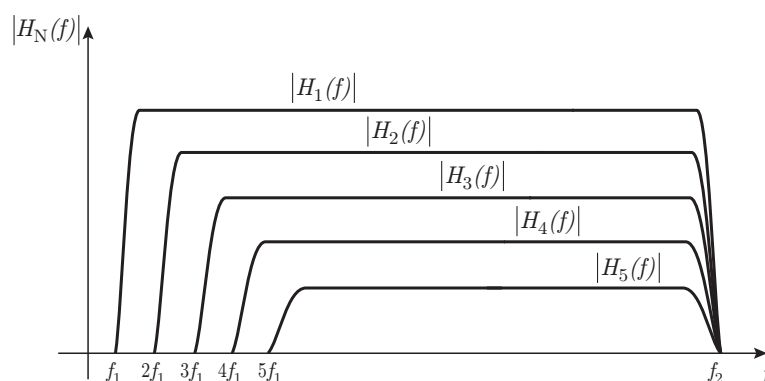


Figure 3.8 – Représentation des réponses impulsionnelles d'ordre supérieur (d'après Novák [2009]).

De plus, Novák *et al.* [2009] ont proposé d'étendre le support temporel du filtre inverse afin que chaque réponse fréquentielle identifiée d'ordre N soit représentée sur l'intervalle $[Nf_1, Nf_2]$. Le nombre maximum d'ordre qu'il est possible d'identifier est alors donné par le théorème de Nyquist-Shannon qui relie la fréquence maximale de l'identification et la fréquence d'échantillonnage de l'enregistrement du signal à la sortie du système. Le nombre de réponses identifiables est alors donné par la relation $Nf_2 < f_e/2$. Les réponses fréquentielles $H_N(f)$ obtenues avec cette modification sont représentées sur la figure 3.9.

3.3.2 Synthèse : Décomposition sur la base des polynômes de Chebyshev

Les réponses fréquentielles non linéaires obtenues à partir de la méthode d'identification améliorée par Novák *et al.* [2009] peuvent alors être utilisées pour la synthèse du modèle qui doit pouvoir simuler le comportement du boomer pour n'importe quel signal audio. La réponse non

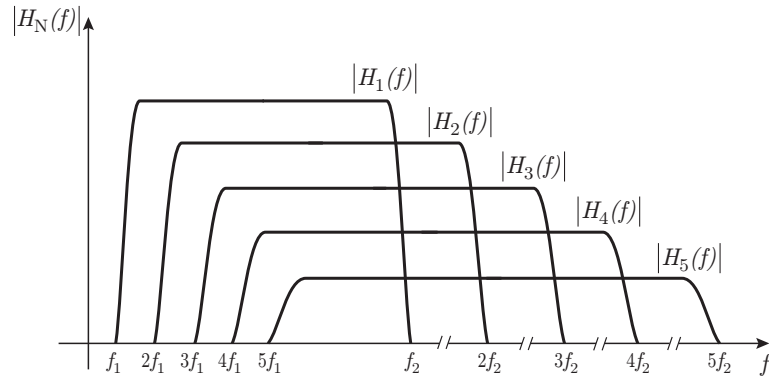


Figure 3.9 – Représentation des réponses fréquentielles d'ordre supérieur en prenant en compte l'extension du filtre inverse dans le processus d'identification (d'après Novák [2009]).

linéaire du boomer est ainsi exprimée sous forme d'une décomposition en harmoniques supérieures du signal sur la base des polynômes de Chebychev [Novák *et al.*, 2010]. Ces polynômes ont la particularité de former une base orthogonale sur l'intervalle $[-1,1]$ permettant une décomposition rapide et unique de n'importe quel signal d'entrée.

Les polynômes de Chebyshev sont définis par la relation [Thompson, 1994] :

$$T_n(x) = \cos(n\theta), \quad \theta = \arccos(x). \quad (3.3)$$

Ils peuvent être exprimés par la relation de récurrence suivante :

$$T_n(x) = 2xT_{n-1}(x) - T_{n-2}(x), \quad \text{pour } n \geq 2 \quad (3.4)$$

avec les conditions initiales :

$$T_0(x) = 1, \quad T_1(x) = x. \quad (3.5)$$

Les expressions des premiers polynômes de Chebyshev (pour $n = 1$ à $n = 6$) sont données ci-dessous et leurs représentations graphiques sont présentées sur la figure 3.10.

$$\begin{aligned}
 T_0(x) &= 1 \\
 T_1(x) &= x \\
 T_2(x) &= 2x^2 - 1 \\
 T_3(x) &= 4x^3 - 3x \\
 T_4(x) &= 8x^4 - 8x^2 + 1 \\
 T_5(x) &= 16x^5 - 20x^3 + 5x \\
 T_6(x) &= 32x^6 - 48x^4 + 18x^2 - 1
 \end{aligned}$$

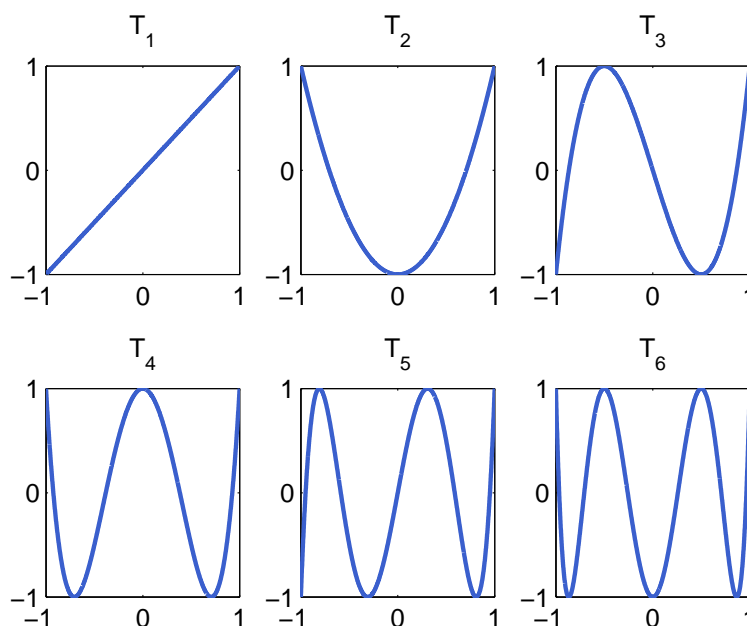


Figure 3.10 – Représentation des polynômes de Chebyshev T_n pour $n = 1$ à $n = 6$

3.3.3 Modélisation du boomer

Dans un premier temps, l'identification du boomer a été effectuée à partir de la méthode améliorée par Novák *et al.* [2009]. Le boomer de l'enceinte a donc été enregistré dans une chambre anéchoïque en appliquant la méthode de séparation des voies. Le signal d'entrée était un sweep choisi pour couvrir la plage de fréquence de 1 à 4400 Hz. La fréquence la plus élevée du sweep était ainsi, bien supérieure à la fréquence de coupure du filtre passif de l'enceinte située à 1800 Hz.

La figure 3.11 présente le montage utilisé pour l'identification du boomer précisant la position du microphone. La mesure a été effectuée avec un microphone de mesure à une distance de 2 mm du plan de la suspension périphérique. Cette mesure en champ proche permet d'obtenir un meilleur rapport signal/bruit qu'avec une mesure effectuée à une distance plus importante et correspond directement au débit du haut-parleur. Le boomer a été enregistré dans une chambre anéchoïque au moyen d'un microphone 1/4" G.R.A.S 40BF suivi d'un pré-amplificateur G.R.A.S. 26AC. Les signaux étaient émis et enregistrés grâce à un convertisseur audio Fostex VC-8 couplé à une carte son RME DIGI9652 avec une fréquence d'échantillonnage de 44100 Hz. La restitution du sweep et son enregistrement ont été effectués de manière synchrone, ce qui est indispensable pour effectuer l'opération de convolution entre l'entrée et la sortie du système.

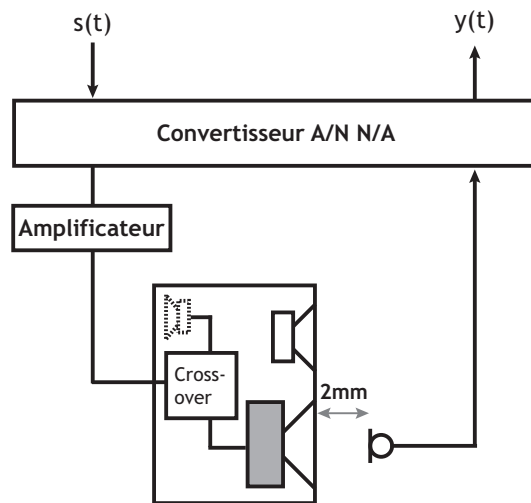


Figure 3.11 – Mesures effectuées pour identifier le boomer de l'enceinte. Le convertisseur Analogique/Numérique Numérique/Analogique permettait une mesure synchrone entre le sweep d'entrée et celui enregistré.

Pour l'identification du boomer, la fréquence d'échantillonnage étant de 44100Hz, les réponses non linéaires ont donc été estimées jusqu'à l'ordre 5. Une fréquence d'échantillonnage supérieure permettrait une identification à des ordres plus élevés mais la carte son que nous avons utilisée était limitée à $f_e=44100$ Hz. En raison d'un accès limité à la chambre anéchoïque, l'identification du boomer n'a été effectuée que pour un seul niveau du sweep d'entrée.

Les 5 réponses $h_n(f)$ identifiées caractérisant le comportement du boomer ont été introduites dans le modèle présenté sur la figure 3.5. Le modèle de boomer peut alors être utilisé pour simuler un signal proche de celui qui aurait été retransmis par le boomer identifié.

3.3.4 Vérification objective du modèle de boomer

La comparaison entre le signal réellement enregistré avec le boomer et le signal provenant du modèle, pour un même signal d'entrée, permet de vérifier de manière objective l'efficacité

du modèle à reproduire le comportement du boomer réel. Cette vérification a essentiellement pour but de valider la chaîne d'analyse-synthèse de réponse non linéaire. Nous avons effectué deux vérifications en fonction du signal d'entrée appliqué : la première fait intervenir des extraits musicaux et la deuxième vérification a été réalisée avec un sweep.

3.3.4.a Vérification avec des extraits musicaux

La première vérification a été effectuée avec des signaux musicaux en comparant le signal enregistré à celui qui a été synthétisé. La figure 3.12 présente le principe de comparaison entre le signal SM issu du modèle et le signal réel SR enregistré. Le signal SR a été enregistré dans des conditions identiques à l'identification : l'enregistrement était réalisé avec un microphone de mesure 1/4" G.R.A.S. 40BF et un pré-amplificateur G.R.A.S. 26AC situé dans l'axe des haut-parleurs à une distance axiale de 2 mm comme indiqué sur la figure 3.12. La restitution était réalisée avec un convertisseur audio Fostex VC-8 et une carte son RME DIGI9652 et l'acquisition était effectuée avec une carte Focusrite Saffire Pro 24, le tout avec une fréquence d'échantillonnage de 44100 Hz.

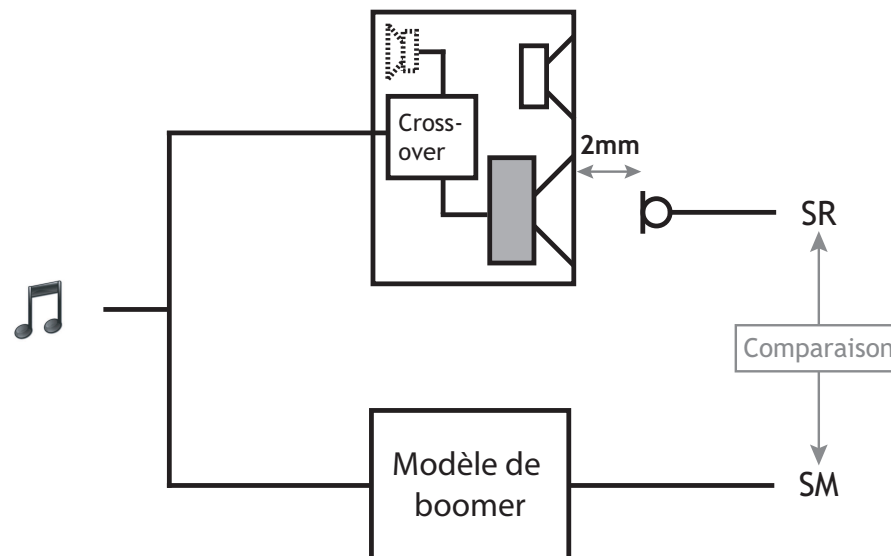


Figure 3.12 – Vérification du modèle de boomer. Comparaison entre le signal modélisé et le signal réellement enregistré pour différents extraits musicaux.

Au total, 9 extraits de morceaux de musique ou de parole ont été enregistrés et passés à travers le modèle :

- Francisco Guerrero, "Requiem". Durée : 50s.
- Kan'nida, "Konsyans". Durée : 46s.
- Mc Coy Tyner, "Miss Bea", Durée : 1min.06s.
- Orchestre des contre basses, Les Cargos, "Noire est la nuit...". Durée : 55s.

- Keith Jarret, "Köhlh Koncert". Durée : 52s.
- Rebecca Pidgeon, "Grandmother". Durée : 41s.
- Philippe Paindavoine. Durée : 1min.
- Santa Cruz, "Astrée-Auvidis". Durée : 47s.
- Georges Canévet, voix parlée enregistrée. Durée : 43s.

Le critère de comparaison est basé sur l'erreur quadratique moyenne MSE (pour Mean Square Error), calculée pour chaque paire d'un même extrait. La MSE correspond à la valeur moyenne de l'erreur estimée échantillon par échantillon entre les deux signaux [Wang et Bovik, 2009]. Nous avons utilisé une adaptation de ce critère, que nous appelons NMSE, qui permet d'obtenir une valeur normalisée de l'erreur comprise entre 0 et 1. Le résultat est soit 1 si les deux signaux sont différents sur la totalité de leur durée et 0 si les deux signaux sont identiques (la valeur maximale de ce critère peut atteindre 2 lorsque le critère est calculé sur deux signaux étant en opposition de phase). Ce critère, défini par l'équation 3.6, permet ainsi une mesure globale de l'erreur sur les N échantillons des deux extraits.

$$MSE(x, y) = \frac{\sum_{i=1}^N (x_i - y_i)^2}{\sum_{i=1}^N (x_i)^2 + \sum_{i=1}^N (y_i)^2}. \quad (3.6)$$

La valeur NMSE moyenne sur l'ensemble des 9 extraits est de 0.0131 qui correspond à une erreur relative de -18 dB. Ce résultat indique que le modèle reproduit assez bien mais pas exactement le signal du boomer. Cette erreur est commentée dans le paragraphe après la section concernant la deuxième vérification.

Pour donner un exemple visuel, la figure 3.13 présente un signal temporel de boomer enregistré et ce même signal synthétisé. Pour cette représentation sur une durée de 0.5 s, il est difficile de distinguer les deux courbes. L'étape d'identification permet donc de caractériser correctement le comportement du boomer et la synthèse basée sur une décomposition en polynômes avec seulement 5 coefficients semble retransmettre assez fidèlement le signal musical.

3.3.4.b Vérification avec un sweep

La deuxième vérification est basée sur l'utilisation d'un sweep couvrant la plage fréquentielle de 1 à 4400 Hz, identique à celui employé lors de l'identification du comportement du boomer. La figure 3.14 présente le principe de cette vérification avec un sweep. La comparaison entre les réponses impulsionnelles $h_R(t)$ et $h_M(t)$ permet de vérifier que le modèle de synthèse reconstruit toutes les informations temporelles et fréquentielles dans le sweep synthétisé. La corrélation linéaire de Pearson a été calculée entre les 5 premières réponses de l'identification. Les résultats sont présentés dans le tableau 3.1.

Les valeurs des corrélations entre les harmoniques supérieures sont assez élevées, ce qui indique

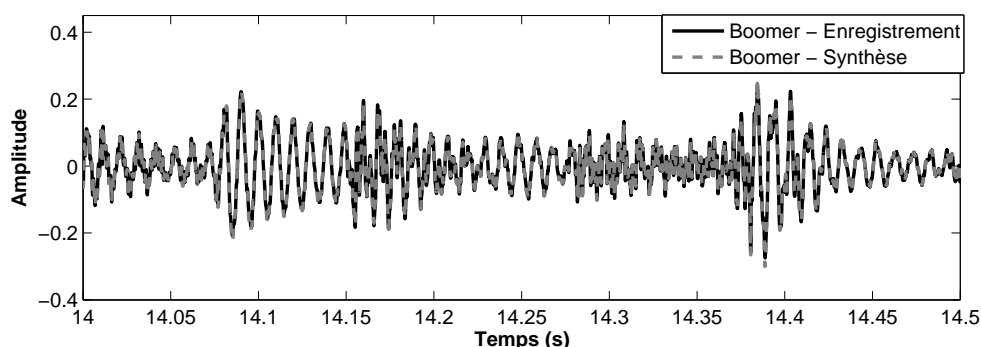


Figure 3.13 – Représentation temporelle d’un extrait enregistré à travers un boomer réel et synthétisé avec le modèle de boomer. L’extrait entier dure 46 s et la valeur RMS sur la totalité du signal est 0.0777.

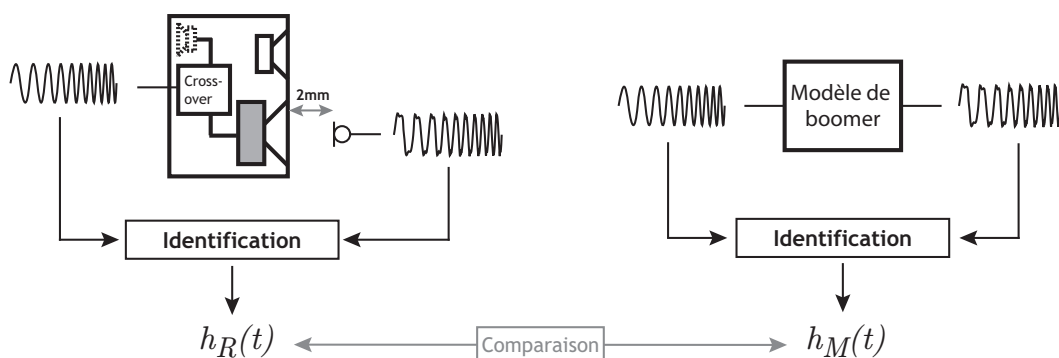


Figure 3.14 – Comparaison des réponses impulsionnelles obtenues à partir du boomer réel et à partir du modèle de boomer.

Ordre N	Corrélation entre les $h_R(t)$ et $h_M(t)$
1	0.989
2	0.980
3	0.964
4	0.979
5	0.997

Tableau 3.1 – Comparaison entre les réponses impulsionnelles obtenues à partir de la mesure $h_N R(t)$ et celles obtenues à partir du modèle $h_N M(t)$.

que la réponse au sweep synthétisée à travers le modèle de boomer est très proche de l’enregistrement. Cette vérification montre ainsi que peu d’informations sont perdues entre l’identification et la synthèse d’un signal. Cependant, la somme RMS des écarts par rapport à des sweeps identiques correspond à une erreur relative de -27 dB.

L’erreur NMSE évaluée sur les extraits sonores et l’erreur évaluée sur les réponses impulsion-

nelles sont du même ordre de grandeur. Ces deux vérifications montrent que le modèle obtenu est proche du boomer réel mais leur différence pourrait être audible. Étant donné que ce modèle de boomer est ensuite associé à l'enregistrement du tweeter, il se peut que cette erreur n'ait pas d'influence notable sur le signal de l'enceinte recomposée. De plus, pour le test final, nous allons introduire de la distorsion clairement audible qui masquera a priori cette erreur entre le modèle et le transducteur réel. Nous avons considéré donc à ce stade le modèle de boomer comme suffisant. De plus, une vérification perceptuelle de l'enceinte recomposée (modèle de boomer et enregistrement du tweeter) par rapport à l'enceinte originale a été effectuée afin de s'assurer du réalisme de l'élaboration d'échantillons synthétisés. Cette vérification est présentée dans la section 3.6.

3.4 Modèle d'enceintes virtuelles

Nous pouvons définir une "enceinte virtuelle", désignant en fait un échantillon sonore, comme le signal issu de la combinaison du modèle de boomer et de l'enregistrement du tweeter. Ainsi, nous cherchons à recréer le signal tel qu'il aurait été enregistré en face de l'enceinte simulée.

L'élaboration de l'enceinte virtuelle se fait à partir des différentes étapes récapitulées sur la figure 3.15. Pour l'étape 1, le signal du boomer provient du modèle que nous avons précédemment décrit et l'enregistrement du tweeter est lui détaillé dans cette partie. Les sweeps courts avant le signal symbolisés sur la figure permettent l'étape de synchronisation.

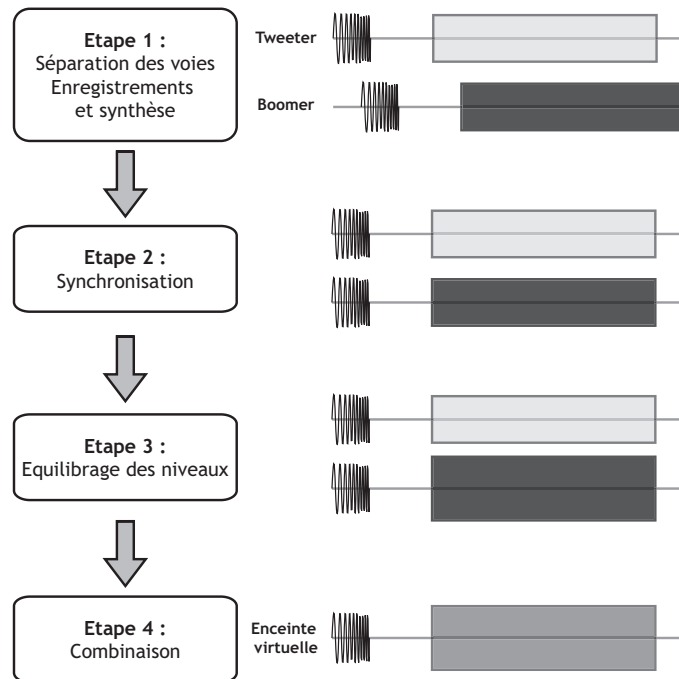


Figure 3.15 – Les différentes étapes dans l'élaboration de l'enceinte virtuelle

Pour développer les différentes étapes de l'élaboration, nous utilisons plusieurs signaux issus

de divers enregistrements ou du modèle du boomer. La figure 3.16, répertorie l'obtention de ces différents signaux.

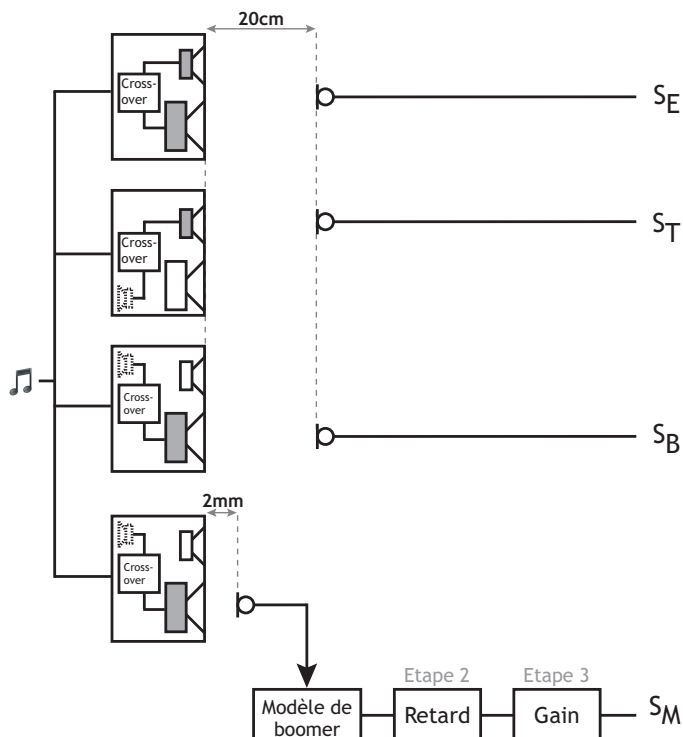


Figure 3.16 – Différents enregistrements utilisés pour l'élaboration de l'enceinte virtuelle. La représentation des haut-parleurs séparés a été choisie pour faire apparaître une différence entre l'enregistrement du boomer et du tweeter. En réalité, l'enceinte est coaxiale.

3.4.1 Enregistrement du tweeter

Nous avons enregistré le signal rayonné par le tweeter avec un microphone de prise de son Texen "Statex P48" positionné à une distance de 20 cm dans l'axe des haut-parleurs (S_T sur la figure 3.16). Le signal musical était envoyé au tweeter à l'aide d'un convertisseur audio Fostex VC-8 et une carte son RME DIGI9652 et l'enregistrement était réalisé par une carte son Focusrite Saffire Pro 24, le tout avec une fréquence d'échantillonnage de 44100 Hz.

3.4.2 Synchronisation des voies

Les deux signaux du boomer et du tweeter ont été obtenus à partir de deux enregistrements indépendants effectués à des distances différentes. Le premier provient de l'enregistrement du tweeter et le deuxième correspond au modèle de boomer. Pour reconstruire le son rayonné par l'enceinte, les deux voies ont donc été associées en recalant temporellement les deux signaux. Pour cela, lors des enregistrements des voies séparées, nous avons ajouté un sweep avant chaque

morceau pour permettre ce recalage a posteriori. Ce sweep s'étendait de 300 à 9000 Hz, pour pouvoir recouvrir les domaines fréquentiels du boomer et du tweeter (la fréquence de coupure du filtre étant de 1800Hz). La méthode que nous avons utilisée est basée sur le calcul d'intercorrélation entre deux signaux $x(t)$ et $y(t)$. Le maximum de cette fonction C_{xy} indique le décalage entre x et y. Son calcul est alors effectué à partir du sweep issu du modèle de boomer et celui issu de l'enregistrement du tweeter.

Afin d'illustrer la précision du calage temporel, nous avons mesuré l'erreur NMSE entre une enceinte réelle et l'enceinte "virtuelle", respectivement S_E et $S_M + S_T$ sur la figure 3.16. A partir du signal du tweeter (S_T) et celui du boomer (S_M) de l'enceinte virtuelle calés avec la technique d'intercorrélacion, nous avons introduit un décalage temporel entre les deux signaux puis calculé la NMSE pour chaque valeur du décalage supplémentaire. Nous avons mesuré cette NMSE pour deux configurations. Dans un cas, le signal du boomer est fixé et celui du tweeter est décalé temporellement. Dans l'autre cas, le signal du tweeter est fixé et celui du boomer subit un décalage. Dans les deux cas, les deux signaux sont ensuite recombines pour recréer l'enceinte virtuelle qui est comparée à l'enceinte réelle S_E . Cette illustration du calage temporel a été effectuée sur l'extrait du morceau Kan'nida sur une durée totale de 46 s.

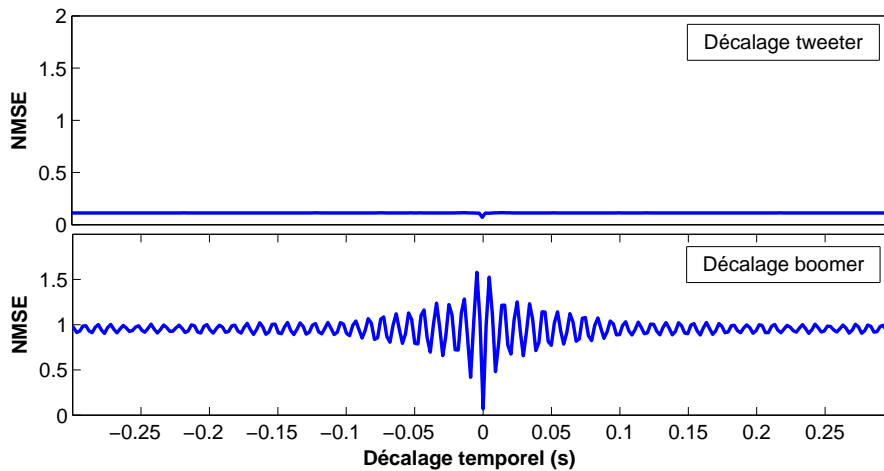


Figure 3.17 – Valeurs de l'erreur NMSE entre les signaux de l'enceinte réelle et de l'enceinte "virtuelle" pour différentes valeurs de décalage temporel entre le signal du tweeter et du boomer de l'enceinte synthétisée.

La figure 3.17 montre les valeurs de NMSE en fonction du décalage introduit sur le signal du tweeter (en haut) et sur le signal du boomer (en bas). Les deux figures indiquent que la valeur minimale de NMSE est obtenue quand les signaux du tweeter et du boomer coïncident temporellement. De plus, lorsque les deux signaux sont décalés, la valeur de NMSE est largement inférieure lorsque le décalage est appliqué sur le signal du tweeter. La forme temporelle du signal

de l'enceinte virtuelle reconstituée est régie principalement par le contenu du signal du boomer. Lorsque le décalage est appliqué au signal du boomer, la NMSE est proche de zéro quand le boomer et le tweeter sont calés et oscillent autour de 1 pour les autres valeurs de décalage. Ces oscillations sont dues au caractère périodique de l'extrait musical Kan'nida qui est un morceau de percussion africaine.

Les valeurs minimales de NMSE obtenues ainsi illustrent l'efficacité de la technique de calage temporel basée sur l'intercorrélacion qui permet bien de synchroniser avec précision les 2 voies de l'enceinte "virtuelle".

3.4.3 Équilibrage des niveaux

Pour obtenir une enceinte virtuelle similaire à l'enceinte originale, les niveaux des deux voies ont dû être équilibrés. Cette opération est nécessaire afin de garantir que l'enceinte reconstituée ne présente pas une contribution plus importante du tweeter ou du boomer par rapport à l'équilibre original entre les voies.

Pour ajuster les niveaux des deux voies nous nous sommes basés sur des enregistrements de référence. Nous avons enregistré dans les conditions identiques à celles de l'enregistrement du tweeter S_T , le boomer S_B et l'enceinte complète S_E . Les signaux S_E , S_B et S_T sont tous les trois issus de l'enregistrement avec le microphone Statex. Nous les avons utilisés pour équilibrer le niveau de l'enceinte "virtuelle" constitué du tweeter S_T et du signal issu du modèle de boomer que nous appelons S_M . La figure 3.16 illustre l'obtention des différents signaux.

Pour l'équilibrage des niveaux, nous avons pris comme niveau de référence, le niveau RMS mesuré sur les enregistrements de l'enceinte complète S_E . La combinaison des prises de son du tweeter S_T et du boomer S_B permet de recréer l'enceinte complète et donne donc un niveau RMS identique au niveau RMS de l'enceinte S_E . Pour l'enceinte virtuelle ($S_M + S_T$), le niveau du modèle de boomer S_M a donc été équilibré par rapport au niveau mesuré sur l'enregistrement du boomer S_B . Le niveau de l'enceinte virtuelle reconstituée est ainsi similaire au niveau de l'enceinte complète enregistrée ($S_E = S_B + S_T = S_M + S_T$).

3.5 Auralisation

Les enregistrements du tweeter et celui du boomer, qui ont permis d'élaborer le modèle, ont été effectués en chambre anéchoïque en un seul point. Pour pouvoir spatialiser ces signaux monophoniques d'enceintes virtuelles, nous avons utilisé la technique d'auralisation. Ceci consiste à convoluer le signal monophonique avec la réponse impulsionnelle d'une salle. Nous avons décidé d'appliquer la réponse impulsionnelle de la salle dans laquelle Lavandier [2005] a enregistré le deuxième corpus de son étude du timbre. Cette salle d'écoute appartenait au constructeur d'enceinte Mosquito et avait été mise à disposition pour le temps de la série d'enregistrements.

Nous nous sommes servis des enregistrements stéréophoniques réalisés par Lavandier [2005] avec la technique AB-ORTF pour obtenir la RI stéréophonique de cette salle d'écoute en utilisant la technique de convolution présentée par Farina [2000]. Nous avons donc effectué la convolution entre le filtre inverse associé au sweep original et l'enregistrement stéréophonique effectué par Lavandier [2005]. Des informations sur la salle d'écoute et des détails concernant la réponse impulsionnelle obtenue sont présentés en annexe C.

Nous avons cependant rencontré un problème pour isoler la RI de la salle. La RI obtenue par la méthode présentée ci-dessus inclut les caractéristiques de la salle mais aussi celles de l'enceinte. Idéalement, il aurait été préférable de déconvoluer la réponse de l'enceinte pour extraire la RI de la salle. Cela aurait été éventuellement possible en utilisant le même matériel (ordinateur, carte son, amplificateur, microphones, carte d'acquisition) et les conditions de l'acquisition dans la salle (distance entre enceinte et microphones) pour refaire cette mesure en chambre anéchoïque. Une fois la réponse de l'enceinte obtenue, il aurait été théoriquement possible de la déconvoluer dans la RI. Une telle mesure est par contre difficilement reproductible à l'identique, notamment en ce qui concerne la position de l'enceinte et des microphones. De plus, une nouvelle mesure dans la salle d'écoute de Mosquito n'a pas été possible car cette salle n'est plus disponible.

Nous avons donc tenté d'utiliser directement la RI identifiée dans la salle, que nous qualifions de RI "colorée" (par la réponse de l'enceinte). En l'appliquant à l'enceinte virtuelle, cela revient à faire agir la réponse de l'enceinte une deuxième fois. Une vérification perceptive est alors nécessaire pour voir si la convolution du signal monophonique de l'enceinte virtuelle avec la RI stéréophonique "colorée" est relativement comparable à l'enceinte réellement enregistrée dans la salle.

3.6 Vérification perceptive : enceinte virtuelle vs. enceinte réelle

Afin de vérifier le processus d'élaboration des enceintes virtuelles et de l'auralisation, deux tests d'écoute ont été menés pour comparer l'enceinte réelle et notre enceinte virtuelle, et ainsi évaluer l'influence de la RI que nous avons choisi d'utiliser. La figure 3.18 résume les différentes étapes menant à l'obtention des deux signaux à comparer.

Pour évaluer si une différence est perçue entre l'enceinte réelle et l'enceinte virtuelle, un test ABX a été mené. Le test ABX présente l'avantage d'être très discriminant par rapport à une quelconque différence audible entre des extraits sonores [Clark, 1982]. Pour chaque essai, sont présentés à l'auditeur 3 sons A, B et X. Les sons A et B sont deux sons différents (soit une enceinte réelle et une enceinte virtuelle, soit l'inverse). Le son X correspond soit au son A soit au son B. La tâche de l'auditeur consiste à dire si X est perçu identique à A ou à B. Deux tests ont été menés avec différents extraits et différents nombres de présentations.

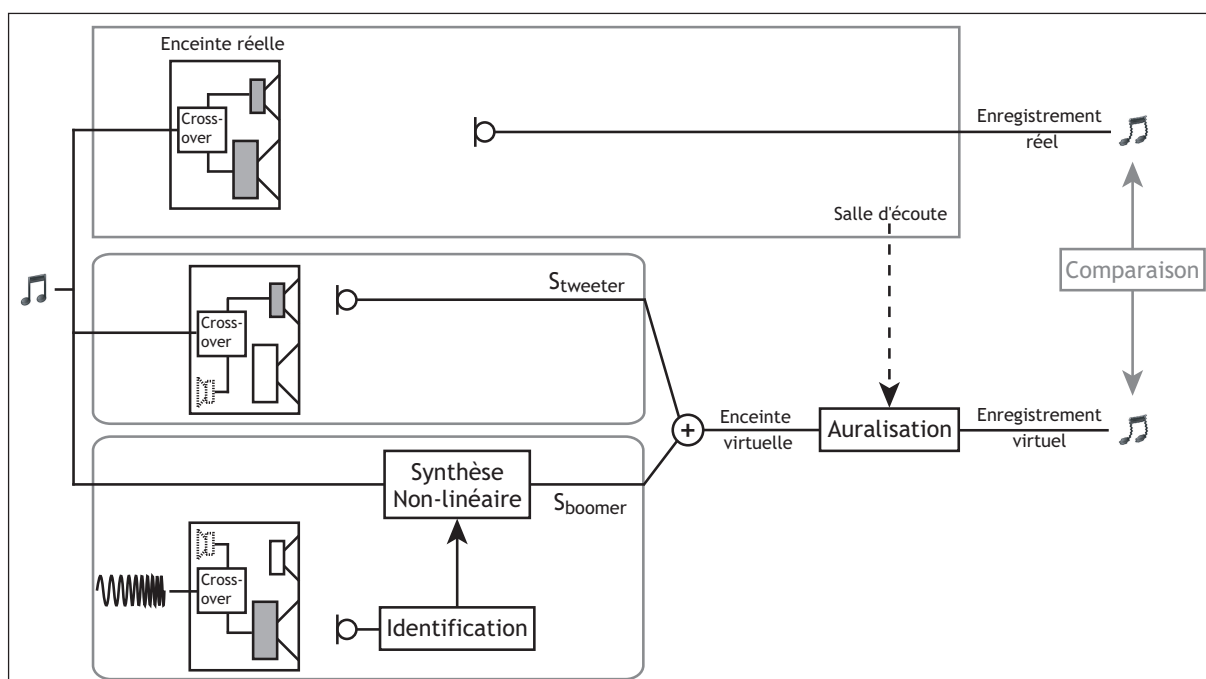


Figure 3.18 – Schéma présentant un résumé de l'élaboration de l'enceinte virtuelle, de l'auralisation et de la comparaison avec un enregistrement réel.

3.6.1 Procédure

Les tests ABX se sont déroulés à partir d'une interface composée de 3 boutons virtuels. Au cours de chaque essai, les trois sons sont joués successivement à l'auditeur. Ensuite, il peut réécouter les sons autant de fois que nécessaire avant de répondre, en pressant le bouton correspondant. Les tests se déroulaient dans une cabine audiométrique et les stimuli étaient joués via une carte son (M-Audio Delta 66) et reproduits au casque (Stax SR Lambda Professional).

3.6.2 Enregistrements réels et virtuels

Les enregistrements virtuels correspondent à plusieurs extraits musicaux obtenus avec le processus de création de l'enceinte virtuelle avec auralisation. Les extraits sont une portion de quelques secondes de certains morceaux. Les enregistrements réels correspondent aux mêmes extraits, joués à travers l'enceinte et enregistrés par Lavandier [2005] avec une technique stéréophonique AB-ORTF dans la salle d'écoute de Mosquito. La distance entre l'enceinte et les microphones était de 2,40 m. Des informations supplémentaires sur la salle d'écoute sont présentées en annexe C.

3.6.3 Tests ABX

Afin d'évaluer si une différence est perçue entre les enceintes réelle et virtuelle, nous avons mené deux tests ABX complémentaires. Ces deux tests permettent de mettre en avant la précision du test ABX en fonction du nombre de présentations soumises à l'auditeur au cours d'un test. Le nombre de stimuli, d'auditeurs et la description de leur conception sont décrits dans cette section.

TEST 1 :

Stimuli : Pour ce premier test ABX, 4 extraits ont été utilisés :

- Voix parlée : Georges Canévet, voix parlée enregistrée, Durée : 2.4s
- Chorale : Francisco Guerrero, "Requiem", Durée : 4s
- Musique classique : W.A. Mozart, Flûte Quartet, "Adagio", Durée : 3.2s
- Guitare et voix féminine : Rebecca Pidgeon, "Grandmother", Durée : 3.8s.

Conception : Pour chaque extrait, toutes les combinaisons de présentation des sons A, B et X (ABA, ABB, BAA et BAB) étaient évaluées. Le test entier comprenait 16 essais (4 extraits et 4 combinaisons) présentés à chaque auditeur dans un ordre aléatoire. Pour chaque auditeur, le test durait environ 10 minutes.

Auditeurs : Douze auditeurs ont participé au test 1. Ils étaient tous normo-entendants.

TEST 2 :

Stimuli : Un seul extrait a été utilisé pour ce deuxième test ABX. Il s'agissait d'un extrait du morceau de jazz de McCoy Tyner, "Miss Bea", qui durait 3s.

Conception : Dans ce test, nous avons voulu évaluer l'influence du nombre de présentations. Ainsi pour l'extrait testé, les 4 combinaisons possibles étaient répétées 5 fois. Pour chaque auditeur, le test entier comprenait alors 20 essais présentés à chaque auditeur dans un ordre aléatoire. La durée de ce deuxième test était aussi d'environ 10 minutes.

Auditeurs : Douze auditeurs ont participé au test 2. Ils étaient tous normo-entendants et 6 d'entre eux avaient participé au test 1.

3.6.4 Résultats

Le résultat d'un test ABX correspond au pourcentage de bonnes réponses données par l'auditeur. Si une différence est clairement audible entre les extraits réels et virtuels alors, le score atteint 100%. Si il n'y a pas de différence perçue entre les extraits, alors le résultat du test ABX avoisine 50%, ce qui correspond à la probabilité de répondre au hasard dans ce type de tâche. Pour chaque test, les résultats exposés représentent la moyenne des résultats des auditeurs.

Le premier test ABX fournit un score moyen pour tous les extraits confondus de 52% avec un écart-type de 16%. Le résultat du deuxième test est de 55% avec un écart-type de 13%. Pour comparer les résultats des deux tests, les valeurs moyennes et les écart-types obtenus pour chaque extrait sont présentés sur la figure 3.19. Les écart-types individuels du test 1 sont plus importants que les écart-types obtenus dans le test 2 avec un nombre de présentations cinq fois supérieur, ce qui illustre qu'il est préférable d'effectuer des tests ABX avec un nombre élevé de présentations afin de réduire l'incertitude des résultats.

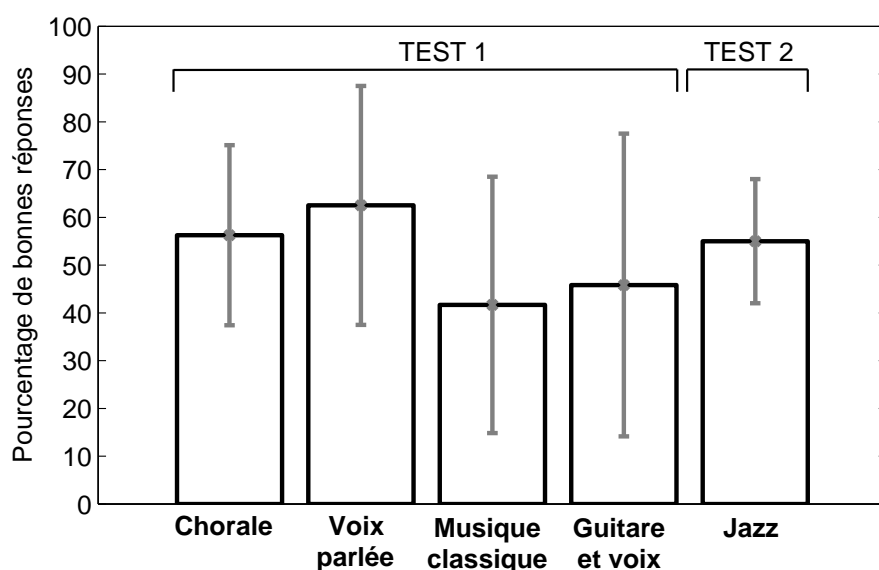


Figure 3.19 – Résultats des tests ABX. Les 4 premiers extraits étaient évalués lors du test 1 et l'extrait de jazz était évalué lors du test 2.

Il est intéressant de noter que parmi les auditeurs, un professionnel de l'audio a participé aux deux tests. Il s'est avéré qu'il percevait dans la majorité des présentations la différence entre l'enregistrement réel et l'enceinte virtuelle. Cette différence audible pour une personne entraînée à une écoute analytique pourrait provenir de la RI colorée que nous avons été amenés à utiliser pour l'auralisation. Cependant, tous les autres auditeurs ne percevaient en moyenne aucune différence. Les résultats moyens des deux tests ABX sont donc très encourageants et valident a posteriori le modèle du boomer (malgré l'erreur de synthèse remarquée), l'élaboration de l'enceinte virtuelle, et la convolution pour l'auralisation (malgré la coloration de la RI de la salle par l'enceinte).

A partir de la séparation des deux voies d'une enceinte, nous avons montré la possibilité de créer des enceintes virtuelles. Nous nous sommes assurés d'effectuer les opérations nécessaires de recalage temporel ou de normalisation pour simuler au mieux le comportement de l'enceinte originale. Les différentes validations objectives ont montré que la synthèse du boomer et l'association de l'enregistrement du tweeter avec le modèle de boomer fonctionnaient correctement. La

validation perceptive a permis d'évaluer le rendu des enceintes virtuelles après auralisation. Il semblerait que les enceintes virtuelles auralisées soient difficiles à discerner des enceintes réelles dans la même pièce. Ce procédé semble donc permettre de créer des stimuli très réalistes, pouvant même au besoin être comparés à des enregistrements d'enceintes réelles.

3.7 Contrôle des non-linéarités dans les enceintes virtuelles

Afin d'étudier la distorsion non linéaire, nous avons ensuite modifié l'enceinte virtuelle pour pouvoir introduire différents types de distorsion. Le principe est de s'écarter de l'enceinte originale X pour créer d'autres enceintes X' , X'' dont les caractéristiques sont modifiées de manière contrôlée. Pour créer un corpus présentant une grande variété de distorsions, nous avons utilisé différentes techniques pour générer la distorsion dans les enceintes virtuelles. La première consistait à modifier les non-linéarités préalablement identifiées à partir du boomer réel, pour rester proche du comportement d'une enceinte. Dans un deuxième temps, nous avons adapté le modèle pour pouvoir créer des échantillons sonores moins réalistes en introduisant des phénomènes artificiels sans mémoire basés sur des courbes entrée/sortie. Nous avons appliqué la technique utilisée par Tan *et al.* [2003] ou Geddes et Lee [2003a] pour simuler diverses non-linéarités artificielles appliquées à la réponse du boomer mais aussi à la réponse complète de l'enceinte.

3.7.1 Modification des non-linéarités identifiées sur le boomer

Notre première technique de création d'échantillons permet de modifier les non-linéarités identifiées sur le boomer. A partir de l'étape d'identification, nous avons obtenu 5 réponses associées à des polynômes non linéaires. En les introduisant dans le modèle de boomer, nous pouvons reproduire le son qu'aurait produit le boomer en un point pour n'importe quel extrait musical. Pour s'écarter du boomer original et modifier sa distorsion non linéaire, nous pouvons soit modifier les courbes de réponses fréquentielles, soit appliquer un coefficient à chaque réponse, soit attribuer un seul coefficient commun à toutes les réponses. Nous avons décidé de changer l'amplitude des réponses fréquentielles d'ordre 2 à 5.

Afin de garder un paramétrage le plus simple possible, nous avons choisi pour une première série de tests de modifier l'influence de toutes les réponses non linéaires par rapport à la réponse linéaire en appliquant à chaque branche un même coefficient α . Le modèle utilisé permettant de modifier les réponses identifiées est présenté sur la figure 3.20 pour les enceintes virtuelles référencées "I". Par rapport au modèle de boomer qu'a utilisé Klippel [2006] pour introduire les non-linéarités spécifiques, cette technique permet de modifier globalement l'influence des non-linéarités mesurées dans la restitution du boomer mais ne garantit pas que cette modification ait un sens physique.

Pour recréer l'enceinte virtuelle, le signal issu du modèle de boomer modifié est associé à l'enregistrement du tweeter en prenant en compte les étapes d'équilibrage des niveaux et de synchronisation présentées dans les sections 3.4. Il est important de préciser que les coefficients introduits pour modifier les réponses fréquentielles sont ensuite modifiés par l'équilibrage des niveaux, puis l'égalisation en sonie lors des tests d'écoute. Notre approche fixe donc uniquement un niveau relatif moyen entre la réponse linéaire et les réponses non linéaires. Les enceintes "I" ainsi obtenues correspondent donc à des variantes de l'enceinte originale dans laquelle la distorsion d'amplitude du boomer est graduée.

3.7.2 Introduction de non-linéarités artificielles

La deuxième technique de création d'enceintes virtuelles consiste à introduire des non-linéarités dites "artificielles", définies à partir de leurs courbes entrée/sortie. Ces non-linéarités artificielles n'ont aucun rapport avec le comportement de l'enceinte utilisée pour l'étude, mais permettent de créer une plus grande variété de signaux distordus. Nous avons cependant utilisé la même technique de synthèse que précédemment pour introduire des non-linéarités artificielles issues de différentes courbes entrée/sortie. L'idée est de combiner l'approche académique des courbes sans mémoire avec une réponse impulsionnelle pour conserver des stimuli proches du rayonnement d'une enceinte acoustique. Ce modèle mixte est composé de deux étapes.

Approximation des relations entrée/sortie :

Nous avons approximé les relations entrée/sortie pour pouvoir les intégrer dans le modèle polynomial. N'importe quelle fonction $f(x)$ continue, définie dans l'intervalle $[-1,1]$ peut être projetée sur la base des polynômes orthogonaux de Chebyshev. La représentation d'une telle fonction correspond alors à une série de coefficients associés à ces polynômes. Une approximation est possible en tronquant cette série à un ordre N donné. La relation entre la fonction et la série de Chebyshev s'écrit alors :

$$f(x) \approx f_N(x) = \sum_{k=0}^N a_k T_k(x). \quad (3.7)$$

Cette approximation d'une fonction sur la base des polynômes de Chebyshev permet d'obtenir un vecteur de coefficients correspondant aux ordres 1 à N , l'ordre N pouvant être choisi en fonction de la précision d'approximation souhaitée. Ainsi, les fonctions présentant des changements brusques de comportement (comme les non-linéarités de Clipping ou de Zero Crossing) nécessitent un grand nombre de coefficients. Le critère de troncature que nous avons choisi est basé sur une valeur "seuil" des coefficients définie arbitrairement. Nous déterminons dans la série le dernier coefficient supérieur à cette valeur "seuil" et tronquons la série au delà de cet ordre. Pour les différents types de non-linéarité que nous avons utilisés dans les tests d'écoute exposés

dans le chapitre 6, le nombre de coefficients de la série et l'erreur d'approximation sont présentés en fonction de la valeur "seuil" dans la dernière partie de l'annexe B. L'erreur d'approximation correspond à l'erreur absolue entre la courbe entrée/sortie et son approximation. A partir des résultats obtenus, nous avons choisi une valeur seuil égale à 0.1%, ce qui permet de limiter considérablement le nombre de polynômes nécessaires tout en gardant une erreur d'approximation relativement faible.

Les vecteurs de coefficients d'approximation a_N des relations entrée/sortie ont alors été insérés dans le processus de synthèse. Contrairement à la technique de synthèse basée sur l'identification du boomer, le nombre de branches du modèle dépend ici du nombre de coefficients nécessaires à l'approximation des courbes entrée/sortie.

Application d'un comportement temporel :

Les non-linéarités artificielles permettent de générer une multitude de phénomènes de distorsion dans le signal d'entrée, mais pour se rapprocher du son émis par une enceinte il est nécessaire d'ajouter une réponse impulsionnelle décrivant le comportement linéaire d'une enceinte. Nous avons donc appliqué ces non-linéarités artificielles de deux manières. D'un coté, les non-linéarités sont appliquée au boomer et de l'autre elles sont appliquées à la totalité de l'enceinte (signaux large bande, sans séparation du tweeter). L'idée provient de l'étude de Tan *et al.* [2003] qui ont mesuré la qualité perçue de différents types de non-linéarités sur la totalité du spectre du signal testé et sur certaines régions fréquentielles spécifiques. Dans notre étude, cette technique permet de comparer au sein d'un même test les mêmes phénomènes appliqués dans des régions fréquentielles différentes du fonctionnement d'une enceinte.

3.7.2.a Non-linéarités artificielles appliquées au boomer : Enceintes "B"

Le modèle que nous avons utilisé pour introduire les non-linéarités artificielles est présenté sur la figure 3.20. Les coefficients a_N proviennent de l'approximation des courbes entrée/sortie et sont appliqués aux différents polynômes de Chebyshev. Puis, pour se rapprocher du rayonnement du boomer, la réponse linéaire identifiée du boomer $h_1(t)$ est convoluée après l'étape approximation. Le résultat obtenu correspond à l'extrait musical distordu à partir de la non-linéarité choisie, mais avec un timbre proche de celui du boomer de l'enceinte "génératrice". Pour obtenir l'enceinte virtuelle, l'enregistrement du tweeter a été ensuite associé au signal distordu du boomer.

3.7.2.b Non-linéarités artificielles appliquées à l'enceinte complète : Enceintes "E"

Pour évaluer les distorsions sur une gamme fréquentielle plus étendue, nous avons appliqué les différentes non-linéarités au signal large bande, avant le filtre caractérisant le comportement de l'enceinte complète. Dans ce deuxième cas, le modèle était uniquement modifié en remplaçant la réponse fréquentielle du boomer par celle de l'enceinte $h(t)$. La réponse $h(t)$ provient de

l'enregistrement d'un sweep de 10 à 22050 Hz reproduit par l'enceinte complète en chambre sourde.

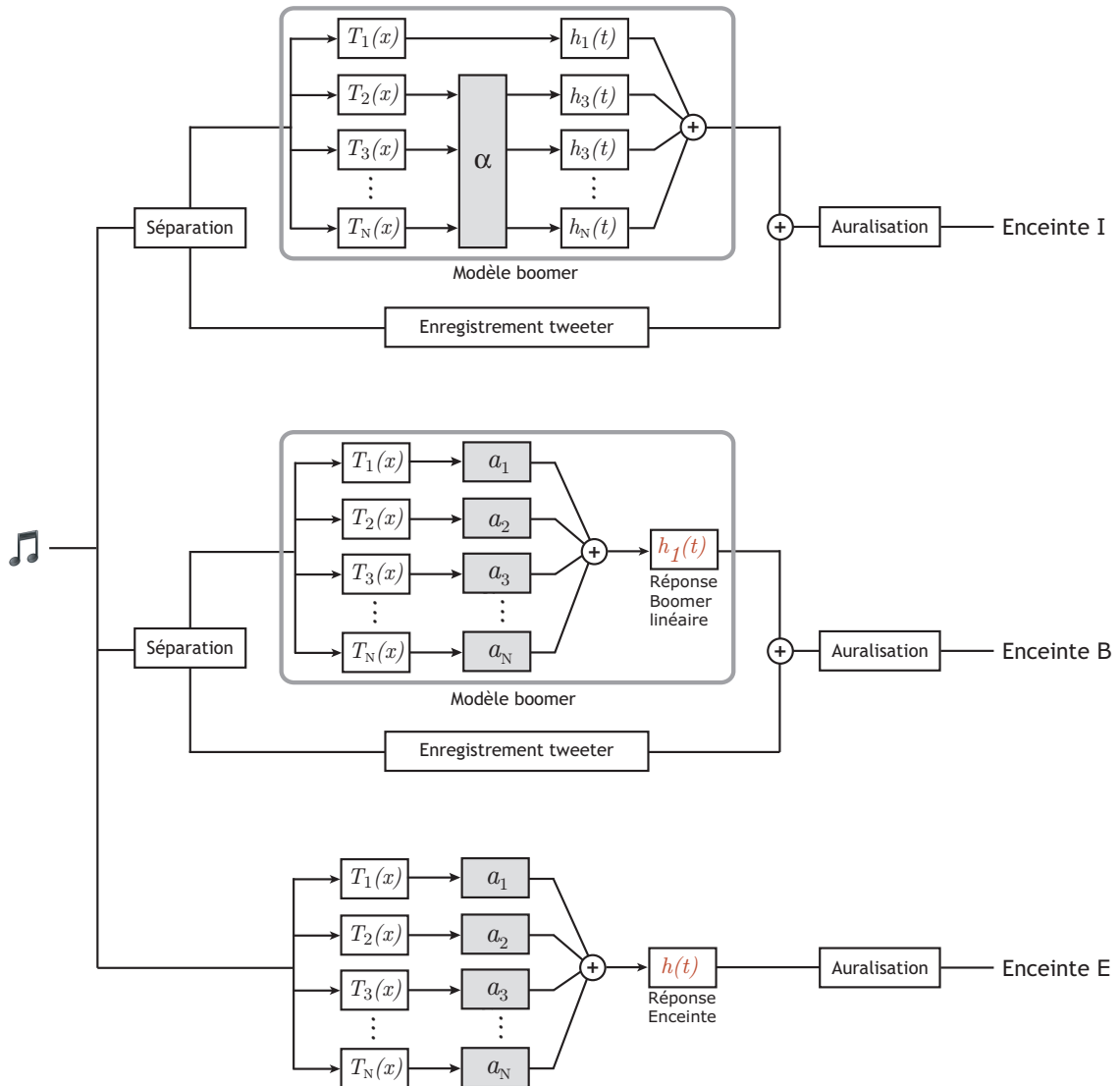


Figure 3.20 – Trois modèles de synthèse permettant de générer les enceintes I (issues de l'identification du boomer), les enceintes B (issues des courbes entrée/sortie appliquées au boomer) et les enceintes E (issues des courbes entrée/sortie appliquées à l'enceinte complète). Les deux premiers modèles de boomer sont associés à l'enregistrement du tweeter pour recomposer l'enceinte virtuelle.

A partir de ces trois techniques permettant d'introduire des non-linéarités dans les enceintes virtuelles, il est possible d'élaborer un large corpus d'enceintes présentant des types de distorsions différents. Ces enceintes virtuelles modifiées ont toutes été auralisées pour permettre l'évaluation réaliste dans un environnement adapté à l'écoute des enceintes acoustiques.

Chapitre 4

Méthode de comparaison avec permutation de références

Sommaire

4.1	Adaptation à l'évaluation de stimuli audio	83
4.2	Particularités de la méthode	84
4.3	Simulations	86
4.3.1	Simulation 1 : Influence du nombre de stimuli	88
4.3.2	Simulation 2 : Influence du nombre d'auditeurs	88
4.4	Résultats	88
4.4.1	Relation entre les dissemblances connues et estimées	88
4.4.2	Nombre de stimuli	89
4.4.3	Nombre d'auditeurs simulés	90
4.4.4	Caractérisation du biais	91

Au cours du chapitre 2, nous avons évoqué le besoin d'utiliser une méthode permettant l'évaluation de nombreux stimuli. La méthode que nous proposons est la méthode de comparaison avec permutation de références issue de l'étude de Rogowitz *et al.* [1998]. Dans ce chapitre nous présentons plus en détail la méthode de comparaison avec permutation de références proposée par Rao et Katz [1971] et utilisée par Rogowitz *et al.* [1998] lors d'un test de dissemblance d'images. Étant donné que le seul exemple d'application que nous avons est l'expérience de Rogowitz *et al.* [1998], nous avons, dans un premier temps, adapté cette méthode à l'évaluation de stimuli sonores. Ensuite nous avons réalisé des simulations de cette méthode pour pouvoir explorer ses limites et avoir une idée sur ses capacités à fournir des valeurs de dissemblances.

4.1 Adaptation à l'évaluation de stimuli audio

Afin d'adapter la méthode de comparaison avec permutation de références à l'évaluation de la dissemblance entre stimuli sonores, certaines précautions ont été prises. Dans l'expérience menée par Rogowitz *et al.* [1998], les participants ont une vision globale de la référence ainsi que des 8 images à comparer. Pour des tests d'écoute, les stimuli sont nécessairement présentés successivement, et les auditeurs doivent mémoriser les sons afin d'établir des comparaisons. Il semble ainsi nécessaire de réduire le nombre de stimuli présenté à chaque essai afin de prendre en compte la capacité de mémoire auditive, en cherchant un compromis entre la durée d'un essai et la durée du test entier.

Pour un corpus composé de n sons, chaque référence est comparée aux $(n - 1)$ autres sons de comparaison. Ces $(n - 1)$ sons sont distribués en p groupes de sons de comparaison à partir d'un tirage aléatoire effectué sans remise. Cette distribution correspond à une présentation partielle de toutes les possibilités de tirer p parmi les $(n - 1)$ sons. Le nombre p correspond au nombre de sons de comparaison évalués au cours d'un essai et son choix définit le nombre d'essais à effectuer pour chaque référence $((n - 1)/p)$. Chaque son apparaissant à son tour comme référence, le test entier est composé de $n \cdot (n - 1)/p$ sons.

Le tableau 4.1 présente l'influence de ce nombre p sur le nombre d'essais effectués pour chaque référence et sur le nombre total d'essais du test pour l'évaluation d'un corpus de 40 stimuli. Pour comparaison, l'évaluation d'un corpus de 40 stimuli avec la méthode de comparaison par paires conduit à un test entier de 780 essais.

Le nombre d'essais d'un test entier diminue rapidement avec le nombre de stimuli de comparaison présentés au cours d'un essai. Le choix du nombre de stimuli au sein d'un essai et le nombre total d'essais du test entier est un compromis pour prendre en compte les contraintes de l'évaluation de stimuli sonores et celles liées au design du test. Nous avons ainsi choisi d'adapter la méthode en comparant chaque référence avec 3 stimuli de comparaison. Ainsi, pour chaque référence, les $(n - 1)$ autres stimuli sont répartis en $(n - 1)/3$ trios. Étant donné que chaque

Nb de stimuli de comparaison	Nb. d'essais / référence	Nb. d'essais total
p	$(n - 1)/p$	$n.(n - 1)/p$
2	19	780
3	13	520
4	9	360
5	7	280
6	6	240

Tableau 4.1 – Nombre d'essais présentés pour chaque référence et nombre d'essais du test entier en fonction du nombre de stimuli de comparaison présentés au sein d'un essai.

stimulus apparaît comme référence, pour un corpus de n stimuli, un test entier est composé de $n.(n - 1)/3$ essais.

La figure 4.1 présente le déroulement du test en fonction du nombre de stimuli et du nombre d'auditeurs choisis pour un nombre de stimuli de comparaison fixé à 3. Le nombre d'essais indiqué sur la figure correspond au nombre d'essais à effectuer pour chaque stimulus de référence.

Pour chaque auditeur, les tests sont différents car la répartition des sons de comparaison en trios est aléatoire. Etant donné que la méthode ne permet qu'une présentation partielle de trios pour chaque référence, il est nécessaire de moyenniser les matrices individuelles de plusieurs auditeurs pour obtenir une matrice de dissemblance moyenne que nous analysons ensuite avec la technique de MDS.

4.2 Particularités de la méthode

La méthode de comparaison avec permutation de références dispose de plusieurs particularités qui nous semblent intéressantes pour l'évaluation de la dissemblance pour des corpus étendus de stimuli audio.

La tâche :

Au cours d'un essai il est demandé aux participants de choisir parmi les stimuli de comparaison celui qui semble le plus similaire à la référence. Cette tâche de regroupement permet de s'affranchir de l'estimation de la dissemblance entre deux stimuli sur une échelle linéaire, comme dans le cas d'une comparaison par paires. De plus, c'est une tâche à choix forcé car l'auditeur est obligé de choisir un stimulus comme réponse. Cette caractéristique permet d'encadrer les jugements des auditeurs par rapport à une tâche de classification libre.

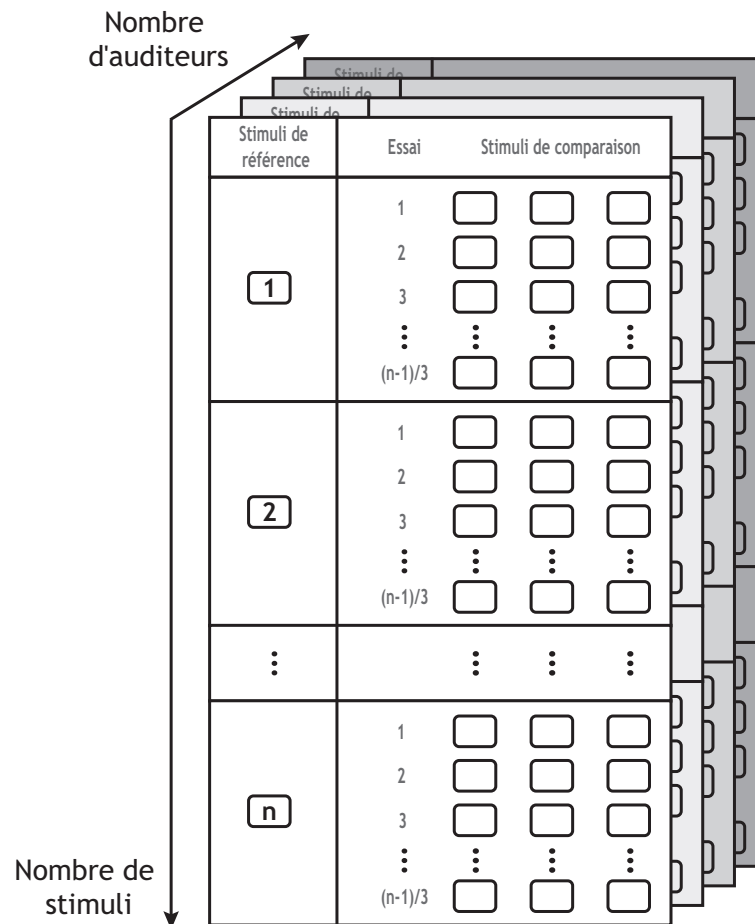


Figure 4.1 – Déroulement d'un test de comparaison avec permutation de références. Au sein d'un essai sont présentés un son de référence et 3 sons de comparaison.

La référence :

Certaines méthodes d'évaluation audio font aussi intervenir un stimulus de référence. C'est le cas de la méthode MUSHRA : "Multi Stimulus test with Hidden Reference and Anchors" [ITU BS.1534-1, 2001]. Cette méthode est utilisée pour estimer l'influence d'un codec audio sur un signal sonore. Les auditeurs doivent comparer tous les sons dégradés par le codec au signal original (référence). La tâche consiste à évaluer la qualité sonore sur une échelle allant de "mauvais" à "bon". Par contre, la méthode reste focalisée sur la qualité sonore et propose une évaluation par rapport à une référence absolue.

Notre méthode de comparaison avec permutation de références présente l'avantage d'utiliser chaque stimulus comme référence. Il n'est pas nécessaire de définir un stimulus comme la référence absolue et de juger la dissemblance par rapport à ce stimulus. La présence d'une référence à chaque essai rend les essais indépendants les uns des autres et permet de réduire le biais séquentiel évoquée par Poulton [1979]. Cette caractéristique nous permet également de diviser le test entier en plusieurs sessions sans craindre que les auditeurs ne changent

d'échelle d'évaluation entre les sessions.

La matrice de dissemblance :

La matrice de dissemblance est remplie d'une manière identique à celle de la classification libre ou celle de la méthode des triades. La matrice de similarité s initialisée à 0 est incrémentée en fonction de la réponse de l'auditeur. La matrice de dissemblance individuelle correspond à l'opposé de la matrice de similarité ($d = 1 - s$). Pour des raisons de clarté, nous ne parlerons dans le reste du document que de matrices de dissemblance (la matrice de dissemblance individuelle initialisée à 1 est décrétementée selon la réponse de l'auditeur). Pour deux stimuli i et j , les jugements reportés dans une matrice individuelle peuvent varier selon que i ou j soit présenté en tant que référence. En fait, une paire (ij) peut apparaître sous trois présentations différentes. Soit i, j jouent un rôle identique et font partie des stimuli de comparaison, soit i est le stimulus de référence et j est parmi les stimuli de comparaison, soit l'inverse. Ces deux derniers cas peuvent générer une asymétrie dans la matrice car le choix du son qui est le plus similaire à la référence dépend des deux autres sons de comparaison présents dans le trio (par exemple j a été choisi comme le plus similaire à la référence i lors d'un essai alors que i n'a pas été choisi pour la référence j dans un deuxième essai). Ceci donne des valeurs $d_{ij} \neq d_{ji}$ conduisant à une matrice individuelle asymétrique. Rogowitz *et al.* [1998] proposent de symétriser les termes diagonalement opposés de la matrice de dissemblance individuelle. La matrice de dissemblance symétrique est alors obtenue : $D_{ij} = \frac{d_{ij} + d_{ji}}{2}$.

La répartition des sons de comparaison en trios conduit à des matrices individuelles incomplètes et l'attribution des valeurs de dissemblances est faite de manière binaire. Pour réduire l'influence de ces deux points, il est nécessaire d'effectuer la moyenne des matrices individuelles de plusieurs auditeurs pour obtenir une matrice de dissemblance moyenne convenablement remplie. Le fait que les matrices individuelles soient différentes ne nous permet pas d'évaluer les différences interindividuelles et d'employer des méthodes de représentation MDS de type INDSCAL ou CLASCAL qui permettent de mettre en évidence des spécificités (voir annexe D). Pour estimer l'influence du nombre de stimuli et du nombre d'auditeurs sur l'obtention de la matrice moyenne nous avons effectué une série de simulations.

4.3 Simulations

Dans le but d'analyser les performances de la méthode de comparaison avec permutation de références pour l'évaluation de la dissemblance, nous avons simulé les jugements d'auditeurs "idéaux". Les auditeurs simulés sont considérés comme idéaux car leurs jugements de dissemblances correspondent à de vraies dissemblances et sont supposés parfaits. Pour obtenir ces

jugements de dissemblances connues, nous utilisons des mesures de distances. Ainsi, un échantillon aléatoire de n points est disposé dans un espace pour lequel le nombre de dimensions a été arbitrairement fixé à 3. La distance entre chaque paire de points est définie comme la valeur de dissemblance placée dans une matrice $n \times n$. Pour le reste du chapitre, cette matrice originale est appelée "matrice de dissemblances connues". Les valeurs de dissemblances connues ont été utilisées lors de la simulation d'un essai pour savoir quel stimulus de comparaison aurait été choisi comme le plus semblable à la référence par les auditeurs idéaux. Soient A, B, C et D, quatre stimuli intervenant dans la simulation d'un essai, avec A la référence, et B, C et D le trio de stimuli de comparaison. d_{AB} , d_{AC} et d_{AD} correspondent respectivement aux valeurs de dissemblances connues entre A et B, A et C puis A et D. La plus petite dissemblance entre d_{AB} , d_{AC} et d_{AD} correspond à la plus grande similarité et indique quel stimulus parmi B, C et D aurait été choisi par un auditeur idéal comme le plus similaire à la référence A dans cet essai. La simulation de plusieurs essais permet de remplir une matrice de dissemblances comme indiqué précédemment. Dans le reste du document, cette matrice est appelée "matrice de dissemblances estimées". La figure 4.2 illustre le principe de simulation basé sur la matrice de dissemblances connues.

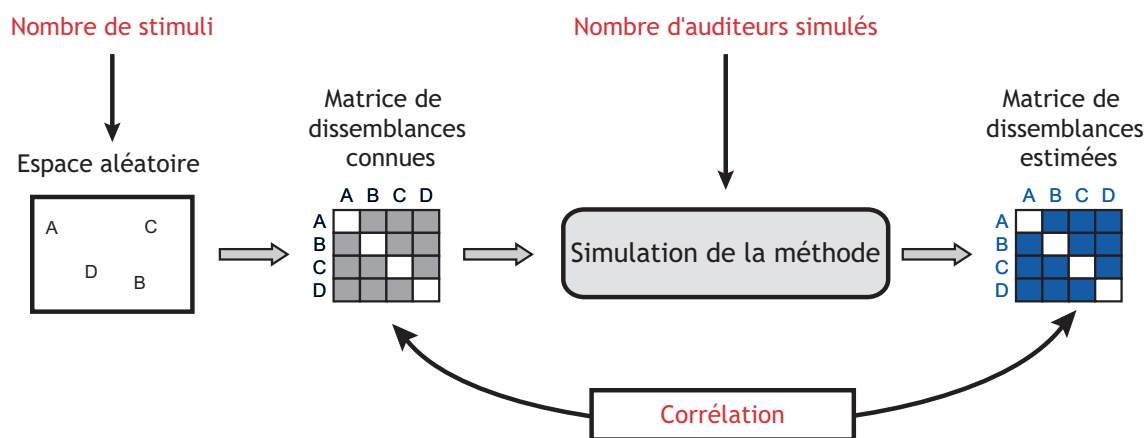


Figure 4.2 – Schéma du principe de la simulation basé sur un espace aléatoire. La méthode est testée en évaluant la relation entre la matrice de dissemblances connues et la matrice de dissemblances estimées.

La comparaison entre la matrice de dissemblances estimées et la matrice de dissemblances connues permet d'évaluer la méthode. Deux simulations ont été menées. La première a permis de simuler un test idéal lors duquel toutes les présentations (référence + trio) seraient jugées. La deuxième simulation a permis de simuler le test réaliste au cours duquel chaque auditeur n'évalue que quelques présentations possibles. A l'aide de ces deux simulations, l'influence du nombre de stimuli et du nombre d'auditeurs simulés a pu être déterminée.

4.3.1 Simulation 1 : Influence du nombre de stimuli

La simulation 1 consiste à évaluer l'influence du nombre de stimuli sur l'évaluation de la dissemblance par la méthode. Nous avons donc simulé toutes les combinaisons possibles de trois sons qui pourraient être comparés à chaque référence. En présentant toutes les combinaisons de trois parmi n stimuli, ce qui correspond à un total de $(n - 1)! / (3!((n - 1) - 3)!)$, la méthode est capable de fournir toutes les informations de dissemblances disponibles. Un unique auditeur simulé est nécessaire pour obtenir cette matrice de dissemblances estimées. Cette simulation 1 a été menée pour un nombre de stimuli allant de 10 à 50. Pour chaque cas, la corrélation entre les matrices de dissemblances connues et estimées a été calculée.

4.3.2 Simulation 2 : Influence du nombre d'auditeurs

La simulation 2 a pour but d'estimer le nombre d'auditeurs nécessaires pour obtenir une matrice moyenne de dissemblances équivalente à celle obtenue dans la simulation 1. Nous avons donc entrepris de simuler des tests plus réalistes en faisant varier le nombre d'auditeurs simulés. Lors d'un tel test, un stimulus parmi n est désigné comme référence puis comparé aux autres $(n - 1)$ aléatoirement distribués en $(n - 1)/3$ trios. Chaque stimulus est à son tour utilisé comme référence et comparé au reste du corpus. Pour chaque auditeur, la méthode ne permet pas de remplir complètement la matrice individuelle de dissemblances. Pour simuler des tests réalistes, plusieurs auditeurs sont donc nécessaires afin d'obtenir une matrice de dissemblances moyenne. Pour cette simulation, le nombre de stimuli était fixé arbitrairement à 40. Pour chaque référence, les 39 stimuli de comparaison étaient distribués en 13 trios. Un test entier comptant $n(n - 1)/3$ essais, le nombre d'essais pour la simulation 2 était de 520 (à noter que la simulation 1 effectuée sur un corpus de 40 stimuli entraîne un nombre total de 365560 essais).

4.4 Résultats

4.4.1 Relation entre les dissemblances connues et estimées

La relation entre les dissemblances connues et estimées a été évaluée. La figure 4.3 représente les dissemblances estimées, résultant de la simulation 1 avec un nombre de 40 stimuli, en fonction des données de dissemblances connues.

La relation entre les deux matrices de dissemblances apparaît comme non linéaire, monotone et croissante. D'après la figure 4.3, la majorité des dissemblances estimées sont surestimées et apparaissent au dessus de la diagonale, qui elle, indiquerait l'égalité entre dissemblances estimées et connues. Cette tendance pourrait évoquer l'existence d'un biais potentiel.

Concernant la comparaison des deux matrices, nous pouvons estimer la corrélation entre les deux séries de valeurs qui constituent ces matrices. Le coefficient de corrélation de Pearson est

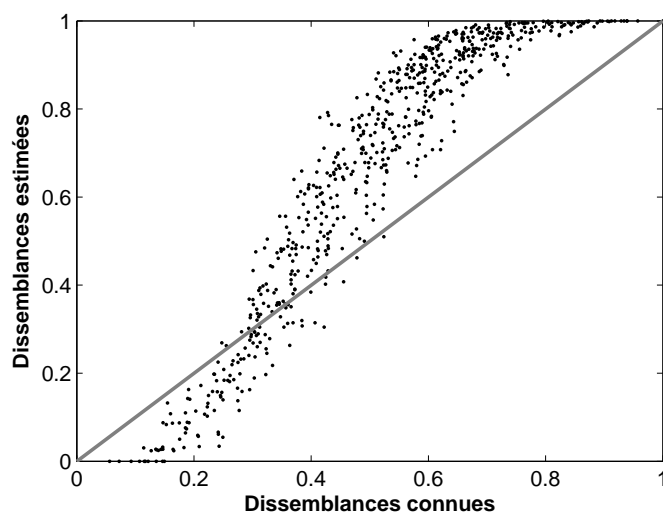


Figure 4.3 – Relation non linéaire entre les valeurs issues des matrices de dissemblances connues et estimées à partir de la simulation 1 pour 40 stimuli.

inadapté, car il s’appuie sur la supposition que la relation entre deux grandeurs est linéaire. Par conséquent, pour les deux simulations, le coefficient de corrélation de Spearman a été préféré pour estimer le lien entre les deux matrices de dissemblances. Ce coefficient de corrélation offre l’avantage d’être adapté aux relations non linéaires étant donné qu’il implique les rangs plutôt que les valeurs des éléments de chaque matrice. Il évalue donc le caractère monotone de leur relation, pas son caractère linéaire. La valeur du coefficient de corrélation de Spearman correspond au coefficient de Pearson calculé sur les rangs des deux séries de données. Pour les résultats de la simulation, nous avons utilisé le coefficient de Spearman. Les matrices de dissemblances étant symétriques, les corrélations ont été calculées sur la moitié de chaque matrice sans prendre en compte la diagonale qui n’est pas modifiée par le test.

4.4.2 Nombre de stimuli

La simulation 1 a été menée en faisant varier le nombre de stimuli. Il en résulte que le coefficient de corrélation de Spearman reste constant pour un nombre de stimuli compris entre 10 et 50. La valeur atteint $\rho = 0.96$ avec un écart-type de 0.01. La simulation 1 montre donc que le nombre de stimuli influence peu la corrélation entre les matrices connues et estimées quand la simulation considère tous les trios possibles pour chaque référence. Par contre, la simulation de la méthode ne permet pas de retrouver exactement les valeurs de dissemblances connues puisque la corrélation $\rho = 1$ n’est pas atteinte même quand tous les trios possibles sont testés. Ce biais est discuté ultérieurement.

4.4.3 Nombre d'auditeurs simulés

Afin d'observer les différences entre les simulations 1 et 2, la corrélation de Spearman entre les matrices de dissemblances issues de chaque simulation a été calculée en fonction du nombre d'auditeurs impliqués dans la simulation 2, pour un nombre de stimuli fixé à 40. La figure 4.4 présente ces résultats pour un nombre d'auditeurs variant entre 1 et 100. Lorsque le nombre d'auditeurs augmente, cette figure indique que les valeurs de dissemblances obtenues avec la simulation 2 convergent vers celles issues de la simulation 1. La corrélation croît rapidement entre 1 et 20 auditeurs, puis se stabilise pour un nombre d'auditeurs plus élevé. D'après la figure 4.4, à partir de 20 auditeurs idéaux, la plupart de l'information de dissemblance est obtenue avec la simulation 2. Ainsi, pour 20 auditeurs idéaux, le coefficient de corrélation de Spearman atteint déjà $\rho = 0.98$, ce qui correspond à un nombre total d'essais égal à 10400 (20×520 , 520 étant le nombre d'essais du test entier pour un auditeur), bien inférieur au 365560 essais nécessaires dans la simulation 1.

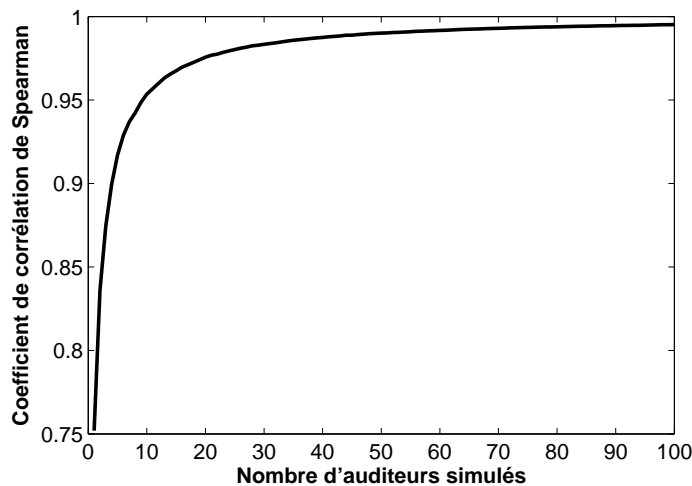


Figure 4.4 – Corrélation entre les matrices de dissemblances estimées issues des simulations 1 et 2 en fonction du nombre d'auditeurs pour 40 stimuli

La nécessité d'un nombre suffisant d'auditeurs résulte du fait que la méthode de comparaison avec permutation de références est appliquée à un corpus de n stimuli où chaque stimulus de référence est comparé aux autres $n - 1$ stimuli distribués en $(n - 1)/3$ trios. Pour une référence, l'auditeur ne peut fournir que $(n - 1)/3$ réponses parmi les $n - 1$ possibles. Par conséquent, pour chaque matrice individuelle de dissemblance, un tiers de la matrice est modifié avec une valeur fixée à 0 et les deux tiers restants conservent leur valeur initiale de 1. La figure 4.4 montre donc que l'influence de ce remplissage partiel de chaque matrice individuelle est réduite efficacement en augmentant le nombre d'auditeurs simulés pour estimer la dissemblance moyenne.

Pour évaluer les performances de la simulation 2, nous avons calculé le coefficient de cor-

relation de Spearman entre les dissemblances estimées et les dissemblances connues pour un nombre de stimuli fixé à 40 en fonction du nombre d’auditeurs simulés. La figure 4.5 représente l’évolution de la corrélation pour un nombre d’auditeurs compris entre 1 et 100. Les valeurs de corrélation augmentent rapidement pour un nombre d’auditeurs simulé inférieur à 20, puis la courbe se stabilise pour un nombre d’auditeurs supérieur.

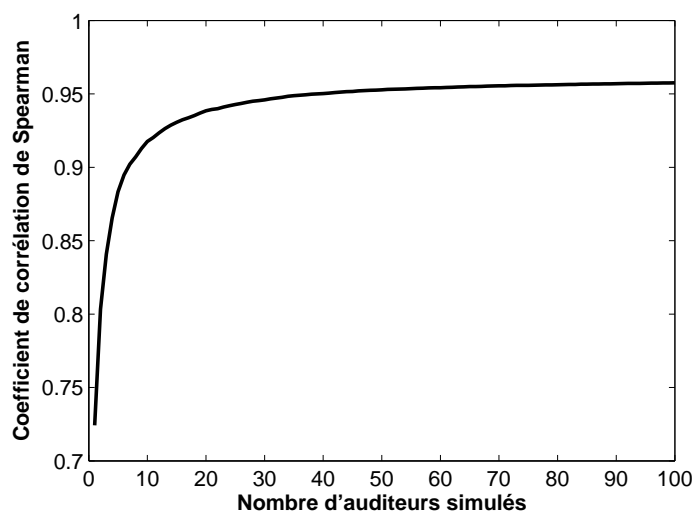


Figure 4.5 – Corrélation entre les matrices de dissemblances connues et estimées avec la simulation 2 en fonction du nombre d’auditeurs pour 40 stimuli

Il est intéressant de noter que même pour 100 auditeurs idéaux, le coefficient de corrélation ne dépasse pas $\rho = 0.96$. Cette valeur limite de corrélation est identique à celle obtenue avec la simulation 1. Elle confirme le résultat obtenu sur la figure 4.4, à savoir que la matrice de dissemblance obtenue par la simulation 2 converge assez rapidement vers la matrice de dissemblance obtenue par la simulation 1. Ce résultat confirme aussi que la méthode ne permet pas de retrouver rigoureusement les valeurs de dissemblances, alors que la manière de remplir partiellement chaque matrice individuelle n’introduit pas de biais par rapport à un remplissage total. Le biais observé serait donc inhérent à la méthode.

4.4.4 Caractérisation du biais

Pour caractériser ce biais, il faut étudier précisément comment les données de dissemblances connues sont transformées en dissemblances estimées. Dans un premier temps, nous avons montré que la relation entre les deux matrices était non linéaire, monotone et croissante. Ainsi, le coefficient de Spearman a été employé pour estimer la relation entre les deux matrices de dissemblances. Afin de comprendre comment la méthode transforme les dissemblances connues en dissemblances estimées, les distributions statistiques des valeurs de chaque matrice sont présentées sur la figure 4.6. Les valeurs de dissemblances estimées proviennent de la simulation 2

effectuée avec 100 auditeurs simulés et sur un corpus de 40 stimuli.

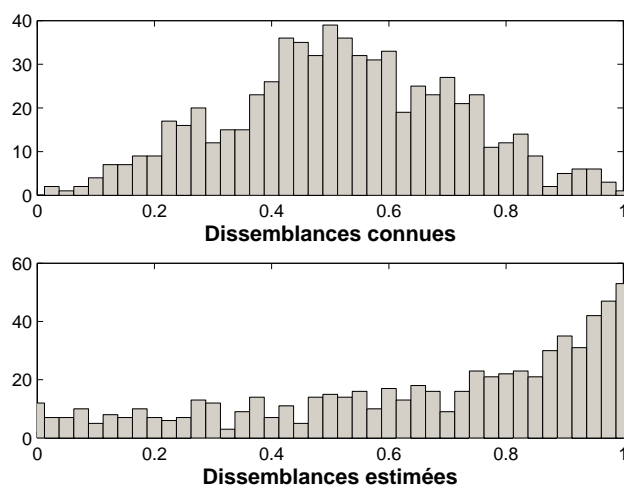


Figure 4.6 – Histogrammes représentant la distribution statistique des valeurs de dissemblances connues et estimées à partir de la simulation 2 avec 40 stimuli et 100 auditeurs.

Les valeurs de dissemblances connues sont distribuées suivant une loi approximativement gaussienne. La distribution des valeurs de dissemblances estimées fait apparaître un décalage vers des valeurs plus grandes (on le remarque aussi sur la figure 4.3, la plupart des points apparaissent au dessus de la diagonale, ligne pour laquelle les dissemblances connues et estimées ont une valeur identique).

Ce décalage est dû à la manière d’attribuer une valeur de similarité entre la référence et les stimuli de comparaison. Lors d’un essai, une information de similarité est donnée entre le stimulus de référence et celui qui est choisi, mais deux informations de dissemblances sont conservées entre la référence et les deux autres stimuli qui n’ont pas été choisis. Un poids deux fois plus important est finalement accordé aux stimuli qui ne sont pas choisis pour chaque essai. Pour chaque auditeur, la matrice de dissemblance contient donc un tiers de réponses notées 0 et les deux tiers restants sont laissés à leur valeur initiale de 1. La comparaison entre la simulation 1 et 2 ont montré que l’influence de ce remplissage partiel est diminuée avec le nombre d’auditeurs suffisant. Le biais provient uniquement des valeurs de dissemblances attribuées au cours d’un essai. Le fait d’avoir deux éléments qui conserve leur dissemblance maximale entraîne la distribution finale des valeurs de dissemblance à tendre vers 1.

L’attribution des valeurs de dissemblances dépend de ce poids et les valeurs de similarité sont sous-estimées par rapport aux valeurs de dissemblances. La courbe représentant la relation entre les dissemblances connues et estimées (figure 4.3) et leurs histogrammes (figure 4.6) confirme le fait que la méthode sous-estime légèrement les petites valeurs et surestime au contraire les grandes. Le biais est donc intrinsèque à la méthode et provient de l’attribution des valeurs

de dissemblances pour les stimuli présents au sein d'un essai. Ceci affecte relativement peu le coefficient de corrélation de Spearman, basé sur des relations d'ordre, mais conduit à une estimation des dissemblances distordues par rapport aux dissemblances initiales.

Pour mesurer l'influence du biais sur l'analyse MDS, nous avons comparé l'espace 3D original avec le résultat de l'analyse MDS des données de dissemblances estimées à partir de la simulation d'un test effectué sur 40 stimuli avec 100 auditeurs. Nous avons effectué cette comparaison de l'espace initial avec deux espaces issus d'analyse métrique et non-métrique des données de dissemblances estimées. Pour l'analyse métrique, nous avons employé le modèle MDSCAL et pour l'analyse nonmétrique le modèle programmé par Matlab. Pour chaque type d'analyse MDS, les coordonnées des points de l'espace initial et celles des points des espaces issus de la simulation sont quasiment identiques suivant les 3 dimensions (coefficient de corrélation linéaire de Pearson supérieur à $r = 0.99$ pour chaque dimension et pour chaque type d'analyse). L'analyse MDS, qu'elle soit métrique ou non, semble alors atténuer l'effet du biais observé sur les valeurs de dissemblances. Même si l'estimation de dissemblance est légèrement biaisée, l'espace MDS résultant de l'analyse de ces dissemblances est identique à l'espace qui nous a permis de générer les dissemblances connues introduites dans la simulation de la méthode.

Les différentes simulations ont permis d'explorer la méthode de comparaison avec permutation de références et d'obtenir des indications sur le nombre de stimuli et le nombre d'auditeurs nécessaires. Nous avons montré que le nombre de stimuli n'influence pas les résultats de la méthode. De plus, les résultats de la simulation 2 indiquent que 20 auditeurs idéaux simulés sont suffisants pour obtenir des valeurs de dissemblances correctes. Néanmoins la relation entre les dissemblances connues et estimées est non linéaire, le coefficient de corrélation de Spearman entre ces deux dissemblances n'atteint pas la valeur maximale de 1 indiquant que la méthode ne permet pas de retrouver la matrice initiale de dissemblances connues, indépendamment du nombre de stimuli et d'auditeurs. Un tel biais n'existe pas pour la méthode de comparaison par paires car la matrice de dissemblances estimées serait remplie directement avec les dissemblances connues en suivant le principe des auditeurs parfaits.

En conclusion, même avec un grand nombre d'auditeurs simulés, la méthode de comparaison avec permutation de références présente un biais. La simulation de la méthode nous a permis de le mettre en évidence et de l'estimer. Même dans le cas où la matrice est idéalement remplie (simulation 1), la méthode de comparaison avec permutation de références surestime les valeurs de dissemblance. Ce biais est alors intrinsèque à la méthode proposée, mais reste relativement faible, et peut donc être acceptable car il ne semble pas avoir d'influence sur l'analyse MDS des dissemblances qui est faite par la suite.

L'utilisation d'auditeurs idéaux nous a permis de mettre en évidence les biais d'estimation de la méthode. Lorsque nous appliquons cette méthode lors de vrais tests, il faut garder à l'esprit que

les jugements ne sont plus idéaux. Le chapitre suivant est dédié à la comparaison de cette méthode avec la méthode de comparaison par paires sur un corpus de 12 stimuli et de classification libre sur un corpus de 37 stimuli.

Chapitre 5

Validation de la méthode de comparaison avec permutation de références

Sommaire

5.1	Test 1 : Évaluation d'un corpus de 12 enceintes	97
5.1.1	Stimuli	97
5.1.2	Auditeurs	97
5.1.3	Procédure	97
5.1.4	Résultats	98
5.2	Test 2 : Évaluation d'un corpus de 37 enceintes	102
5.2.1	Stimuli	102
5.2.2	Auditeurs	102
5.2.3	Procédure	102
5.2.4	Résultats	102
5.3	Influence de la méthode d'évaluation	104

Ce chapitre présente l'application de la méthode de comparaison avec permutation de références pour l'évaluation de la dissemblance entre des enregistrements d'enceintes acoustiques. Le but est de vérifier si le biais estimé dans les simulations est retrouvé et de valider l'utilisation de la méthode pour l'évaluation des dissemblances et l'analyse de ces dernières avec une technique de représentation multidimensionnelle (MDS). Les deux tests que nous avons effectués portent sur l'évaluation du timbre car nous avons la possibilité de comparer nos résultats avec ceux déjà obtenus par Lavandier [2005]. La méthode a donc été testée sur un premier corpus de 12 enceintes pour la confronter à la méthode de comparaison par paires que Lavandier [2005] a effectuée sur ce même corpus. Ensuite un deuxième test sur un corpus plus important de 37 enceintes a été mené. Les résultats obtenus sont comparés avec ceux obtenus sur le premier corpus mais aussi avec les résultats issus de la classification libre menée par Lavandier [2005] sur le deuxième corpus.

5.1 Test 1 : Évaluation d'un corpus de 12 enceintes

5.1.1 Stimuli

Les stimuli sont identiques à ceux du premier corpus constitué de 12 enceintes évalué par Lavandier [2005]. Douze enceintes ont été enregistrées dans une même pièce avec une prise de son stéréophonique AB-ORTF. L'extrait musical enregistré était tiré d'un morceau de Mc Coy Tyner intitulée "Miss Bea" et durait 3.3 secondes. La sonie globale des stimuli reproduits avait été égalisée par Lavandier [2005] à 70 phones.

5.1.2 Auditeurs

Le test d'écoute a été réalisé avec 27 auditeurs (7 femmes et 20 hommes). Tous les auditeurs étaient membres du laboratoire. Tous les auditeurs ont réalisé un audiogramme et aucun ne présentait de troubles auditifs. Le nombre d'auditeurs ayant participé au test de comparaison avec permutation de références était identique au nombre d'auditeurs ayant effectué le test de la comparaison par paires [Lavandier, 2005].

5.1.3 Procédure

Utiliser la méthode de comparaison avec permutation de références avec 12 stimuli implique que chaque son apparaissant comme référence soit comparé aux 11 autres distribués aléatoirement en trios. Ces 11 sons de comparaison ne pouvant pas former un nombre entier de trios, nous avons distribué les stimuli de comparaison en 4 trios, le dernier trio étant composé des deux sons restants et complété avec un son choisi aléatoirement parmi les 9 sons déjà tirés. Le test entier pour évaluer les 12 enregistrements d'enceintes avec la méthode de comparaison avec permutation

de références était ainsi composé de 48 essais (12x4). Pour chaque auditeur la distribution des sons de comparaison en trios, et l'ordre de présentation des essais étaient tirés aléatoirement.

A chaque essai, 4 sons (1 référence + 3 sons de comparaison) étaient présentés automatiquement à l'auditeur. Ensuite les auditeurs avaient la possibilité de réécouter les sons (dans n'importe quel ordre et autant de fois qu'ils le souhaitent) avant de répondre. La tâche était de choisir parmi les 3 sons de comparaison, celui qui leur paraissait le plus semblable au son de référence. Le test a été réalisé avec une interface composée de boutons virtuels représentant chaque son, et les auditeurs pouvaient écouter le son en cliquant sur le bouton correspondant avant de répondre. Le test d'écoute était réalisé dans une cabine audiométrique et les sons étaient joués via une carte son (M-Audio Delta 66) et reproduits au casque (Stax SR Lambda Professionnal).

5.1.4 Résultats

5.1.4.a Comparaison des dissemblances perceptives

La comparaison des dissemblances perceptives consiste ici à calculer la corrélation entre deux matrices de dissemblances. La première est issue de notre test et la deuxième correspond à la matrice de dissemblances obtenue par Lavandier [2005] lors d'un test de comparaison par paires.

Le coefficient de corrélation de Spearman et de Pearson ont été comparés pour estimer la corrélation entre les deux matrices. La valeur de la corrélation de Pearson atteint $r = 0.74$ et $\rho = 0.75$ pour la corrélation de Spearman. Ces valeurs indiquent que les données de dissemblances issues des deux tests sont assez proches mais pas identiques. De plus, les valeurs proches des deux types de corrélation semblent indiquer que le bruit de mesure des deux méthodes ne favorise aucune des deux manières d'estimer la corrélation.

Une manière de confronter visuellement les données de dissemblances est d'effectuer une analyse en arbre hiérarchique. Le dendrogramme résultant de ce type d'analyse permet de représenter la structure des éléments de la matrice de dissemblances reliés par des branches. La dissemblance entre deux éléments est représentée par la hauteur du nœud qui les raccorde. Plus les éléments ont été jugés similaires, plus la hauteur du nœud est faible. La figure 5.1 présente les dendrogrammes des données de dissemblances issues de la méthode de comparaison par paires et de la méthode de comparaison avec permutation de références.

Les deux dendrogrammes font apparaître une structuration des données de dissemblances similaire pour les deux méthodes. Les groupes d'enceintes (7-8), (4-6-11) et (2-5-12) se retrouvent sur les deux dendrogrammes. Le dendrogramme présentant les données issues de la comparaison par paires fait apparaître les enceintes 2, 5 et 12 comme ayant été jugées différentes des autres. Sur le dendrogramme issu de notre test, la même tendance est visible concernant ces enceintes. De plus, les enceintes 7 et 8 correspondent au même modèle d'enceinte. Elles apparaissent bien sur les deux dendrogrammes comme ayant été jugées les plus proches par les auditeurs des deux

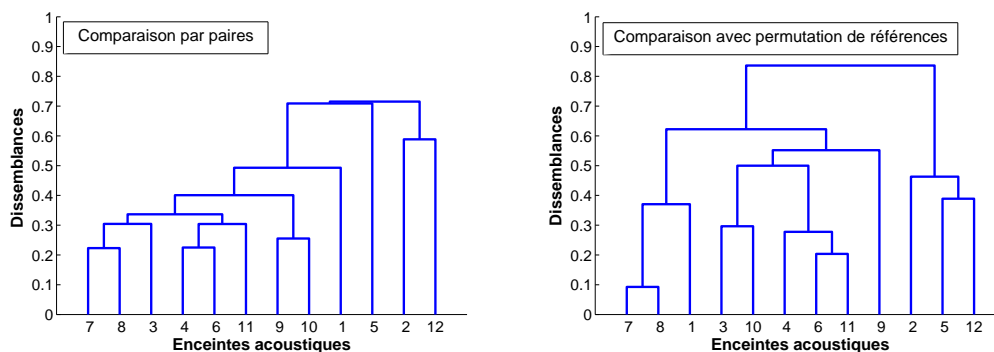


Figure 5.1 – Dendrogrammes résultant de l’analyse en arbre hiérarchique des dissemblances obtenue sur 12 enceintes avec la méthode de comparaison par paires (à gauche) et avec la méthode de comparaison avec permutation de références (à droite).

tests.

Le biais évoqué lors des simulations pourrait être présent mais masqué par le bruit de la mesure sur un nombre restreint d’auditeurs non idéaux. Pour aller plus loin dans la comparaison, nous avons comparé les espaces perceptifs issus des deux expériences afin de déterminer si la méthode de comparaison avec permutation de références permet de construire un espace proche de celui obtenu avec l’analyse d’une matrice de dissemblance issue de comparaisons par paires.

5.1.4.b Comparaison des espaces perceptifs

Les données de dissemblances issues de notre test utilisant la méthode de comparaison avec permutation de références ont été soumises à une analyse MDS. Nous avons vu dans la simulation de la méthode qu’une analyse métrique ou non métrique des données issues de notre méthode modifie très peu les dimensions. Nous avons alors décidé d’utiliser un modèle de MDS identique à celui utilisé par Lavandier [2005]. Le modèle que nous avons utilisé pour les expériences présentées dans ce chapitre est un modèle métrique MDSCAL. Des détails sur le modèle de MDS utilisé et la courbe de stress permettant de déterminer le nombre de dimensions sont présentés dans l’annexe D. L’analyse de la matrice de dissemblance que nous avons obtenue conduit à un espace à deux dimensions. La matrice de dissemblances issue du test avec la méthode de comparaison par paires a également permis de dévoiler deux dimensions à partir de la même technique MDS [Lavandier, 2005]. Pour comparer les deux espaces perceptifs, la corrélation entre les coordonnées des points représentant les enregistrements a été calculée suivant chaque dimension. La corrélation entre les coordonnées des stimuli présents sur les deux espaces perceptifs est de $r = 0.96$ pour la dimension 1, et $r = 0.88$ suivant la dimension 2. La figure 5.2 présente la relation entre les coordonnées issues des deux méthodes pour chaque dimension.

Les valeurs de corrélation obtenues entre les dimensions montrent que les espaces perceptifs résultant de l’analyse MDS sont très proches pour les deux méthodes. Afin de comparer

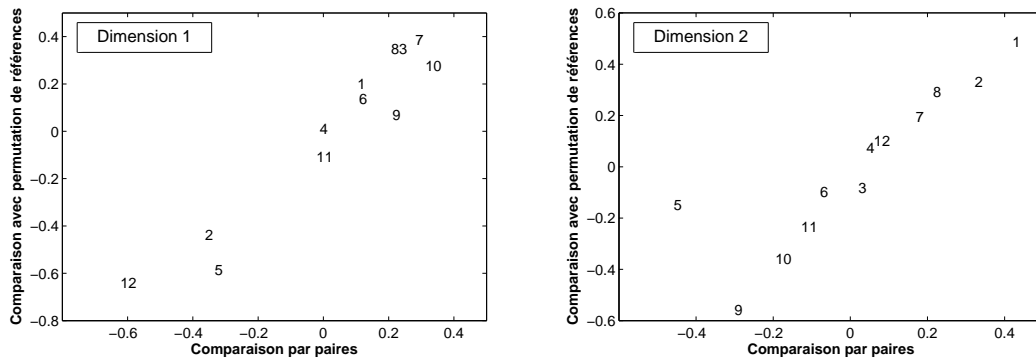


Figure 5.2 – Relation entre les coordonnées des points des deux espaces. La relation pour la dimension 1 est présentée sur la figure de gauche et celle pour la dimension 2 est présentée sur la figure de droite.

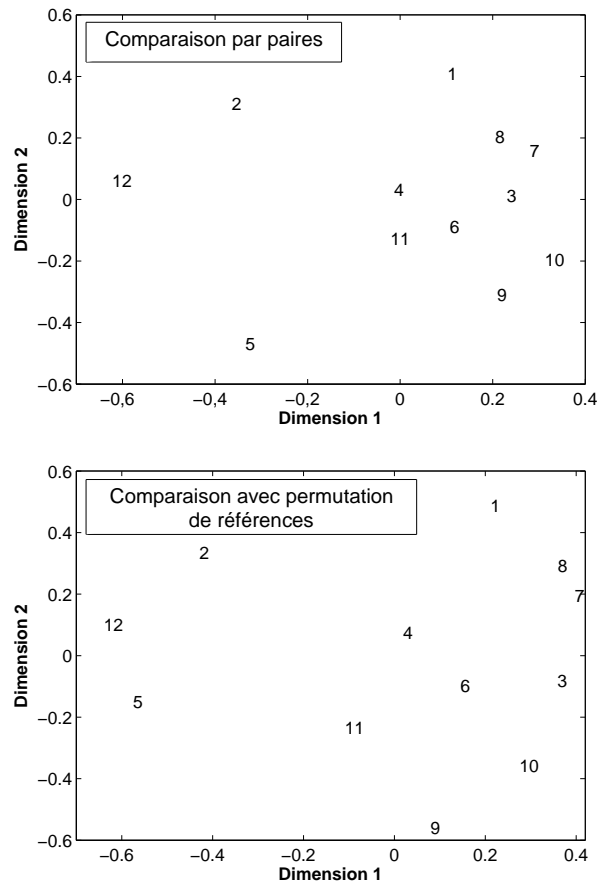


Figure 5.3 – Espaces perceptifs obtenus avec la méthode de comparaison par paires (en haut) et la méthode de comparaison avec permutation de références (en bas). Chaque numéro correspond à l’enregistrement d’une enceinte.

visuellement les dimensions obtenues grâce aux deux méthodes, la figure 5.3 présente les deux espaces perceptifs. La position des points représentant les enregistrements d’enceintes est quasi-

ment identique d'un espace à l'autre. Lors des sessions d'enregistrements, Lavandier *et al.* [2008a] ont volontairement enregistré deux enceintes du même modèle (7 et 8). Ces enregistrements apparaissent bien côte à côte dans les deux espaces perceptifs. Les espaces semblent similaires à l'exception de l'enregistrement de l'enceinte 5. Le comportement particulier de cette enceinte avait déjà été mis en évidence par Lavandier *et al.* [2004] lors d'une expérience cherchant à caractériser l'influence de la prise de son sur les espaces perceptifs. La position des enregistrements de l'enceinte 5 variait dans les différents espaces alors que les autres enceintes conservaient une position sensiblement constante. Cette information se retrouve aussi en observant les relations entre les coordonnées des points de chaque espace suivant les deux dimensions (figure 5.2). L'enceinte 5 se distingue des autres enceintes sur la figure présentant la relation des coordonnées suivant la dimension 2. L'écart de cette enceinte par rapport à la relation linéaire des données issues des deux tests explique aussi la valeur de corrélation plus faible obtenue pour la dimension 2.

Cet enregistrement d'enceinte pourrait avoir des caractéristiques perceptives spécifiques qui ne soient pas partagées par les autres enregistrements. L'algorithme MDS que nous avons utilisé permet de représenter seulement les dimensions communes aux différents stimuli dans l'espace perceptif. Il ne prend pas en compte une dimension particulière associée à un seul stimulus. La spécificité de l'enceinte 5, ignorée par la MDS, pourrait expliquer pourquoi sa position n'est pas stable d'un espace à un autre. Quoiqu'il en soit, cette instabilité a été observée auparavant et ne semble pas être due à la méthode de test utilisée pour les tests d'écoute.

Il est à noter que lors de l'interprétation des dimensions, Lavandier [2005] a trouvé que la première dimension était liée à l'équilibre grave/aigu, alors que la deuxième correspondait à une notion de clarté liée à l'émergence des médiums. Nous avons procédé à une interprétation de l'espace obtenu avec la méthode de comparaison avec permutation de références. En écoutant les sons le long des deux dimensions, nous retrouvons les mêmes caractéristiques. Suivant la dimension 1, les sons perçus comme ayant trop de basses sont situés d'un côté de la dimension, ensuite apparaissent les sons "équilibrés" et enfin les sons perçus comme ayant peu de basses. Les sons le long de la dimension 2 sont organisés par rapport à la présence plus ou moins importante des médiums.

La méthode de comparaison avec permutation de références appliquée au corpus de 12 enceintes évalué par Lavandier [2005] fournit des résultats perceptifs similaires à ceux obtenus avec la méthode de comparaison par paires. Les matrices de dissemblances ne sont pas exactement les mêmes mais l'analyse MDS montre que nous obtenons le même espace perceptif qu'avec la méthode de comparaison par paires sur les 12 enceintes. Cette méthode de comparaison par paires peut être considérée comme une méthode de "référence". Obtenir des dimensions similaires à celles dévoilées à l'aide de la comparaison par paires montre que malgré un biais intrinsèque, la méthode de comparaison avec permutation de références semble bien adaptée pour la construction d'un espace perceptif, via l'évaluation de la dissemblance.

5.2 Test 2 : Évaluation d'un corpus de 37 enceintes

Le premier test a permis d'évaluer un corpus composé de 12 enceintes et d'obtenir un espace perceptif à deux dimensions pour les deux méthodes d'évaluation. Il est encore plus intéressant d'utiliser la méthode de comparaison avec permutation de références pour évaluer un corpus d'enceintes important. Dans un premier temps, nous avons cherché à vérifier si l'évaluation d'un corpus plus important avec la méthode proposée permet de trouver de nouvelles dimensions. Ensuite, nous avons comparé l'espace perceptif avec celui obtenu par Lavandier [2005] effectué sur le même corpus avec la méthode de classification libre pour comparer les répartitions des enceintes le long des différentes dimensions.

5.2.1 Stimuli

Les 37 stimuli évalués au cours de ce test correspondaient aux 37 enceintes acoustiques enregistrées par Lavandier [2005]. Ces enregistrements ont aussi été effectués avec un couple AB-ORTF mais dans une salle différente du test 1. L'extrait musical que nous avons choisi est le même que pour le test 1. Parmi les 37 enceintes, 11 d'entre elles faisaient partie du premier corpus. Ces enceintes communes aux deux tests nous permettent de comparer les espaces perceptifs obtenus pour chaque corpus.

5.2.2 Auditeurs

Pour ce test, vingt-sept auditeurs (5 femmes et 22 hommes) ont participé à l'évaluation des 37 enceintes. Parmi ces auditeurs, treize avaient participé au premier test perceptif. De la même manière que dans le premier test, les auditeurs étaient aussi membres du laboratoire et ne présentaient pas de déficits auditifs.

5.2.3 Procédure

Pour un corpus de 37 enceintes, chaque référence était comparée aux 36 sons restants distribués aléatoirement en 12 trios. Le test entier était donc composé de 12x37 essais, ce qui correspondait à un total de 444 essais. Le test entier était divisé en 5 sessions d'une durée moyenne de 30 minutes chacune. Les conditions de réalisation du test (interface et restitution) étaient identiques à celles du test 1.

5.2.4 Résultats

La matrice de dissemblances moyenne obtenue a été soumise à la même analyse MDS que celle utilisée lors du premier test. L'espace perceptif résultant est composé de 3 dimensions (voir la courbe de stress dans l'annexe D).

Nous avons dans un premier temps comparé les dimensions obtenues à partir des 11 enceintes communes aux tests 1 et 2. Étant donné que les jugements obtenus au cours des tests pour ces 11 enceintes sont dépendants des autres stimuli du corpus, il ne nous semblait pas prudent d'extraire simplement les données de dissemblances des 11 enceintes de chacune des matrices pour ensuite les soumettre à l'analyse MDS. Nous avons donc mis en évidence les positions relatives de ces enceintes sur les espaces perceptifs issus de l'analyse complète des matrices de dissemblance. Les numéros des enceintes n'étant pas les mêmes d'un test à l'autre, nous les avons modifiés pour faciliter la comparaison. La figure 5.4 présente les espaces perceptifs partiels des deux tests suivant les deux dimensions principales.

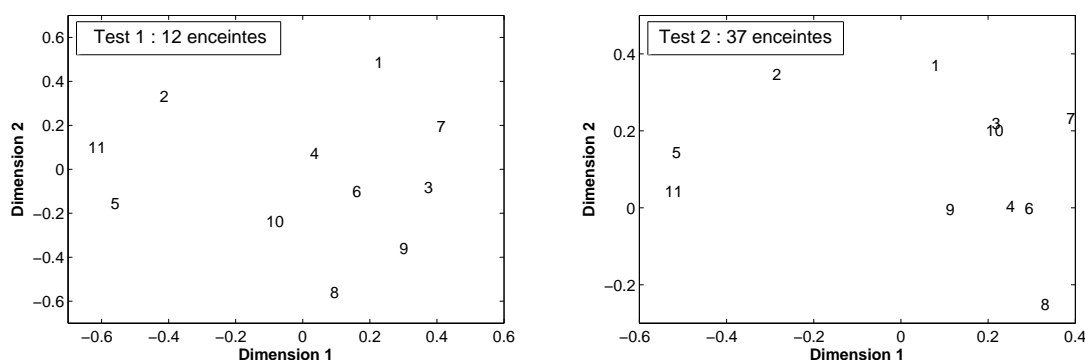


Figure 5.4 – Représentation partielle des 11 enceintes communes aux deux corpus. Les enceintes de chaque corpus ont été renumérotées afin de faciliter la comparaison.

Il est important de rappeler que les 11 enceintes d'un test à l'autre ont été enregistrées avec la même prise de son mais dans deux salles différentes. La position des enceintes dans les espaces perceptifs dépend aussi des autres enceintes impliquées dans les tests. Malgré tout, les positions relatives des enceintes d'un test à l'autre suivant les deux premières dimensions sont assez similaires. Une corrélation de $r = 0.89$ a été obtenue entre les coordonnées des points suivant la dimension 1 et une corrélation de $r = 0.79$ a été obtenue suivant la dimension 2. Ces valeurs indiquent que les dimensions 1 et 2 semblent communes aux deux tests malgré la différence d'environnement d'enregistrement et la taille du corpus évalué.

Afin d'interpréter les différentes dimensions, une écoute informelle des sons suivant les trois dimensions de l'espace obtenues pour les 37 enceintes nous a permis de retrouver les caractéristiques d'équilibre spectral de la dimension 1. Suivant cette dimension, les sons apparaissent d'abord comme ayant trop de basses, puis équilibrés et enfin peu de basses. L'écoute suivant la dimension 2 a révélé une progression de la présence des médiums dans les extraits. La troisième dimension de l'espace perceptif est nouvelle et se caractérise par une sensation d'espace. Elle traduit l'interaction avec la salle, avec d'un côté de la dimension les sons que nous avons qualifié d'"étouffés" et de l'autre les sons ayant une présence spatiale plus importante. Cette écoute

informelle semble donc aussi confirmer que les deux dimensions principales de l'espace perceptif sont communes aux deux tests. L'évaluation d'un corpus plus important avec la méthode de comparaison avec permutation de références fait apparaître une nouvelle dimension.

5.3 Influence de la méthode d'évaluation

La recherche d'une méthode d'évaluation de la dissemblance pour des corpus étendus était motivée par le fait que l'espace perceptif obtenu par Lavandier [2005] avec la méthode de classification libre sur 37 enceintes était en partie différent de celui obtenu avec la méthode de comparaison par paires sur 12 enceintes. Afin d'expliquer ces différences, nous avons comparé les espaces perceptifs résultant des deux tests d'écoute effectués sur le corpus de 37 enceintes.

Le test effectué par Lavandier [2005] sur les 37 enceintes avec la méthode de classification libre a révélé 4 dimensions. La première dimension était liée à la balance spectrale mais son comportement était différent de la dimension obtenue pour le corpus de 12 enceintes. L'écoute le long de cette dimension faisait apparaître des sons manquant de basses puis les sons ayant trop de basses et enfin les sons "équilibrés" spectralement. La deuxième dimension était caractérisée par la sensation d'espace. La troisième dimension était liée à la présence plus ou moins importante des médiums qui pouvait se traduire par une notion de clarté. La quatrième dimension était interprétée par Lavandier [2005] comme ayant un caractère de "douceur" ou de "dureté" notamment sur les cymbales de l'extrait.

Les espaces perceptifs des 37 enceintes suivant les deux premières dimensions pour la méthode de comparaison avec permutation de références et suivant la dimension 1 et 3 pour l'espace issu de la classification libre sont présentés sur la figure 5.5. Cette comparaison est focalisée sur les dimensions d'équilibre spectral et de clarté pour chaque test.

Sur la partie gauche des espaces la position de certaines enceintes change peu d'un espace à l'autre. Cette partie de l'espace correspond aux sons ayant peu de basses. L'observation des positions relatives des enceintes sur la partie droite de l'espace n'est pas évidente car le nombre de stimuli est important. Pour faciliter la comparaison, nous avons tracé sur la figure 5.6 des flèches pour symboliser le déplacement de la position d'une enceinte d'un espace à un autre. L'origine de la flèche correspond à l'espace obtenu avec la méthode de comparaison avec permutation de références. Nous avons noirci les flèches pour mettre en évidence les déplacements importants qui apparaissent sur la partie droite de l'espace perceptif.

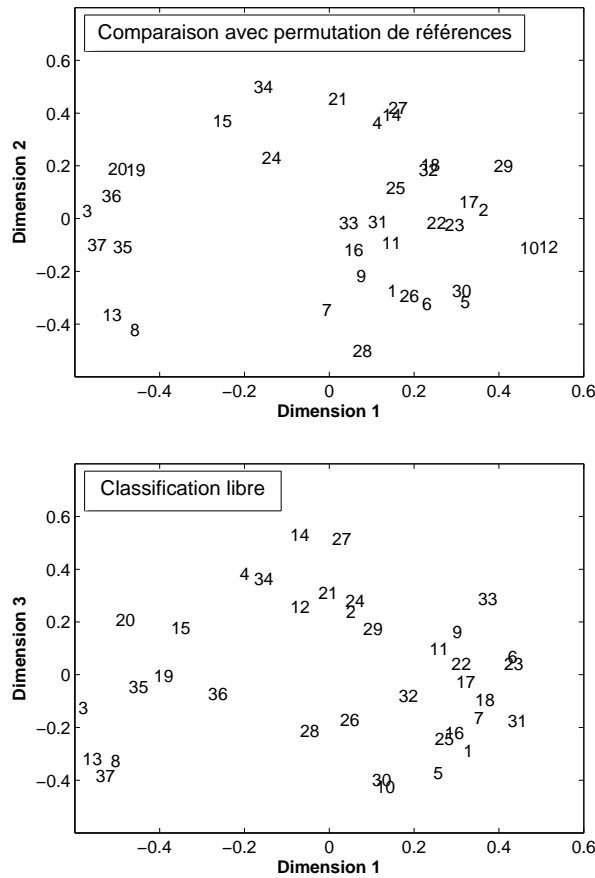


Figure 5.5 – Espaces perceptifs obtenus avec la méthode de comparaison avec permutation de références suivant les dimensions 1 et 2 (en haut) et la méthode de classification libre suivant les dimensions 1 et 3 (en bas) . Chaque numéro correspond à l'enregistrement d'une enceinte.

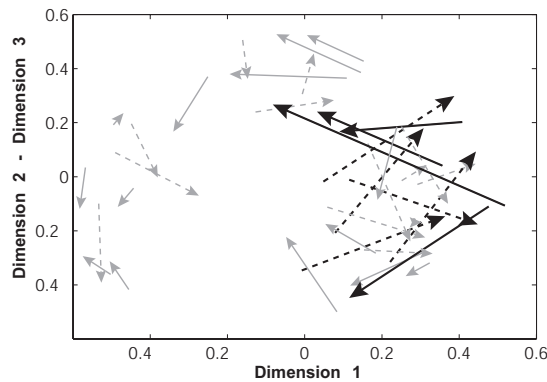


Figure 5.6 – Représentation des déplacements des enceintes de l'espace obtenu avec la méthode de comparaison avec permutation de références vers celui issu de la classification libre. Les flèches sont présentées en trait plein ou en pointillés en fonction du sens de déplacement suivant la dimension 1.

La représentation des déplacements montre que la position de nombreuses enceintes est différente d'un espace à un autre. Les plus grands déplacements sont présents dans la partie droite de l'espace et se produisent suivant la dimension 1 dans les deux sens. Cette inversion des positions des enceintes entre l'espace que nous avons obtenu et celui révélé par Lavandier [2005] semble confirmer les interprétations que nous avons faites. Le comportement de la dimension 1 obtenue avec la classification libre semble bien inversé pour les sons ayant trop de basses et ceux qui sont "équilibrés". Cette orientation avait été révélée par Lavandier [2005] et la comparaison avec la méthode de comparaison avec permutation de références nous permet de conclure sur le fait que cette orientation serait due à la méthode de classification libre.

Pour résumer, nous avons placé dans le tableau 5.1 les différentes dimensions obtenues à partir des deux corpus de 12 et 37 enceintes en fonction de la méthode employée.

Corpus 1 : 12 enceintes		Corpus 2 : 37 enceintes	
Comparaison par paires	Comparaison avec permutation de références		Classification libre
Equilibre spectral	Equilibre spectral	Equilibre spectral	Equilibre spectral *
Clarté	Clarté	Clarté	Sensation d'espace
		Sensation d'espace	Clarté
			Douceur, dureté

Tableau 5.1 – Récapitulatif des différentes dimensions trouvées pour les deux corpus de 12 et 37 enceintes en fonction des méthodes employées. L'astérisque indique un comportement différent de cette dimension par rapport aux autres du même nom.

L'évaluation d'un corpus comprenant plus d'enceintes avec les méthodes de comparaison avec permutation de références et de classification a mené à un espace perceptif comprenant plus de dimensions. Par rapport au deux dimensions mis en évidence pour le corpus de 12 enceintes, l'évaluation d'un corpus de 37 enceintes nous a permis d'obtenir une troisième et avait permis à Lavandier [2005], avec la méthode de classification libre, de révéler un espace comprenant 4 dimensions.

Il est intéressant de noter que les dimensions principales utilisées par les auditeurs ayant participé au test de classification libre ne sont pas les mêmes que celles employées par les auditeurs ayant participé aux autres tests. Les résultats de la classification libre montrent que les auditeurs ont accordé plus d'importance à la sensation d'espace des stimuli plutôt que la clarté des sons. Pour le corpus de 37 enceintes, l'ordre des dimensions de l'espace perceptif que nous avons obtenu avec la méthode de comparaison avec permutation de références est cohérent avec celui mis en

évidence lors des tests sur le corpus de 12 enceintes.

Seule la classification libre conduit à une dimension 1 dont le comportement n'est pas progressif en fonction de l'équilibre spectral. Cette caractéristique a été interprétée par Lavandier [2005] comme pouvant provenir de jugements de préférence des auditeurs ayant participé à la tâche de classification libre. D'après Lavandier [2005], il est possible que les auditeurs aient préféré dans l'ordre les enceintes "équilibrées" puis les enceintes avec une forte présence de basses et enfin les enceintes ayant peu de basses. Cette dimension atypique ne permettait pas de conclure sur les dimensions de l'espace perceptif et Lavandier [2005] s'est demandé si cette différence provenait de la méthode de classification libre, ou des enceintes composant le corpus. Les résultats que nous obtenons sur le même corpus avec la méthode de comparaison avec permutation de références semble indiquer que la méthode de classification serait à l'origine des différences observées dans l'espace perceptif.

A travers ce chapitre, deux points importants sont à souligner. Le premier point est que la méthode de comparaison avec permutation de références semble permettre de retrouver les dimensions découvertes grâce à la méthode standard de comparaison par paires. Le deuxième point concerne l'utilisation de la méthode de comparaison avec permutation de références pour l'évaluation d'un corpus étendu. Nous avons montré que l'espace perceptif résultant de notre test sur les 37 enceintes permettrait d'obtenir deux dimensions communes avec le test sur 12 enceintes, plus une dimension supplémentaire. Ces deux observations semblent valider la capacité de la méthode à évaluer des dissemblances sur un corpus étendu de stimuli.

Chapitre 6

Application du protocole pour l'évaluation de la distorsion non linéaire

Sommaire

6.1	Test 1 : Évaluation d'un corpus de 31 enceintes distordues	111
6.1.1	Stimuli	111
6.1.2	Auditeurs	113
6.1.3	Procédure	114
6.1.4	Résultats	114
6.1.5	Bilan intermédiaire	122
6.2	Test 2 : Évaluation d'un corpus de 19 enceintes distordues	122
6.2.1	Stimuli	122
6.2.2	Auditeurs	122
6.2.3	Procédure	122
6.2.4	Résultats	123
6.3	Discussion	124

Après avoir testé la méthode de comparaison avec permutation de références sur la restitution du timbre au chapitre 5, ce chapitre concerne l'application du protocole complet à la caractérisation des distorsions non linéaires. La méthode de comparaison avec permutation de références a donc été utilisée pour évaluer la dissemblance perçue sur deux corpus de 31 puis 19 enceintes virtuelles distordues. Les espaces perceptifs que nous obtenons permettent de vérifier notre protocole pour l'évaluation de la distorsion non linéaire et donne quelques résultats préliminaires quant aux effets perçus de la distorsion non linéaire.

6.1 Test 1 : Évaluation d'un corpus de 31 enceintes distordues

6.1.1 Stimuli

Le choix de l'extrait sonore est un point important car, en fonction du type de non-linéarité appliqué, le contenu temporel et spectral peut être plus ou moins affecté. Un extrait riche spectralement et dont la forme temporelle présente plusieurs variations de dynamique peut permettre de mettre en évidence de nombreux phénomènes. Après plusieurs écoutes informelles, nous avons choisi, parmi les différents morceaux enregistrés, un extrait du morceau de jazz de Mc Coy Tyner "Miss bea". Cet extrait contient quatre instruments : une batterie, une contrebasse, un piano et un saxophone. De plus, il contient des passages dynamiques avec une combinaison basse-batterie et des passages plus mélodiques avec le piano ou le saxophone. Cet extrait dure 3.77s, ce qui est du même ordre de grandeur que dans les études précédentes sur l'évaluation de la distorsion non linéaire.

A partir de l'extrait choisi, un corpus de 31 stimuli a été constitué. Nous avons créé 30 enceintes virtuelles à partir des trois manières d'introduire de la distorsion présentées dans le chapitre 3. Nous avons aussi intégré au corpus l'enregistrement de l'enceinte réelle effectué par Lavandier [2005] afin de vérifier si elle est bien considérée proche des enceintes virtuelles.

Six enceintes désignées I ont été obtenues en modifiant les amplitudes des réponses d'ordre supérieur issues de l'identification du boomer. Le paramètre associé à ce type de non-linéarité est le coefficient α présenté dans la section 3.7. Il permet de modifier l'amplitude de toutes les composantes non linéaires identifiées sur le boomer.

Les enceintes B et E ont été obtenues en introduisant des phénomènes non linéaires artificiels représentés par des courbes entrée/sortie "sans mémoire" comparables à celles utilisées dans la littérature. Nous avons ainsi choisi quatre types de non-linéarités : les non-linéarités de Clipping, de Zero Crossing, Carrée et Cubique. Les non-linéarités de Clipping et de Zero Crossing sont deux formes souvent employées pour introduire de la distorsion [Tan *et al.*, 2003; Geddes et Lee, 2003a] et nous avons ajouté deux formes polynomiales inspirées des non-linéarités évaluées par de Santis et Henin [2007]. Les représentations des différentes courbes entrée/sortie des non-linéarités

sont présentées sur la figure 6.1.

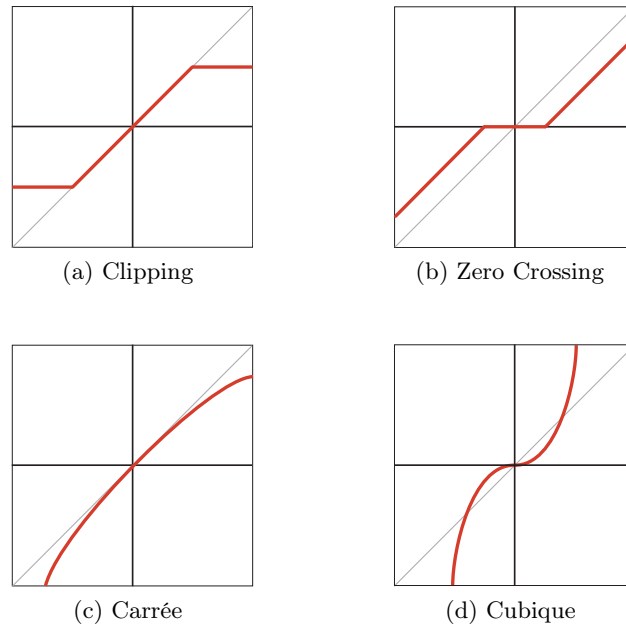


Figure 6.1 – Représentations des différentes non-linéarités artificielles.

Chaque type de non-linéarité est issu d'une équation qui relie le signal de sortie y avec le signal d'entrée x . Les équations 6.1, 6.2, 6.3 et 6.4 expriment la relation entrée/sortie pour ces non-linéarités.

- Clipping

$$y = \begin{cases} -a & \text{si } x < -a \\ x & \text{si } -a < x < a \\ a & \text{si } x > a \end{cases} \quad (6.1)$$

- Zero crossing

$$y = \begin{cases} x + a & \text{si } x < -a \\ 0 & \text{si } -a < x < a \\ x - a & \text{si } x > a \end{cases} \quad (6.2)$$

- Carrée

$$y = -ax^2 + x \quad (6.3)$$

- Cubique

$$y = ax^3 + x \quad (6.4)$$

Douze enceintes B ont été générées en introduisant les différentes courbes entrée/sortie dans la réponse du boomer, et douze enceintes E ont été obtenues en introduisant les mêmes phénomènes mais dans la réponse de l'enceinte complète. Le tableau 6.1 récapitule les différents stimuli avec pour chacun une valeur du paramètre gouvernant la non-linéarité. Pour les non-linéarités de

Clipping, le paramètre présenté dans le tableau 6.1 correspond au pourcentage de temps où la non-linéarité est appliquée au signal. C'est le paramètre qui est généralement utilisé pour exprimer la quantité de signal qui est touchée par ce type de non-linéarité [Tan *et al.*, 2003]. Le paramètre utilisé pour définir le degré de non-linéarité du Zero Crossing est la valeur correspondant à un pourcentage de la valeur RMS du signal d'entrée [Tan *et al.*, 2003]. Pour les deux autres types de non-linéarité, le paramètre a correspond à celui exprimé dans les équations définissant les non-linéarités. Les valeurs de ces paramètres " a " ne sont absolument pas comparables entre les différentes méthodes.

Type d'enceinte	Type de non-linéarité	No.	Paramètres	
Enceinte réelle	—	1	—	
Enceintes virtuelles	Enceintes I		2 à 7	$\alpha = 0, 1, 3, 5, 7, 9$
	Enceintes B	Carrée	8-9-10	$a = 1.2, 1.6, 2$
		Clipping	11-12-13	20, 30, 40 % du temps
		Cubique	14-15-16	$a = 1.4, 2.5, 4.5$
		Zero Crossing	17-18-19	20, 30, 40 % de la valeur RMS
	Enceintes E	Carrée	20-21-22	$a = 0.4, 0.6, 0.8$
		Clipping	23-24-25	2, 5, 15 % du temps
		Cubique	26-27-28	$a = 0.6, 1, 1.4$
Zero Crossing		29-31-31	10, 15, 20 % de la valeur RMS	

Tableau 6.1 – Récapitulatifs de la nature des stimuli du test 1 et du type de non-linéarité appliqué.

Après quelques premières écoutes informelles des stimuli, nous avons ainsi décidé d'appliquer des non-linéarités qui engendrent des distorsions nettement audibles par tous les auditeurs. Certains stimuli s'écartent alors significativement du fonctionnement normal d'une enceinte hi-fi, et de telles enceintes pourraient être mal notées dans un test évaluant la préférence. Ce premier test d'écoute a donc avant tout comme but de tester le protocole proposé.

6.1.2 Auditeurs

Vingt auditeurs (3 femmes et 17 hommes) ont participé à ce test, ils étaient tous membres du laboratoire. Tous les auditeurs présentaient un audiogramme normal. Un seul d'entre eux avait de l'expérience dans l'écoute musicale mais pas précisément dans l'évaluation de la distorsion des enceintes acoustiques.

6.1.3 Procédure

Le test d'écoute a été réalisé avec la méthode de comparaison avec permutation de références pour évaluer la dissemblance perçue entre les différents stimuli. Pour ce corpus de 31 enceintes acoustiques, chaque référence était comparée aux 30 stimuli restant distribués aléatoirement en 10 trios. Le test entier comprenait 310 essais (10x31 stimuli). Chaque auditeur effectuait le test en 4 sessions d'une durée moyenne de 30 minutes. L'interface employée pour le déroulement du test était identique à celle mise au point pour les tests d'écoute effectués lors de l'élaboration de la méthode de comparaison avec permutation de références telle qu'elle est présentée dans le chapitre 5. Le test d'écoute a été réalisé dans une cabine audiométrique et les conditions de restitution (carte son M-Audio Delta 66 et casque Stax SR Lambda Professional) étaient aussi identiques à celles utilisées lors des tests sur la méthode présentés au chapitre 5. Le casque que nous avons utilisé ne semble pas introduire de distorsion notable dans les échantillons. La mesure des niveaux des harmoniques 2 et 3 sur une plage de fréquence de 100 à 7000 Hz sont au minimum à -60 dB. Le protocole expérimental et les détails concernant cette vérification sont donnés dans l'annexe E. Étant donné que la problématique de cette étude porte sur la recherche d'une mesure de distorsion qui soit pertinente d'un point de vue perceptif, la caractérisation du casque est donnée ici à titre indicatif. Nous avons ainsi considéré la distorsion du casque comme assez faible en comparaison de la distorsion des échantillons évalués, donc non susceptible d'empêcher leur comparaison.

Avant d'effectuer le test, tous les stimuli étaient égalisés afin de s'assurer que le niveau de restitution ne rentrerait pas en compte dans les jugements de dissemblance des auditeurs. Nous avons fait une première égalisation à "l'oreille" puis nous avons validé cette égalisation à partir d'un calcul de sonie de sons non-stationnaires proposé par Glasberg et Moore [2002]. Nous avons enregistré les 31 signaux retransmis à travers le casque avec l'oreille artificielle présentée dans l'annexe E. A partir de ces enregistrements, nous avons calculé la sonie au cours du temps pour chaque signal et effectué la moyenne sur la durée. Nous obtenons un niveau moyen pour la voie droite de 83 phons et 82 phons pour la voie gauche, avec un écart-type pour les deux voies égal à 0.3 phons. Cette vérification montre que l'égalisation globale de sonie telle que nous l'avons effectuée est relativement précise. Ce niveau d'écoute a été fixé lors d'écoutes informelles, de manière à permettre une bonne discrimination des signaux. Il a été jugé confortable par tous les auditeurs.

6.1.4 Résultats

A partir des jugements des auditeurs, la matrice de dissemblances a été obtenue en effectuant la moyenne des matrices individuelles. Nous avons représenté sur la figure 6.2 les données de dissemblances sur un dendrogramme issu de l'analyse en arbre hiérarchique. Le dendrogramme

ne fait pas apparaître de regroupements nets de stimuli. Nous définissons un groupe comme un ensemble de stimuli jugés similaires entre eux et apparaissant à l'écart du reste du corpus. Un tel groupe pourrait alors signifier la présence d'une catégorie de sons qui ne partagerait pas les dimensions perceptives des autres sons. A la vue de l'arborescence du dendrogramme, aucun groupement de sons n'est révélé. L'analyse MDS semble donc pertinente pour mettre en évidence des dimensions continues, implicitement utilisées par les auditeurs pour juger la dissemblance entre les sons.

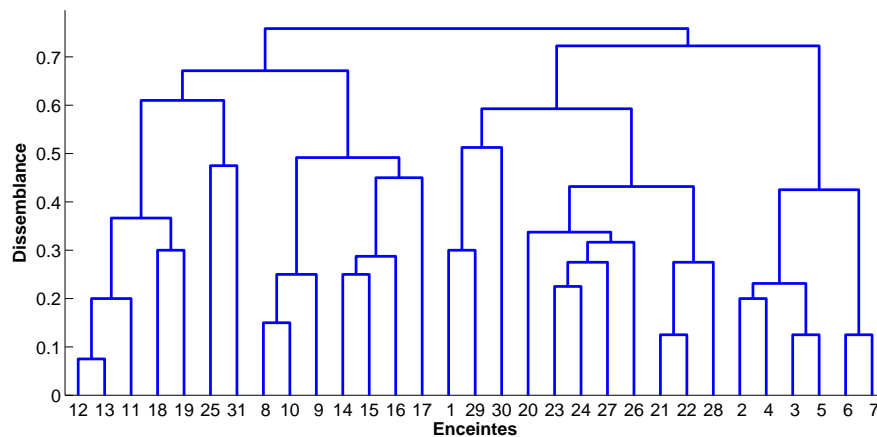


Figure 6.2 – Dendrogramme résultant de l'analyse en arbre hiérarchique des dissemblances issues du test sur les 31 enceintes.

La matrice de dissemblance issue du test a été soumise à une analyse MDS afin de révéler les dimensions de l'espace perceptif. Nous avons utilisé le modèle métrique MDSCAL, similaire à celui employé pour les analyses effectuées au chapitre 5. L'analyse de la courbe de stress indique que 4 dimensions semblent suffire pour représenter les données de dissemblance dans un espace perceptif (voir annexe D).

Pour interpréter ces dimensions d'un point de vue perceptif, nous avons procédé à une écoute informelle des stimuli le long de chaque dimension de l'espace (voir figure 6.3). L'interprétation permet d'évaluer la nature de l'évolution des stimuli suivant chaque dimension et nous avons attribué un terme qui évoque cette dimension (il est évident que les termes que nous avons choisis sont arbitraires, d'autres termes pourraient être utilisés). Nous rappelons que ce test a été mené sur un seul extrait musical et que l'interprétation que nous donnons ne peut être généralisée à l'évaluation de la distorsion non linéaire dans l'absolu.

Dimension 1 : Il semblerait que la dimension 1 soit liée à un degré de détérioration ou de brouillage du signal perçu. Cette dimension évolue depuis des stimuli qui semblent peu affectés (partie droite sur la figure 6.3) vers les stimuli qui présentent une détérioration très audible

(partie gauche sur la figure 6.3). Suivant cette dimension, l'augmentation de détérioration est aussi associée à une région spectrale affectée de plus en plus large. Sur la partie droite de la dimension 1 identifiée ici, sont d'abord touchés par la modification des instruments avec un spectre basse fréquence (basse et grosse caisse). En se déplaçant vers la gauche, les instruments affectés par la distorsion sont de plus en plus nombreux : d'autres instruments à spectre plus aigu (saxophone, piano puis cymbales) sont progressivement détériorés. La détérioration du signal le long de cette dimension s'accompagne de bruit additionnel résultant de l'effet de différents types de distorsion. La combinaison de la détérioration et du bruit donne une sensation de dégradation ayant pour conséquence de brouiller l'écoute en se déplaçant vers les paramétrages de distorsion les plus élevés.

Dimension 2 : La deuxième dimension semble plutôt refléter la "position" dans le spectre des artefacts de la distorsion. En écoutant les sons le long de cette dimension en se basant sur la figure 6.3, pour une coordonnée de la dimension 1 (une intensité de distorsion perçue et une largeur spectrale), la bande fréquentielle de distorsion semble se décaler des fréquences les plus basses vers les plus aiguës.

Dimension 3 : La troisième dimension de l'espace perceptif est liée à l'évolution temporelle du signal musical, plus précisément une modification des transitoires présentes dans l'extrait. Ce sont les instants de l'extrait qui ont les amplitudes les plus fortes qui sont en général modifiés selon cette dimension. La dimension progresse depuis des stimuli possédant des attaques et décroissances qualifiées de "molles" vers des stimuli ayant les attaques et décroissances "normales" puis "exagérées" (partie haute sur la figure 6.3).

Dimension 4 : La quatrième dimension traduit une évolution de la présence des basses fréquences sur les passages de faible niveau. D'un côté de la dimension, sont situés les extraits dont les instruments les plus graves sont fortement présents (bas de la dimension sur la figure 6.3) et de l'autre, cette présence de basses est atténuée.

Dans l'espace perceptif présenté sur la figure 6.3, nous avons fait ressortir les enceintes I et l'enceinte réelle pour une analyse spécifique. Ces enceintes I dont le boomer a été modifié à partir de l'identification sont situées à l'extrémité de la dimension 1. Cette dimension reflète le niveau perçu de détérioration du signal. Le fait que ces enceintes se retrouvent du côté droit de la dimension signifie qu'elle n'ont pas été perçues comme étant détériorées au sens brouillées tel que nous l'avons introduit. Suivant la dimension 2, ces enceintes se retrouvent sous la forme d'un groupe. Étant donné qu'elles proviennent de la modification du boomer, elles se situent dans la partie de l'espace où l'influence de la dimension 2 est plus importante sur le bas du spectre. Suivant la dimension 3, leur répartition est un peu plus étendue mais les enceintes

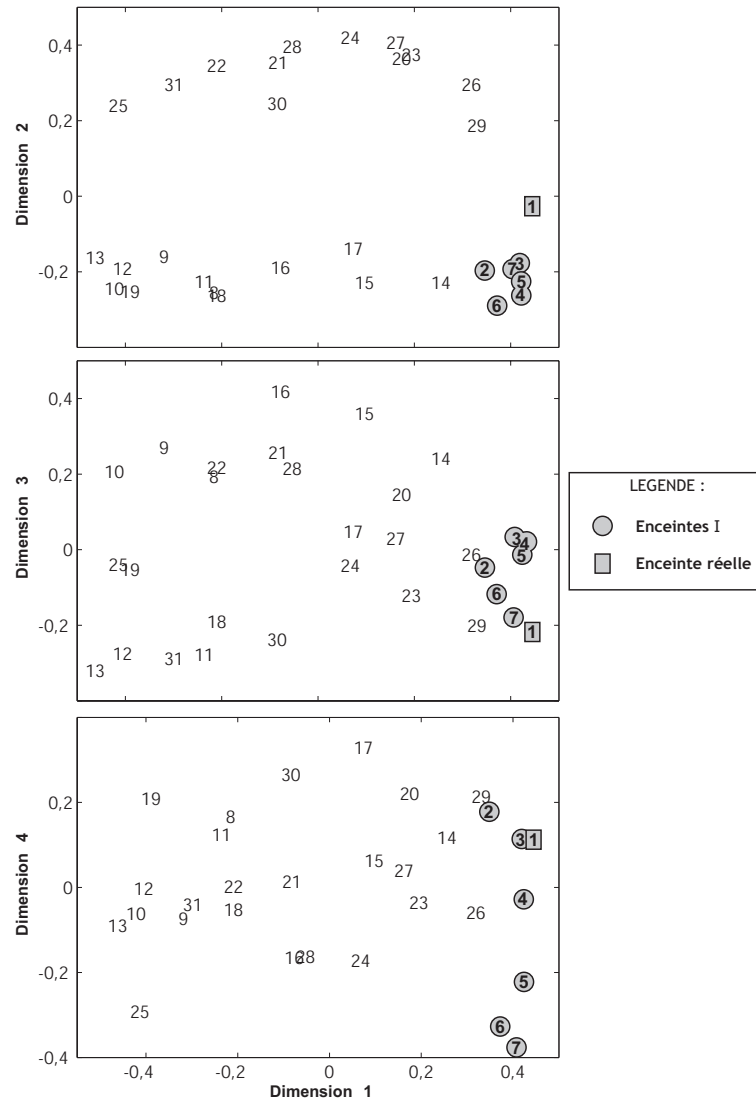


Figure 6.3 – Espace perceptif résultant de l’analyse MDS des données de dissemblances du test 1. L’espace est présenté suivant les dimensions 1 et 2 (en haut), suivant les dimensions 1 et 3 (en bas) et suivant les dimensions.

restent assez proches. L’augmentation du paramètre de la non linéarité a eu un faible effet sur la modification des attaques présentes dans le signal. Il est tout de même possible de retrouver une certaine progression depuis l’enceinte 2 jusqu’à l’enceinte 7. La position de ces enceintes suivant la dimension 4 montre une nette évolution de l’enceinte 2 à l’enceinte 7. L’enceinte 2 correspond à une enceinte dont les non-linéarités du boomer ont été retirées ($\alpha = 0$). En augmentant graduellement les non-linéarités identifiées sur le boomer, les enceintes 3, 4, 5, 6 et 7 sont réparties de manière assez continue le long de la dimension 4 de l’espace perceptif.

La position de ces enceintes dans l’espace perceptif indique que l’évolution du paramétrage n’a pas été perçue comme une sensation de détérioration telle que nous l’avons décrite. Leur

répartition suivant la dimension 4 montre que les auditeurs ont alors perçus principalement le paramétrage comme un renforcement des basses dans l'extrait. Pour ces enceintes obtenues en modifiant les non-linéarités du boomer identifié, l'augmentation du paramètre α se traduit clairement par une augmentation de la présence des basses. L'augmentation des non-linéarités dans le boomer identifié crée ainsi un déséquilibre entre le boomer et le tweeter. Il s'agit alors d'une modification spectrale similaire à une distorsion linéaire. Ce résultat semble provenir du fait que le comportement de l'enceinte au niveau auquel nous l'avons identifié, est faiblement non linéaire.

L'enceinte 3 correspond à l'enceinte modifiée avec un coefficient $\alpha = 1$. Cette enceinte virtuelle, normalement similaire à l'enceinte réelle (numéro 1) n'en est pas forcément proche suivant toutes les dimensions. Elles apparaissent uniquement très proche dans le plan 1-4 de l'espace perceptif. Le fait qu'elles n'ont pas été confondues suivant chaque dimension montre que les auditeurs ont perçu des différences. Les tests ABX menés au cours du chapitre 3 pour valider l'élaboration de l'enceinte virtuelle indiquaient que les auditeurs ne percevaient pas de différences entre une enceinte réelle et une enceinte virtuelle. Il faudrait se pencher davantage sur la tâche correspondant à ces deux méthodes pour expliquer le fait que ces enceintes 1 et 3 soient perçues différemment suivant ces différentes dimensions, mais pas avec le test ABX.

Pour les enceintes B (numéro 8 à 19) et les enceintes E (numéro 20 à 31), l'espace tel qu'il est présenté sur la figure 6.3 ne permet pas de séparer facilement des informations par rapport à la position des enceintes de chaque type de non-linéarité. Afin de faciliter l'analyse, nous avons représenté l'espace suivant les dimensions 1-2, 1-3 et 1-4, les non-linéarités Carrée et de Clipping sur la figure 6.4 et les non-linéarités Cubique et de Zero Crossing sur la figure 6.5, en faisant ressortir les numéros des enceintes correspondants. Sur cette figure, les numéros des enceintes que nous cherchons à mettre en avant ont été entourés et les numéros des autres enceintes ont été grisés. Les numéros entourés d'un trait plein correspondent aux enceintes B et ceux entourés avec un trait en pointillés représentent les enceintes E.

L'enceinte numéro 2 correspond à l'enceinte I issue de l'identification mais pour laquelle le coefficient α est nul, donc son boomer a été linéarisé. Elle est similaire à une enceinte de type B pour laquelle une relation entrée/sortie linéaire aurait été appliquée. Le modèle du boomer est seulement composé du signal d'entrée et d'un filtre $h_1(t)$ caractérisant la partie linéaire du boomer. Elle est donc commune aux deux types d'enceintes et apparaît comme le point de départ des non-linéarités introduites dans les enceintes B. Un point de départ des enceintes E correspondrait à l'enceinte 2 pour laquelle la réponse du tweeter serait aussi linéarisée. En considérant les non-linéarités du tweeter comme négligeables par rapport à celles du boomer, cette enceinte de départ serait similaire à l'enceinte 2. Elle serait dans tous les cas très proche de l'enceinte 2 dans l'espace perceptif. Afin d'observer l'évolution des paramétrages de distorsion, nous avons donc choisi l'enceinte 2 comme point de départ pour les enceintes B et E. Le numéro de cette

enceinte a été entouré d'un double trait.

L'évolution des positions des enceintes dans l'espace nous informe sur le lien entre les paramètres que nous avons choisis pour les différents types de distorsions et la manière dont les auditeurs les ont perçus. Il s'avère que l'augmentation des paramètres se traduit par une évolution graduelle de la position des enceintes suivant les différentes dimensions. Les enceintes élaborées avec les paramètres les plus faibles se retrouvent proches de l'enceinte 2 dont le boomer est linéarisé.

Les différents types de non-linéarités (Carrée, Clipping, Cubique et Zero Crossing) utilisés pour les enceintes B et E n'apparaissent pas sous la forme de groupes dans l'espace perceptif. Ainsi, les auditeurs ont bien perçu les effets communs de ces non-linéarités. Ils ont jugé les sons selon la "quantité" de distorsion introduite et non selon le type de non-linéarité. Dans le plan des dimensions 1-3 et 1-4, les évolutions des enceintes E et B présentent des tendances globalement similaires : pour un type de non-linéarité donné, les enceintes E et B évoluent de manière quasi parallèle dans une même direction de l'espace perceptif. Par contre, suivant la dimension 2, les enceintes E se situent pour chaque type de non-linéarité dans la partie supérieure du plan 1-2, et les enceintes B dans la partie inférieure. De ce fait, cette dimension semble refléter le mécanisme que nous avons introduit.

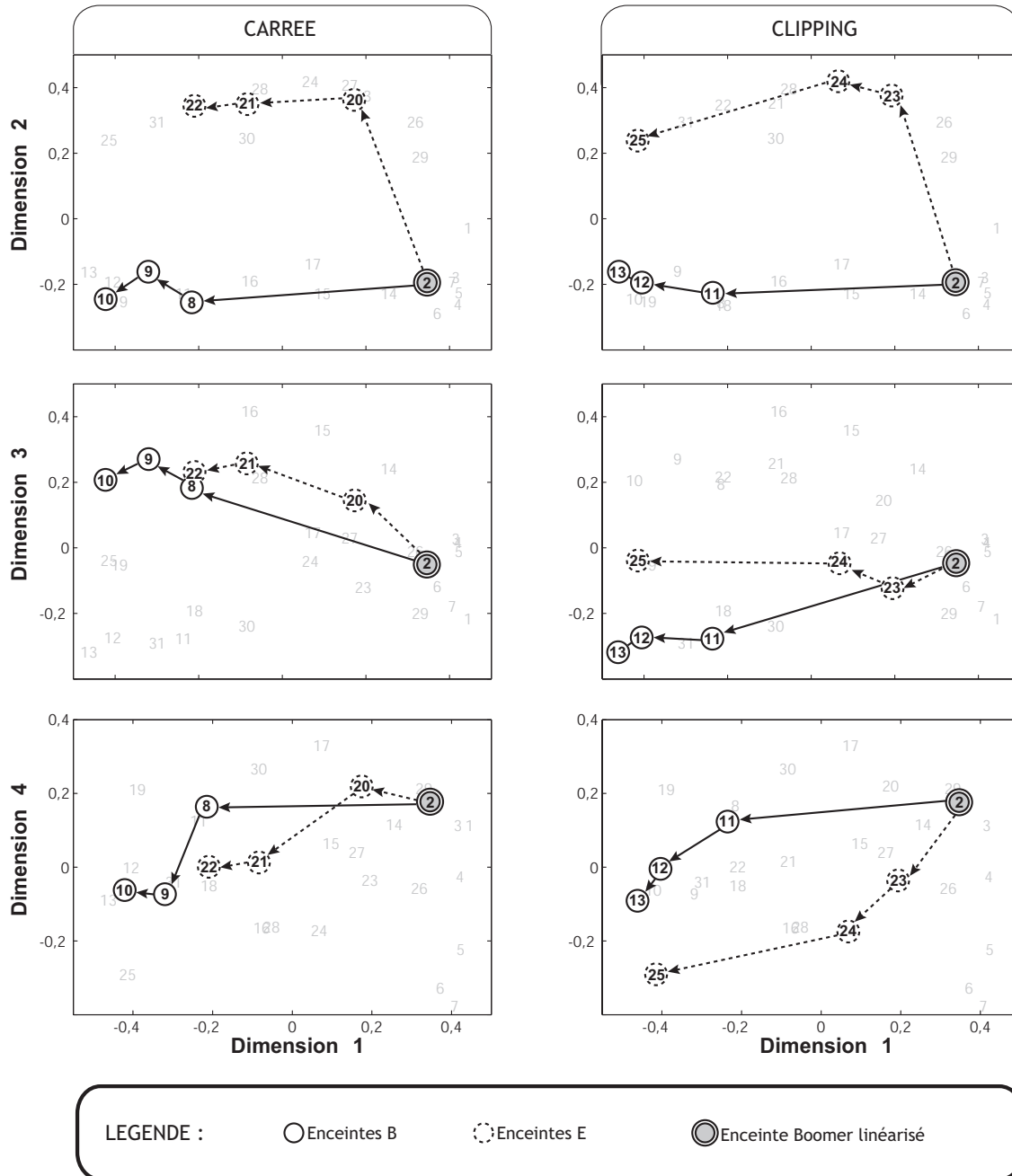


Figure 6.4 – Espace perceptif suivant les dimensions 1-2 (à haut), 1-3 (au milieu) et 1-4 (en bas) pour les non-linéarités Carrée (à gauche) et de Clipping (à droite). Sont accentuées les enceintes B (non-linéarités artificielles appliquées au boomer), les enceintes E (non-linéarités artificielles appliquées à l'enceinte complète) et l'enceinte 2 dont le boomer est linéarisé.

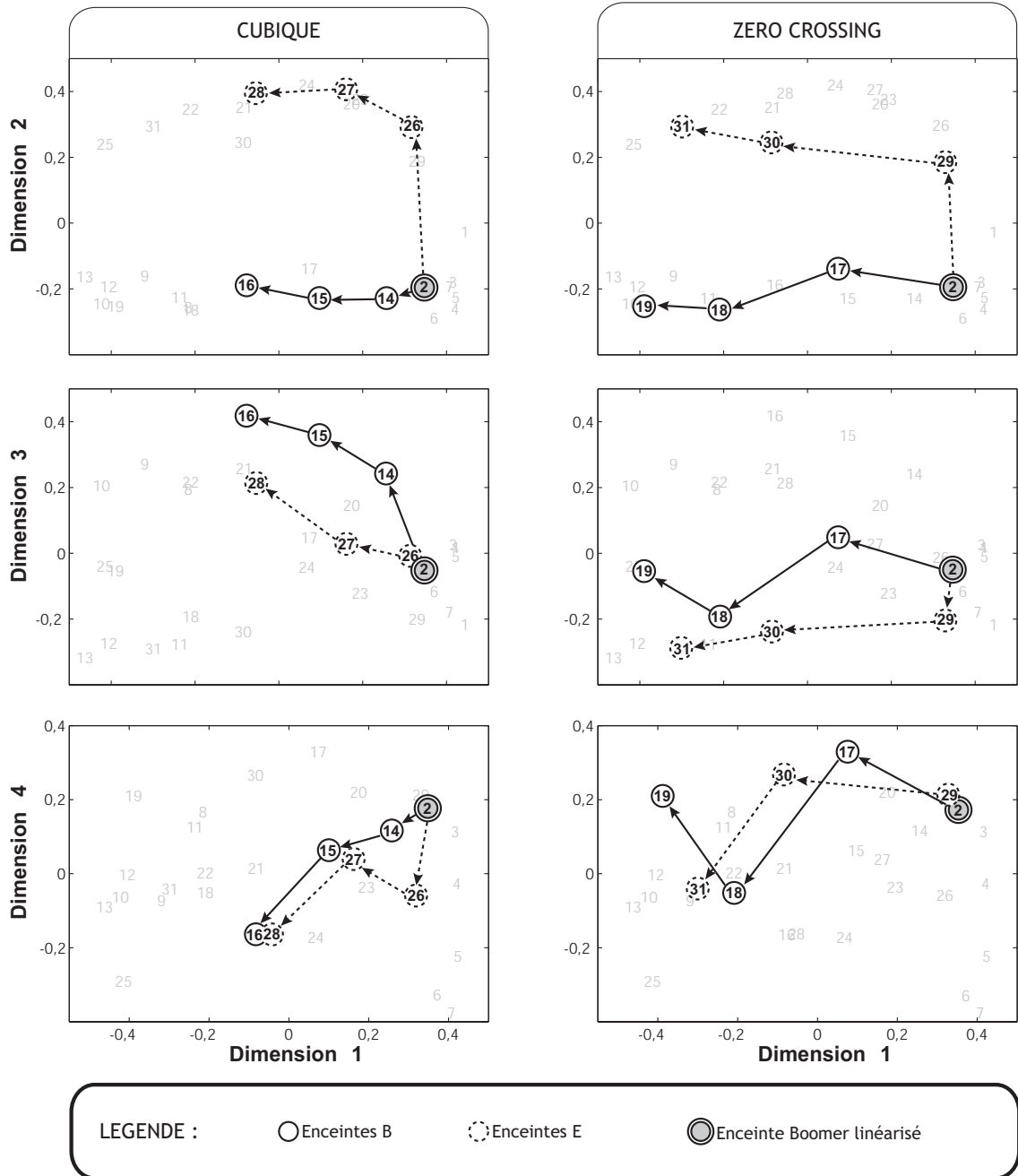


Figure 6.5 – Espace perceptif suivant les dimensions 1-2 (en haut), 1-3 (au milieu) et 1-4 (en bas) pour les non-linéarités Cubique (à gauche) et de Zero Crossing (à droite). Sont accentuées les enceintes B (non-linéarités artificielles appliquées au boomer), les enceintes E (non-linéarités artificielles appliquées à l'enceinte complète) et l'enceinte 2 dont le boomer est linéarisé.

6.1.5 Bilan intermédiaire

Ce premier test était destiné à évaluer différentes enceintes plus ou moins distordues et à comparer les différentes manières de générer des distorsions. On remarque que suivant les dimensions 1-3 et 1-4, les stimuli sont répartis de manière homogène. Les valeurs des paramètres choisis pour chaque type d'enceinte nous permettent de paver correctement l'espace suivant ces deux dimensions. Par contre, suivant la dimension 2, il apparaît une franche divergence entre deux groupes d'enceintes : la partie supérieure de l'espace contient les enceintes distordues sur la totalité du spectre et la partie inférieure celles dont seul le boomer était modifié par les non-linéarités. Il semble alors que cette dimension ne reflète pas les conséquences de la distorsion mais plutôt la manière dont les enceintes ont été créées. Ce groupe d'enceintes E aurait effectivement été jugé différemment car les non-linéarités ont été ajoutées sur la réponse complète de l'enceinte contrairement aux enceintes I et B dont seul le modèle de boomer est modifié. Afin de confirmer l'espace perceptif des distorsions en éliminant l'effet dû à la dimension 2 (position spectrale de la dégradation), nous avons décidé de mener un deuxième test en conservant seulement les enceintes virtuelles dont le boomer est modifié en éliminant les enceintes E.

6.2 Test 2 : Évaluation d'un corpus de 19 enceintes distordues

6.2.1 Stimuli

Pour ce test, nous avons écarté du corpus les enceintes dont les non-linéarités artificielles étaient appliquées sur la totalité du spectre de l'enceinte. Le corpus des 19 enceintes virtuelles restantes était composé des enceintes I et B et de l'enceinte réelle. Les valeurs des paramètres des non-linéarités appliquées au modèle de boomer sont identiques à celles utilisées pour le test 1.

6.2.2 Auditeurs

Vingt auditeurs normo-entendants (4 femmes et 16 hommes) ont participé à ce deuxième test. Tout comme ceux du test 1, ils étaient membres du laboratoire. Douze d'entre eux avait participé au premier test et les autres n'avaient aucune expérience dans l'évaluation de la distorsion.

6.2.3 Procédure

Pour ce test faisant intervenir 19 stimuli, chaque référence était comparée à 18 stimuli distribués aléatoirement en 6 trios pour chaque auditeur. Le test entier était composé de 114 essais répartis en 3 séances de 15 minutes environ. Les conditions de tests étaient identiques à celles du test 1.

6.2.4 Résultats

L'analyse en arbre hiérarchique de la matrice de dissemblance est présentée sur la figure 6.6. Aucun groupe net ne peut être distingué sur ce dendrogramme, ce qui valide à nouveau l'utilisation de la MDS pour l'analyse des données de dissemblance de ce deuxième test.

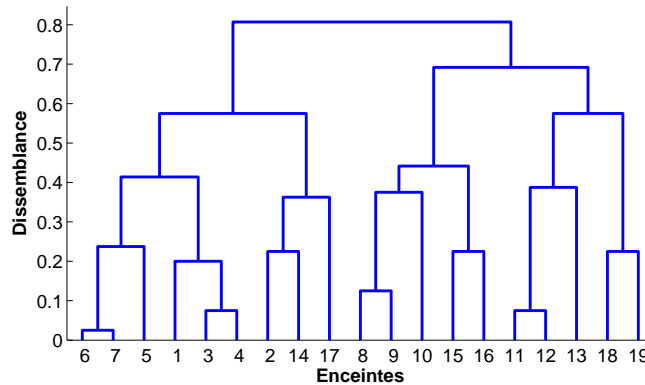


Figure 6.6 – Dendrogramme résultant de l'analyse en arbre hiérarchique des dissemblances issues du test sur les 19 enceintes.

La matrice de dissemblance obtenue à partir des jugements des 20 auditeurs a été soumise à la même analyse MDS que celle utilisée au cours du test 1. L'espace perceptif résultant est composé de 3 dimensions sur lesquelles les auditeurs ont basé leurs jugements (la courbe de stress associée à cette analyse est présentée dans l'annexe D). La figure 6.7 présente l'espace perceptif de ce deuxième test suivant les dimensions 1-2 et 1-3 avec les enceintes I et l'enceinte réelle mises en évidence.

De la même manière que pour le test 1, nous avons écouté les sons le long des trois dimensions de l'espace perceptif issu de ce deuxième test. La première dimension semble similaire à la première dimension obtenue dans le test 1. Cette dimension reste la dimension prépondérante utilisée par les auditeurs pour différencier les stimuli. Le long de cette dimension les caractéristiques ressenties sont à nouveau l'évolution d'un certain niveau de détérioration du signal et l'élargissement de la région spectrale où la distorsion intervient dans le signal (de droite à gauche sur la figure 6.7). La deuxième dimension issue de ce deuxième test est similaire à la dimension 3 du test 1. En effet, les stimuli le long de cette dimension présentent des modifications audibles sur les transitoires du signal. La dimension évolue, de haut en bas sur la figure 6.7, des stimuli avec des attaques et des décroissances "molles" vers des stimuli présentant des attaques et décroissances normales puis "exagérées". La troisième dimension obtenue pour les 19 enceintes est similaire à la dimension 4 du test 1 sur les 31 enceintes. La présence de basses sur les passages de faible niveau dans l'extrait est effectivement modifiée le long de cette dimension.

Sur la figure 6.7, nous avons fait ressortir les enceintes I (numéro 2 à 7) et l'enceinte réelle

pour une première analyse de leur position. Elles sont organisées d'une manière assez similaire dans l'espace issu du test 2 suivant les dimensions 1-2 et 1-3 que dans celui issu du test 1 suivant les dimensions 1-3 et 1-4.

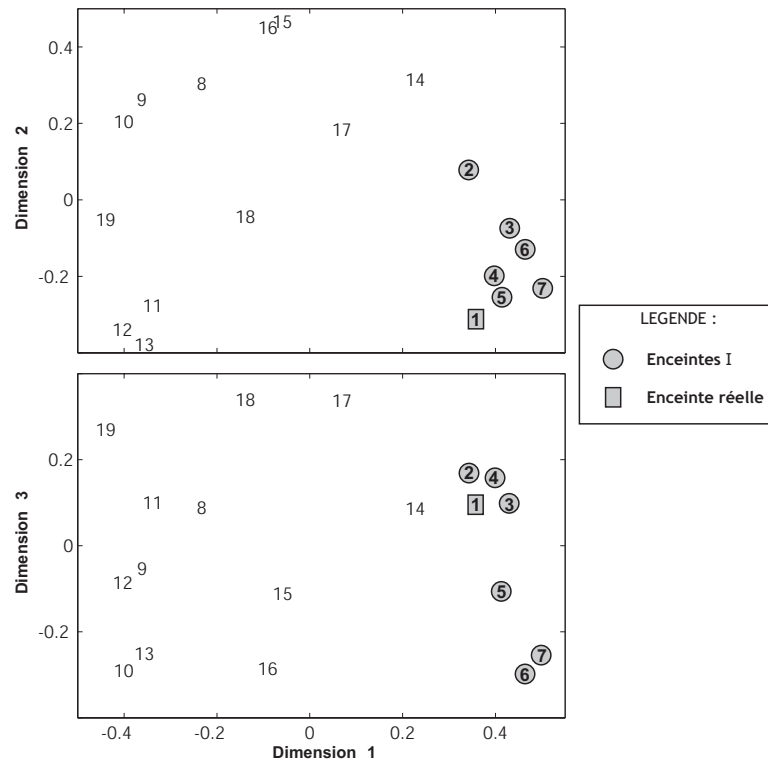


Figure 6.7 – Espace perceptif résultant de l'analyse MDS des données de dissemblances du test 2.

Afin de vérifier la similarité entre les espaces perceptifs des deux tests, les figures 6.8 et 6.9 montrent l'espace perceptif pour chaque type de non-linéarités des enceintes B, suivant les dimensions 1-2 et 1-3 pour ce test comparé à l'espace suivant les dimensions 1-3 et 1-4 pour le test 1. Nous avons fait ressortir la position en entourant les enceintes B (numéro 8 à 19) en fonction du type de non-linéarités.

La comparaison des espaces issus des deux tests effectuée sur les figure 6.8 et 6.9 fait ressortir une certaine ressemblance entre les résultats des deux tests. Pour chaque type de non-linéarité artificielle, l'évolution des enceintes avec les paramètres de non-linéarité présente une tendance similaire à celle des enceintes du test 1.

6.3 Discussion

Bien que préliminaires, ces premiers tests perceptifs effectués avec le protocole que nous proposons illustrent le caractère multidimensionnel de la distorsion non linéaire. La dimension

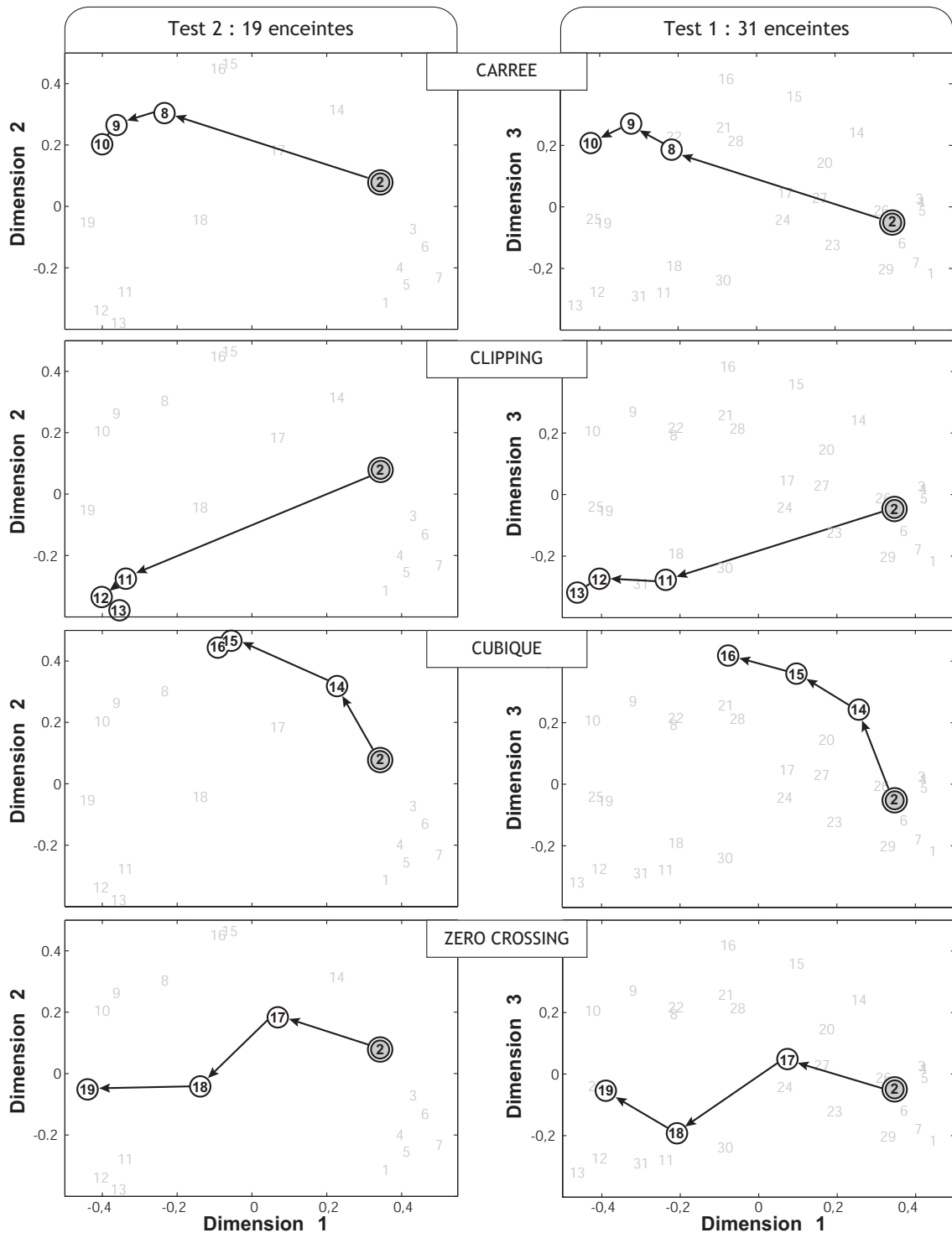


Figure 6.8 – Espace perceptif suivant les dimensions 1-2 pour le test 2 (à gauche) et les dimensions 1-3 pour le test 1 (à droite) pour comparaison. Les enceintes B sont accentuées (non-linéarités artificielles appliquées au boomer) pour comparer l'évolution de chaque type de non-linéarité.

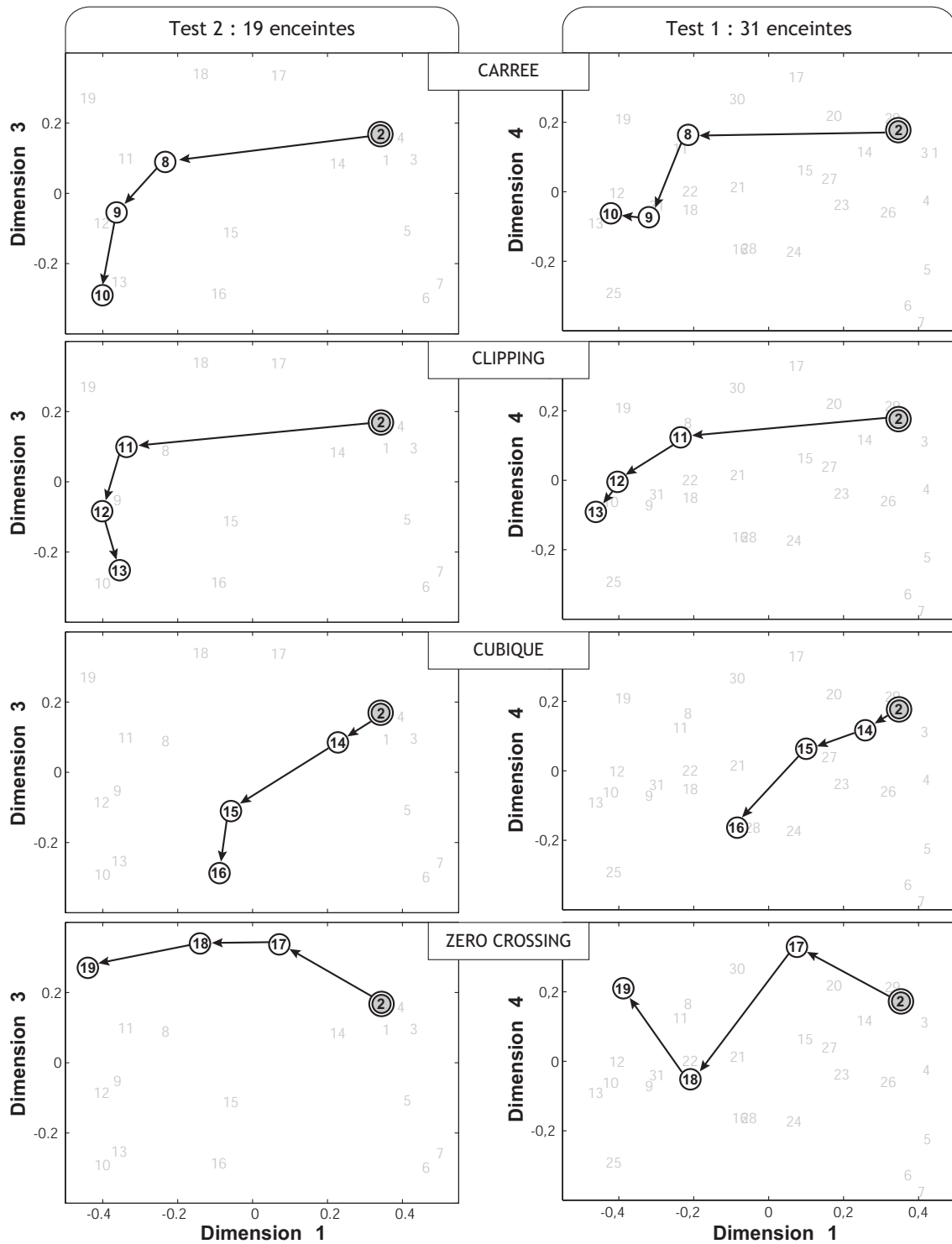


Figure 6.9 – Espace perceptif suivant les dimensions 1-3 pour le test 2 (à gauche) et les dimensions 1-4 pour le test 1 (à droite) pour comparaison. Les enceintes B sont accentuées (non-linéarités artificielles appliquées au boomer) pour comparer l'évolution de chaque type de non-linéarité.

prépondérante est la sensation de détérioration perçue et les deux autres dimensions présentent un caractère timbral et temporel. Nous retrouvons des caractéristiques liées à la modification du signal évoquées par Schmitt [1995] : effets de modification du spectre ou coloration, de dynamique, de structure temporelle et l'ajout de bruit additionnel.

Les dimensions de détérioration perçue (dimension 1), d'accentuation de dynamique (dimension 2) et de modification spectrale (dimension 3) sont communes aux deux tests que nous avons menés. La dimension 2 du test 1, liée à la position spectrale de la détérioration, n'a pas été révélée dans les jugements de dissemblances obtenus sur le deuxième corpus. Le fait que cette dimension n'ait pas été obtenue dans les résultats du test 2 indique qu'elle a perdu de l'importance ou qu'elle provenait uniquement de la présence des deux types d'enceintes issues de la modification du boomer ou de l'enceinte complète au sein d'un même corpus. Les jugements des auditeurs auraient alors été orientés vers cette dimension "artificielle" car provenant directement du mécanisme que nous avons choisi d'appliquer à nos stimuli.

L'espace perceptif obtenu pour le test 2 sur les 19 enceintes montre que les paramètres que nous avons choisis pour l'élaboration du corpus d'enceintes distordues permettent de remplir de manière homogène l'espace suivant trois dimensions. Les coefficients appliqués aux différents types de non-linéarités ont permis d'obtenir une évolution graduelle des enceintes au sein de l'espace.

Les enceintes obtenues en introduisant les non-linéarités artificielles correspondent, pour certaines, à des enceintes extrêmement distordues. Pour cette série d'expériences préliminaires nous avons fait en sorte que les différentes distorsions soient audibles par tous les auditeurs. La méthode d'élaboration des enceintes distordues permet ainsi d'étendre un corpus d'enceintes à des échantillons moins réalistes. Cette caractéristique offre la possibilité d'explorer l'espace perceptif pour des "taux" de distorsion peu réalistes et difficilement réalisables lors d'un test sur de vraies enceintes. Nous avons pu évaluer au sein d'un même test des enceintes très distordues avec des enceintes moins distordues et une enceinte réelle.

Il est à noter que les enceintes issues de l'identification du boomer ont été obtenues pour une identification effectuée à un seul niveau de restitution et pour une seule enceinte. Ces enceintes dont l'influence des non-linéarités du boomer était augmentée ne se déplacent pas suivant la dimension liée à la détérioration. Cela montre que le comportement du boomer est peu non linéaire au niveau auquel nous l'avons identifié, et qu'une identification à plus fort niveau serait sans doute nécessaire pour pouvoir avoir davantage de comparaison avec des non-linéarités artificielles.

Ce dernier chapitre a permis de mettre en application le protocole que nous avons élaboré pour la caractérisation de la distorsion non linéaire. La génération d'échantillons distordus (enceintes virtuelles) nous a permis d'étudier divers types de distorsions et nous avons révélé 3

dimensions caractérisant le corpus d'enceintes virtuelles créées. Cette première expérience valide alors l'utilisation de ce protocole pour l'étude des effets de distorsion. Les dimensions que nous obtenons ne sont valables que pour les types de distorsions que nous avons créés. D'autres tests sont nécessaires pour confirmer ces dimensions caractérisant les effets de la distorsion non linéaire, ici, réduit aux phénomènes que nous avons introduits dans nos stimuli. Il faut tout de même noter que les dimensions perceptives sont en accord avec les descriptions mises en évidence par les entretiens informels menés par Schmitt [1995]. Le protocole que nous proposons offre la possibilité d'étendre le nombre et la variété d'échantillons afin de mener des expériences supplémentaires en faisant intervenir davantage de stimuli virtuels et/ou réels. Outre le fait de simuler l'enceinte identifiée avec une composante non linéaire plus importante, d'autres courbes entrée/sortie et d'autres paramètres pourraient être appliqués, il faudrait aussi évaluer d'autres modèles d'enceintes acoustiques. L'application du protocole à ces différentes possibilités pourrait ainsi mettre en évidence de nouvelles dimensions liées aux phénomènes non linéaires de la reproduction d'un extrait musical. Pour compléter la caractérisation de la distorsion non linéaire, il faudrait ensuite trouver les dimensions objectives qui décrivent les dimensions perceptives que nous avons révélées.

Conclusion et perspectives

L'étude présentée dans ce document porte sur l'évaluation perceptive de l'effet des non-linéarités dans la restitution d'un signal musical par des enceintes acoustiques. La majorité des études sur le sujet s'est concentrée sur des mesures de seuils ou des mesures de qualité perçue pour quelques phénomènes de distorsion. Nous avons orienté nos recherches vers une approche multidimensionnelle de la distorsion non linéaire. Dans cette optique, nous avons cherché à révéler les dimensions perceptives mises en jeu lors de jugements de différence perçue entre plusieurs signaux résultant de phénomènes de distorsion différents. Nous avons alors proposé un protocole qui permet d'élaborer un corpus d'enceintes virtuelles dont les phénomènes de distorsion sont paramétrables, et nous avons adapté une méthode d'évaluation de la dissemblance appropriée à l'évaluation de la dissemblance pour des corpus comprenant un grand nombre d'échantillons.

Étant donné la difficulté de contrôler la distorsion d'enceintes réelles, nous avons eu recours à des signaux de synthèse. Nous avons ainsi créé un signal d'enceinte "virtuelle" composé du signal issu d'un modèle de boomer et de l'enregistrement du tweeter reproduit au casque. Les tests d'écoute que nous avons menés montrent qu'en moyenne les auditeurs ne perçoivent pas de différences notables entre des enregistrements d'extraits musicaux joués à travers une enceinte réelle dans une salle donnée et ces mêmes extraits joués à travers notre enceinte virtuelle auralisée avec la réponse impulsionnelle de la même salle. A partir de ce résultat nous avons paramétré le modèle de synthèse pour générer une multitude d'enceintes se différenciant uniquement par leur distorsion. Nous avons associé au sein d'un même corpus des non-linéarités réelles identifiées dans le comportement du boomer et des non-linéarités artificielles. Ces dernières ont été générées à partir de courbes entrée/sortie décrivant le comportement d'un système non linéaire "sans mémoire" auquel nous avons ajouté dans un cas la réponse du boomer et dans l'autre la réponse de l'enceinte complète. Le corpus que nous avons ainsi élaboré comprend des échantillons sonores proches du rayonnement d'une enceinte mais dont les caractéristiques de distorsion étaient très variées.

Afin d'évaluer un corpus comprenant un grand nombre de stimuli, nous avons adopté une méthode alternative que nous avons appelée méthode de comparaison avec permutation de références. Celle-ci consiste à évaluer les différences perçues entre un son de référence et trois autres

sons de comparaison. Cette tâche à choix forcé nous est apparue plus simple que l'évaluation de la dissemblance sur une échelle comme c'est le cas pour la comparaison par paires. Cette méthode permet une évaluation indirecte de la dissemblance et la matrice résultant des jugements des auditeurs est adaptée à une analyse MDS. Nous avons vérifié la capacité de cette méthode à fournir des valeurs de dissemblances à partir de simulations, et nous avons montré que cette méthode permet d'obtenir un espace perceptif similaire à celui obtenu avec la méthode de comparaison par paires sur un même corpus de 12 enceintes. Enfin, l'évaluation d'un corpus de 37 enceintes a permis de montrer que la méthode que nous proposons est adaptée à l'évaluation d'un corpus plus important.

Nous avons enfin effectué deux tests perceptifs avec la méthode de comparaison avec permutation de références sur un corpus d'enceintes que nous avons généré en modifiant ou en introduisant de la distorsion non linéaire. Ces tests ont permis de mettre en application le protocole que nous proposons et de vérifier qu'il est adapté à l'évaluation de la distorsion non linéaire pour l'approche multidimensionnelle que nous avons choisie. A partir de l'analyse MDS des valeurs de dissemblances mesurées, nous avons aussi révélé les dimensions sur lesquelles les auditeurs s'appuient pour leurs jugements de dissemblances. Nous avons révélé trois dimensions liées à la distorsion non linéaire perçue. La répartition des échantillons au sein de l'espace obtenu indique que la procédure de génération des échantillons distordus permet de paver de manière homogène l'espace perceptif. Les résultats que nous avons obtenus ne sont valables que pour les types de non-linéarités évalués et l'échantillon musical choisi. Ces deux tests préliminaires ne fournissent donc pas des résultats définitifs en ce qui concerne l'évaluation des non-linéarités des enceintes acoustiques, mais ils ont permis de mettre en avant l'intérêt du protocole proposé pour étudier les caractéristiques de la distorsion d'un point de vue multidimensionnel plutôt que qualitatif.

Le protocole tel que nous l'avons établi pourrait être amélioré et appliqué pour approfondir l'évaluation des effets de la distorsion dans la restitution des enceintes acoustiques. De plus, il ouvre des pistes de recherche sur l'évaluation de la restitution sonore. L'avantage de notre protocole est d'être composé d'éléments indépendants : choix de l'enceinte génératrice, élaboration et contrôle de l'enceinte virtuelle à partir d'un modèle de synthèse, auralisation et méthode d'évaluation.

Le point de départ de l'élaboration des enceintes virtuelles est le choix d'une enceinte génératrice. Il serait intéressant d'utiliser d'autres enceintes ayant des caractéristiques mécaniques bien différentes de celles de l'enceinte que nous avons utilisée pour cette étude. L'évaluation d'un nouveau corpus issu d'autres enceintes permettrait de vérifier les dimensions perceptives que nous avons révélées et éventuellement de découvrir de nouvelles dimensions.

Les tests d'écoute préliminaires que nous avons menés comprenaient un seul extrait musical. Il serait maintenant nécessaire d'élargir les tests avec d'autres extraits sachant que les phénomènes

non linéaires sont très sensibles au contenu du signal utilisé.

L'enceinte virtuelle que nous avons élaborée est la combinaison d'un modèle de boomer et de l'enregistrement du tweeter d'une même enceinte. Nous n'avons pas pu identifier totalement le comportement du boomer car son modèle a été obtenu qu'à un seul niveau de restitution. Une identification à des niveaux différents nous donnerait la possibilité de modifier davantage le comportement non linéaire du boomer synthétisé.

L'adaptation du modèle pour introduire des distorsions artificielles "sans mémoire" permet d'explorer une plus grande partie de l'espace perceptif et surtout de générer des enceintes qui distordent beaucoup (comportement qui serait difficile à obtenir pour une enceinte de qualité, même via une identification à des niveaux élevés). D'autres formes de relations entrée/sortie pourraient être étudiées, dans le but de rechercher d'autres dimensions éventuelles de l'espace perceptif liée à la distorsion non linéaire. Pour se rapprocher du fonctionnement normal d'une enceinte, il serait intéressant de mener un test d'écoute avec ce type de non-linéarités, mais pour des valeurs plus faibles de distorsion.

L'utilisation d'autres modèles de haut-parleurs serait aussi intéressante pour élargir les possibilités de contrôle des enceintes du corpus à évaluer. Notre protocole pourrait ainsi inclure d'autres modèles de boomer, comme celui proposé par Klippel [2001]. Cette perspective pourrait permettre d'explorer les dimensions de l'espace perceptif plus directement liées à la modification de paramètres de conception des haut-parleurs.

Lors de la phase d'auralisation, nous avons aussi utilisé une réponse impulsionnelle (RI) pour permettre une écoute dans des conditions réalistes. Cette RI provenait d'enregistrements stéréophoniques dans une salle, mais était colorée par la réponse de l'enceinte utilisée. Malgré tout, le test ABX que nous avons effectué a montré qu'une enceinte réellement enregistrée dans cette salle et notre enceinte virtuelle étaient perçues comme très proches. Il serait intéressant de mener une expérience pour s'assurer que la RI caractérise uniquement la réponse de la salle. Cet élément du protocole pourrait aussi être appliqué pour des expériences sur l'étude de la restitution du timbre en modifiant les caractéristiques de la salle ou des prises de sons utilisées pour l'auralisation, comme celle menée par Lavandier [2005] pour la restitution du timbre. Cette possibilité permettrait à partir d'un enregistrement anéchoïque de plusieurs enceintes de faire varier différents facteurs en post-traitement et d'analyser leur influence sur les dimensions de l'espace perceptif.

La méthode de comparaison avec permutation de références que nous avons utilisée pour l'estimation des dissemblances entre les enceintes pourrait être améliorée. Nous avons mis en évidence la présence d'un biais dans l'estimation de la dissemblance lors des simulations. Malgré ce biais, les résultats perceptifs sont cohérents avec ceux obtenus avec des méthodes usuelles. Pour éventuellement réduire ce biais, il faudrait, à partir de nouvelles simulations, étudier la loi qui régit la transformation de dissemblances connues en dissemblances estimées, afin de "linéari-

ser" l'estimation des dissemblances. La méthode de comparaison avec permutation de références pourrait aussi être modifiée par exemple pour laisser à l'auditeur le choix du nombre de réponses possibles au cours d'un essai. Cette variante pourrait éventuellement réduire le biais provenant de l'attribution des valeurs de dissemblance au cours d'un essai.

Enfin, les résultats perceptifs issus du protocole que nous proposons pourraient permettre d'orienter les évaluations physiques de la distorsion afin de trouver des métriques qui soient plus pertinentes d'un point de vue perceptif. Notre protocole pourrait éventuellement être fusionné avec le protocole développé par Lavandier [2005] afin de relier des évaluations physiques et perceptives pour la restitution du timbre et celle de la distorsion non linéaire. Ceci permettrait peut être de proposer des métriques caractérisant simultanément les distorsions linéaire et non linéaire dans la restitution des enceintes acoustiques. De telles métriques permettraient à terme aux constructeurs d'enceintes de modifier la conception de leurs produits sur la base d'évaluations perceptives.

Bibliographie

- AES20 [1996]**. AES recommended practice for professional audio - Subjective evaluation of loudspeakers (Reaffirmed 2007), Audio Engineering Society.
- Aldrich, K.M., Hellier, E.J. et Edworthy, J. [2009]**. What determines auditory similarity? The effect of stimulus group and methodology. *The quarterly journal of experimental psychology*, 62(1), 63–83.
- Atsushi, M. et Martens, W.L. [2002]**. Multidimensional perceptual calibration for distortion effects processing software. *In Proc. AES 113th Convention*, n ° 5708.
- Bech, S. et Zacharov, N. [2006]**. *Perceptual audio evaluation - Theory, method and application*. Wiley.
- Belcher, R.A. [1978]**. A new distortion measurement - Better subjective/objective correlation than given by THD. *In Wireless World*, pp. 36–41.
- Bergman, P., Sköld, A., Västfjäll, D. et Fransson, N. [2009]**. Perceptual and emotional categorization of sound. *J. Acoust. Soc. Am.*, 126(6), 3156–3167.
- Bijmolt, T.H.A. et Wedel, M. [1995]**. The effects of alternative methods of collecting similarity data for multidimensional scaling. *International Journal of Research in Marketing*, 12(4), 363–371.
- Blauert, J. et Lindemann, W. [1986]**. Auditory spaciousness : Some further psychoacoustic analyses. *J. Acoust. Soc. Am.*, 80(2), 533–542.
- Boer, M.A., Nijmeijer, A.G.J., Schurer, H., Druyvesteyn, W. F., Slump, C.H. et Hermann, O.E. [1998]**. Audibility of nonlinear distortion in loudspeakers. *In Proc. AES 104th Convention*, n ° 4718.
- Bonebright, T.L. [1996]**. An investigation of data collection methods for auditory stimuli : Paired comparison versus a computer sorting task. *Behavior Research Methods, Instruments, and Computers*, 2(28), 275–278.

- Bonebright**, T.L. [2001]. Percptual structure of everyday sounds : a multidimensional scaling approach. *In Proceeding of the 2001 International Conference on Auditory Display*.
- Borg**, I. et **Groenen**, P. [1997]. *Modern Multidimensional Scaling, Theory and Applications*. Springer, New York.
- Bose**, A.G. [1968]. On the design, measurement, and evaluation of loudspeakers. *In Proc. AES 35th Convention*, n ° 622.
- Bryan**, M.E. et **Parbrook**, H.D. [1960]. Just audible thresholds for harmonic distortion. *Acustica*, 10, 87–91.
- Cabot**, R.C. [1990]. Audibles effects vs. objective measurements in the eslectrical signal path. *In Proc. AES 8th International Conference*, n ° 8.
- Cabot**, R.C. [1999]. Fundamentals of modern audio measurement. *J. Audio Eng. Soc.*, 47(9), 738–744, 746–762.
- Caclin**, A., **McAdams**, S., **Smith**, B.K. et **Winsberg**, S. [2005]. Acoustic correlates of timbre space dimensions : A confirmatory study using synthetic tones. *J. Acoust. Soc. Am.*, 118(1), 471–482.
- Canevet**, G., **Habault**, D., **Meunier**, S. et **Demirdjian**, F. [2004]. Auditory perception of sounds radiated by a fluid-loaded vibrating plate excited by a transient point force. *Acta Acustica United with Acustica*, 90(1), 181–193.
- Cermak**, G.W. [1979]. Exploratory laboratory studies of the relative aversiveness of traffic sounds. *J. Acoust. Soc. Am.*, 65(1), 112–123.
- Clark**, D. [1982]. High-resolution subjective testing using a double-blind comparator. *J. Audio Eng. Soc.*, 30(5), 330–338.
- Czerwinski**, E., **Voishvillo**, A., **Alexandrov**, S. et **Terekhov**, A. [2001a]. Multitone testing of sound system components - Some results and conclusions, Part 1 : History and theory. *J. Audio Eng. Soc.*, 49(11), 1011–1048.
- Czerwinski**, E., **Voishvillo**, A., **Alexandrov**, S. et **Terekhov**, A. [2001b]. Multitone testing of sound system components - Some results and conclusions, Part 2 : Modeling and application. *J. Audio Eng. Soc.*, 49(12), 1181–1192.
- de Santis**, M.E. et **Henin**, S. [2007]. Perception & thresholds of nonlinear distortion using complex signals, Section of Acoustics, Institute of Electronic Systems, Aalborg University.

-
- Farina, A.** [2000]. Simultaneous measurement of impulse response and distortion with a swept-sine technique. *In Proc. AES 108th Convention*, n ° 5093.
- Fielder, L.D.** et **Benjamin, E.M.** [1988]. Subwoofer performance for accurate reproduction of music. *J. Audio Eng. Soc.*, 36(6), 443–455.
- Fryer, P.A.** [1975]. Intermodulation distortion listening tests. *In Proc. AES 50th Convention*, n ° L-10.
- Fryer, P.A.** et **Millward, G.P.** [1980]. Audibility of amplifier clipping. *In Proc. AES 65th Convention*, n ° 1572.
- Gabrielson, A.** et **Lindstrom, B.** [1985]. Perceived sound quality of high fidelity loudspeakers. *J. Audio Eng. Soc.*, 33(1/2), 33–53.
- Gabrielson, A.** et **Sjögren, H.** [1972]. Detection of amplitude distortion in flute and clarinet spectra. *J. Acoust. Soc. Am.*, 52(2A), 471–483.
- Geddes, E.R.** et **Lee, L.W.** [2003a]. Auditory perception of nonlinear distortion. *In Proc. AES 115th Convention*, n ° 5891.
- Geddes, E.R.** et **Lee, L.W.** [2003b]. Auditory perception of nonlinear distortion - Theory. *In Proc. AES 115th Convention*, n ° 5890.
- Giordano, B.L.**, **Guastavino, C.**, **Murphy, E.**, **Ogg, M.**, **Smith, B.K.** et **McAdams, S.** [2011]. Comparison of methods for collecting and modeling dissimilarity data : Applications to complex sound stimuli. *Multivariate Behavioral Research*, 46(5), 779/811.
- Glasberg, B.R.** et **Moore, B.C.J.** [2002]. A model of loudness applicable to time-varying sounds. *J. Audio Eng. Soc.*, 50(5), 331–342.
- Graef, J.** et **Spence, I.** [1979]. Using distance information in the design of large multidimensional experiment. *Psychological Bulletin*, 80, 60–66.
- Grey, J.M.** [1977]. Multidimensional perceptual scaling of musical timbres. *J. Acoust. Soc. Am.*, 61(5), 1270–1277.
- Gunnarsson, V.** [2010]. *Assessment of nonlinearities in loudspeakers*, Rapport de master, Chalmers University.
- IEC 60268-13** [1998]. Equipements pour systèmes électroacoustiques - partie 13 : Essais d'écoute des haut-parleurs, International Electrotechnical Commission, Geneva, Switzerland.

- IEC 60268-5 [1989]**. Equipements pour systèmes électroacoustiques - partie 5 : Haut-parleurs, International Electrotechnical Commission, Geneva, Switzerland.
- ITU BS.1387-1 [2001]**. Method for objective measurements of perceived audio quality, International Telecommunication Union Recommendation, Geneva, Switzerland.
- ITU BS.1534-1 [2001]**. Method for the subjective assessment of intermediate quality levels of coding systems, International Telecommunication Union.
- ITU-T P.862 [2001]**. Perceptual evaluation of speech quality (PESQ) : An objective method for end-to-end speech quality assessment of narrow-band telephone networks and speech codecs, International Telecommunication Union Recommendation, Geneva, Switzerland.
- Karjalainen, M. [1982]**. Measurement of distortion in an audio signal channel based on psychoacoustic models. *In Proc. of NAS-82*, pp. 141–144.
- Kates, J.M. et Arehart, K.H. [2005]**. Coherence and the speech intelligibility index. *J. Acoust. Soc. Am.*, 117(4), 2224–2237.
- Kates, J.M. et Kozma-Spytek, L. [1994]**. Quality ratings for frequency-shaped peak-clipped speech. *J. Acoust. Soc. Am.*, 95(6), 3586–3594.
- Keele Jr., D.B. [1974]**. Low frequency assessment by nearfield sound-pressure measurement. *J. Audio Eng. Soc.*, 22(3), 154–162.
- Klippel, W. [2001]**. Speaker auralization - Subjective evaluation of nonlinear distortion. *In Proc. AES 110th Convention*, n° 5310.
- Klippel, W. [2003]**. Measurement of impulsive distortion, rumble and buzz and other disturbances. *In Proc. AES 114th Convention*, n° 5734.
- Klippel, W. [2006]**. Tutorial : Loudspeaker nonlinearities - Causes, parameters, symptoms. *J. Audio Eng. Soc.*, 54(10), 907–939.
- Klippel, W. [2011]**. Measurement of turbulent air noise distortion in loudspeaker systems. *J. Audio Eng. Soc.*, 59(7/8), 475–486.
- Kristoffersen, R., Kleiner, M. et Västfjäll, D. [1999]**. Subjectively perceived sound quality in audio systems as a function of distribution and number of loudspeakers used in playback. *In Proc. AES 106th Convention*, n° 4876.
- Kruskal, J.B. [1964]**. Non metric multidimensional scaling : a numerical method. *Psychometrika*, 29(2), 115 – 129.

-
- Kruskal, J.** et **Wish, M.** [1978]. *Multidimensional scaling*. Sage Publications, Newbury Park.
- Lavandier, M.** [2005]. *Différence entre enceintes acoustiques : une évaluation physique et perceptive*. Thèse de doctorat, Université Aix-Marseille II.
- Lavandier, M., Herzog, P.** et **Meunier, S.** [2004]. The restitution of timbre by loudspeakers in a listening room : perceptual and physical measurements. *In Proc. of AES 117th Convention*.
- Lavandier, M., Herzog, P.** et **Meunier, S.** [2008a]. Comparative measurements of loudspeakers in a listening situation. *J. Acoust. Soc. Am.*, 123(1), 77–87.
- Lavandier, M., Meunier, S.** et **Herzog, P.** [2005]. Perceptual and physical evaluation of differences among a large panel of loudspeakers. *In Proc. of Forum Acusticum*.
- Lavandier, M., Meunier, S.** et **Herzog, P.** [2008b]. Identification of some perceptual dimensions underlying loudspeaker dissimilarities. *J. Acoust. Soc. Am.*, 123(6), 4186–4198.
- Lemaitre, G., Susini, P., Winsberg, S.** et **McAdams, S.** [2007]. The sound quality of car horns : A psychoacoustical study of timbre. *Acta Acustica United with Acustica*, 93(3), 457–468.
- Lipshitz, S.P.** et **Vanderkooy, J.** [1981]. The great debat : Subjective evaluation. *J. Audio Eng. Soc.*, 29(7/8), 482 – 491.
- McAdams, S., Chaigne, A.** et **Roussarie, V.** [2004]. The psychomechanics of simulated sound sources : Material properties of impacted bars. *J. Acoust. Soc. Am.*, 115(3), 1306–1320.
- McAdams, S., Roussarie, V., Chaigne, A.** et **Giordano, B.L.** [2010]. The psychomechanics of simulated sound sources : Material properties of impacted thin plates. *J. Acoust. Soc. Am.*, 128(3), 1401–1413.
- McDermott, B.J.** [1969]. Multidimensional analyses of circuit quality judgments. *J. Acoust. Soc. Am.*, 45(3), 774–781.
- McKay, C.M., McDermott, H.J.** et **Clark, G.M.** [1996]. The perceptual dimensions of single-electrode and nonsimultaneous dual-electrode stimuli in cochlear implantees. *J. Acoust. Soc. Am.*, 99(2), 1079–1090.
- Moore, B.C.J.** [2003]. *An introduction of the psychology of hearing*. Academic Press, Londres, Royaume-Uni, 5ème édition.
- Moore, B.C.J., Glasberg, B.R.** et **Baer, T.** [1997]. A model for the prediction of thresholds, loudness, and partial loudness. *J. Audio Eng. Soc.*, 45(4), 224–240.

- Novák, A.** [2009]. *Identification of nonlinear systems in acoustics*. Thèse de doctorat, Université du Maine.
- Novák, A., Simon, L., Kadlec, F. et Lotton, P.** [2009]. Nonlinear system identification using exponential swept-sine signal. *IEEE Transactions on Instrumentation and Measurement*, 59(8), 2220–2229.
- Novák, A., Simon, L., Lotton, P. et Gilbert, J.** [2010]. Chebyshev model and synchronized swept sine method in nonlinear audio effect modeling. *In Proc. 13th Int. Conference on Digital Audio Effects (DAFx-10)*.
- Parizet, E. et Koehl, V.** [2012]. Application of free sorting tasks to sound quality experiments. *Applied Acoustics*, 73(1), 61–65.
- Petri-Larmi, M., Ojala, M. et Lammasniemi, J.** [1980]. Psychoacoustic detection threshold of transient intermodulation distortion. *J. Audio Eng. Soc.*, 28(3), 98–104.
- Poulton, E. C.** [1979]. Models of biases in judging sensory magnitude. *Psychological Bulletin*, 86(4), 777–803.
- Poulton, E. C.** [1982]. Biases in quantitative judgements. *Applied Ergonomics*, 13(1), 31–42.
- Preis, D.** [1984]. Linear distortions : Measurement, methods and audible effects - a survey of existing knowledge. *In Proc. AES 2th International Conference*, n ° C1005.
- Quaegebeur, N.** [2007]. *Vibrations non linéaires et rayonnement acoustiques de structures minces de type haut-parleur*. Thèse de doctorat, Ecole Polytechnique.
- Rao, V.R. et Katz, R.** [1971]. Alternative multidimensional scaling methods for large stimulus sets. *Journal of Marketing Research*, 8(4), 488–494.
- Rauhala, J., Ahonen, J., Tikander, M. et Karjalainen, M.** [2008]. Perception and physical behavior of loudspeaker nonlinearities at bass frequencies in closed vs. reflex enclosures. *In Proc. AES 124th Convention*, n ° 7346.
- Rogowitz, B.E., Frese, T., Smith, J.R., Bouman, C.A. et Kalin, E.** [1998]. Perceptual image similarity experiment. *Human Vision and Electronics Imaging III*.
- Scavone, G.P., Lakatos, S., Cook, P.R. et C.R., Harbke.** [2001]. Perceptual spaces for sound effects obtained with an interactive similarity rating program. *In Proceeding of the 2001 International Symposium on musical acoustics*.

-
- Scavone, G.P., Lakatos, S. et C.R., Harbke. [2002].** The sonic mapper : An interactive program for obtaining similarity ratings with auditory stimuli. *In Proceeding of the 2002 International Conference on Auditory Display.*
- Schmitt, R. [1995].** Audibility of nonlinear loudspeaker distortions. *In Proc. AES 98th Convention, n ° 4016.*
- Shorter, D. [1950].** The influence of high order products on nonlinear distortion. *Electronic Engineering, 22(4), 152–153.*
- Shurer, H. [2007].** *Linearization of electroacoustic transducers.* Thèse de doctorat, Université de Twente.
- Spence, I. et Domoney, D.W. [1974].** Single subject incomplete designs for nonmetric multidimensional scaling. *Psychometrika, 39(4).*
- Subkoviak, M. et Roecks, A.L. [1976].** A closer look at the accuracy of alternative data-collection methods for multidimensional scaling. *Journal of Educational Measurement, 13(4), 309–317.*
- Susini, P., McAdams, S. et Winsberg, S. [1999].** A multidimensional technique for sound quality assessment. *Acta acustica united with Acustica, 85(5), 650–656.*
- Takane, Y. [1982].** The method of triadic combinations : A new treatment and its applications. *Behaviormetrika, 11, 37–48.*
- Takane, Y., Jung, S. et Oshima-Takane, Y. [2009].** Multidimensional scaling. *In Millsap, R. E. et Maydeu-Olivares, A., éditeurs : Handbook of quantitative methods in psychology, pp. pp. 219–242.* Sage Publications, London.
- Tan, C.T., Moore, B.C.J. et Zacharov, N. [2003].** The effect of nonlinear distortion on the perceived quality of music and speech signals. *J. Audio Eng. Soc., 51(11), 1012–1031.*
- Tan, C.T., Moore, B.C.J., Zacharov, N. et Mattila, V.-V. [2004].** Predicting the perceived quality of nonlinearly distorted music and speech signals. *J. Audio Eng. Soc., 52(7/8), 699–711.*
- Temme, S., Brunet, P. et Keele Jr, D.B. [2009].** Practical measurements of loudspeaker distortion using a simplified auditory perceptual model. *In Proc. AES 127th Convention, n ° 7905.*
- Thibaut, J.-P. [1997].** Similarité et catégorisation. *L'année psychologique, 97(4), 701–736.*
- Thompson, W.J. [1994].** Chebyshev polynomials : After the spelling the rest is easy. *Computers in physics, 8(2), 161–165.*

- Toole**, F.E. [1982]. Listening tests - turning opinion into fact. *J. Audio Eng. Soc.*, 30(6), 431–445.
- Torgerson**, W.S. [1952]. Multidimensional scaling : I. theory and method. *Psychometrika*, 17(4), 401–419.
- Tsogo**, L., **Masson**, M.H. et **Bardot**, A. [2000]. Multidimensional scaling methods for many-object sets : A review. *Journal Multivariate Behavioral Research*, 35(3), 307 – 319.
- Voishvillo**, A. [2002]. Nonlinearity in horn drivers - Where the distortion comes from ? *In Proc. AES 113th Convention*, n ° 5641.
- Voishvillo**, A. [2011]. Assessment of nonlinearity in transducers and sound systems - from thd to perceptual models. *In Proc. AES 121st Convention*, n ° 6910.
- Voishvillo**, A., **Terekhov**, A., **Czerwinski**, E. et **Alexandrov**, S. [2004]. Graphing, interpretation, and comparison of results of loudspeaker nonlinear distortion measurements. *J. Audio Eng. Soc.*, 52(4), 332–357.
- Wang**, Z. et **Bovik**, A.C. [2009]. Mean squared error : Love it or leave it ? a new look at signal fidelity measures. *IEEE Signal Processing Magazine*, 26(1), 98–117.
- Young**, F.W., **Null**, C.H., **Sarle**, W.S. et **Hoffman**, D.L. [1982]. *Proximity and preference : problems in the multidimensional analysis of large data sets*, chapitre 1.2 - Interactively ordering similarities among a large set of stimuli, pp. 10–28. University of Minnesota Press.
- Zahorik**, P. [2009]. Perceptually relevant parameters for virtual listening. *J. Acoust. Soc. Am.*, 126(2), 776–791.
- Zoltogorski**, B. [1999]. Non-linear distortions of loudspeaker radiators in closed enclosures. *In Proc. AES 106th Convention*, n ° 4894.
- Zuccatti**, C. et **Bandiera**, M. [2009]. Dynamics distortion : Loudspeaker sensitivity modulation generated by audio signals. *J. Audio Eng. Soc.*, 57(5), 338–348.
- Zwicker**, E. et **Fastl**, H. [1999]. *Psychoacoustics - Facts and models*. Springer-Verlag, Heidelberg, Allemagne, 2ème édition.

Annexe A

Enceinte Tannoy System 600

Cette annexe présente les caractéristiques techniques de l'enceinte Tannoy System 600 que nous avons utilisée comme enceinte génératrice. C'est l'enceinte qui est la source des différents échantillons que nous avons créés pour les tests d'écoute. Cette enceinte a la particularité d'être coaxiale, un élément indispensable au protocole que nous utilisons. Cette caractéristique permet de mesurer indépendamment sur le même axe, face des deux haut-parleurs concentriques, le champ rayonné par le boomer et le tweeter. Le plan de l'enceinte est présenté la figure A.1. Lors de l'élaboration des enceintes virtuelles nous avons bouché les deux événements. La réponse en fréquence et les caractéristiques techniques fournies par le constructeur sont présentées respectivement sur la figure A.2 et dans le tableau A.1.

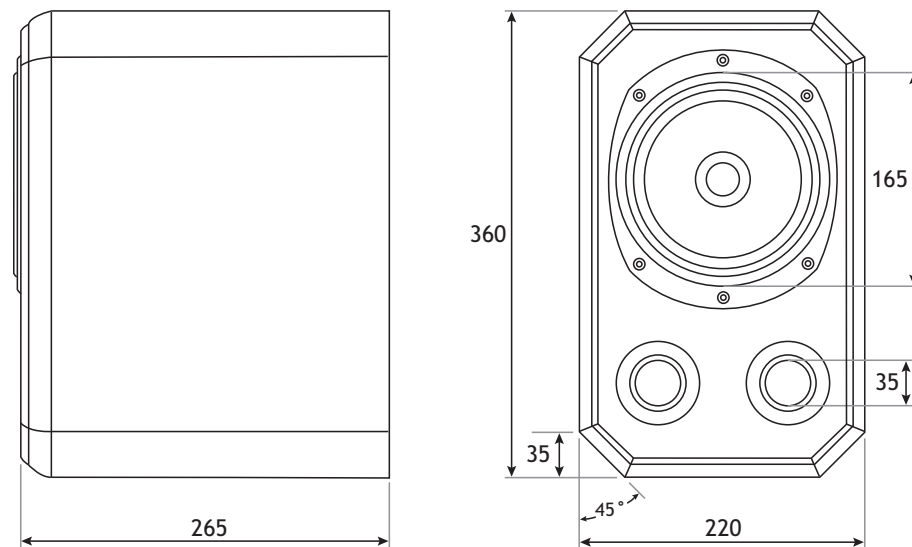


Figure A.1 – Plan de l'enceinte Tannoy System 600, vue de face et de droite.

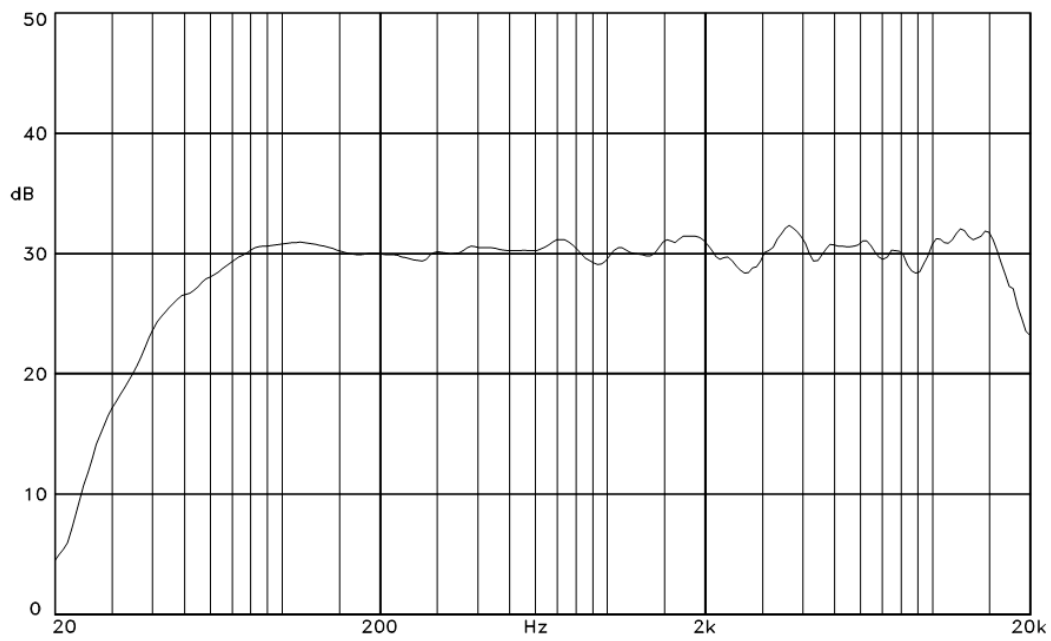


Figure A.2 – Réponse fréquentielle de l'enceinte Tannoy System 600.

Enceinte	
Réponse fréquentielle	52Hz - 20kHz
Puissance de l'amplificateur recommandée	50 à 150 Wms en 8Ω
Impédance Nominale	8Ω
Sensibilité	90dB SPL / 1W @ 1m
Distortion	<0.8%
Dispersion (@ -6dB)	90 ° conique
Fréquence de coupure	1800Hz
Cabinet	
Haut-parleurs	6.5" Dual Concentric type 1678
Conception pour les basses fréquences	Chargement bass-reflex optimisé, 13 litres
Construction cabinet	MDF (36mm) panneaux avant et arrière
Finition Cabinet	Vinyl Spray 'F'
Dimension Cabinet	220mm x 360mm x 267mm
Poids cabinet	7.5kg

Tableau A.1 – Caractéristiques techniques de l'enceinte Tannoy System 600.

Annexe B

Approximation à l'aide des polynômes de Chebyshev

Cette annexe présente des données supplémentaires sur l'utilisation des polynômes de Chebyshev. L'approximation d'une fonction sur la base des polynômes de Chebyshev retourne un certain nombre de coefficients dont il est possible de limiter le nombre en tronquant la série obtenue à un ordre plus bas. Nous avons défini comme critère de troncature une valeur de coefficient seuil : la série est tronquée au-delà du dernier coefficient de la série égal à cette valeur seuil. Nous avons estimé pour les 4 types de non-linéarités le nombre de coefficients de la série et l'erreur absolue entre la fonction et son approximation, pour différentes valeurs du paramètre a présents dans les équations les définissant.

Paramètre a	Seuil	Nombre de coefficients	Erreur
0.5	—	2000	$5.34 \cdot 10^{-14}$
	0.0001	106	$1.14 \cdot 10^{-4}$
	0.001	34	$7.88 \cdot 10^{-4}$
	0.01	10	0.0051
	0.1	4	0.0180
0.2	—	2000	$2.32 \cdot 10^{-14}$
	0.0001	104	$1.44 \cdot 10^{-4}$
	0.001	28	0.0012
	0.01	10	0.0067
	0.1	2	0.0646

Tableau B.1 – Valeurs du paramètre, nombres de coefficients et valeurs de l'erreur absolue pour l'approximation de la fonction Clipping.

Paramètres a	Seuil	Nombre de coefficients	Erreur
1	—	2000	$5.31 \cdot 10^{-14}$
	0.0001	112	$1.31 \cdot 10^{-4}$
	0.001	26	0.0011
	0.01	10	0.0091
	0.1	2	0.0372
0.2	—	2000	$0.59 \cdot 10^{-14}$
	0.0001	104	$1.39 \cdot 10^{-4}$
	0.001	36	0.0010
	0.01	6	0.0093
	0.1	2	0.0199

Tableau B.2 – Valeurs du paramètre, nombres de coefficients et valeurs de l'erreur absolue pour l'approximation de la fonction Zero Crossing.

Paramètres	Seuil	Nombre de coefficients	Erreur
$\forall a$	—	3	0
	0.0001	3	0
	0.001	3	0
	0.01	3	0
	0.1	3	0

Tableau B.3 – Valeurs des paramètres, nombres de coefficients et valeurs de l'erreur absolue pour l'approximation de la fonction Carrée.

Paramètres	Seuil	Nombre de coefficients	Erreur
$\forall a$	—	3	0
	0.0001	3	0
	0.001	3	0
	0.01	3	0
	0.1	3	0

Tableau B.4 – Valeurs des paramètres, nombres de coefficients et valeurs de l'erreur absolue pour l'approximation de la fonction Cubique.

Annexe C

Salle d'écoute et réponse impulsionnelle

Cette annexe présente la salle d'écoute dans laquelle Lavandier [2005] a effectué ses enregistrements. Nous les utilisons pour l'obtention de la réponse impulsionnelle de cette salle d'écoute.

Le fabricant d'enceintes Mosquito avait mis à disposition sa salle d'écoute, permettant ainsi à Lavandier [2005] de réaliser une série d'enregistrements au cours de sa thèse. La salle présentait la caractéristique d'avoir les murs qui n'étaient pas parallèles et différents traitements acoustiques étaient appliqués. Le sol de la salle était recouvert de moquette. Lavandier [2005] a mesuré une durée de réverbération de 0.4 s en moyenne pour des fréquences de 500 Hz à 5000 Hz. La prise de son binaurale AB-ORTF a été effectuée à 1 m du sol et à une distance de 2.40 m en face de l'enceinte. Le plan de la salle et la position de l'enceinte et des microphones sont présentés sur la figure C.1.

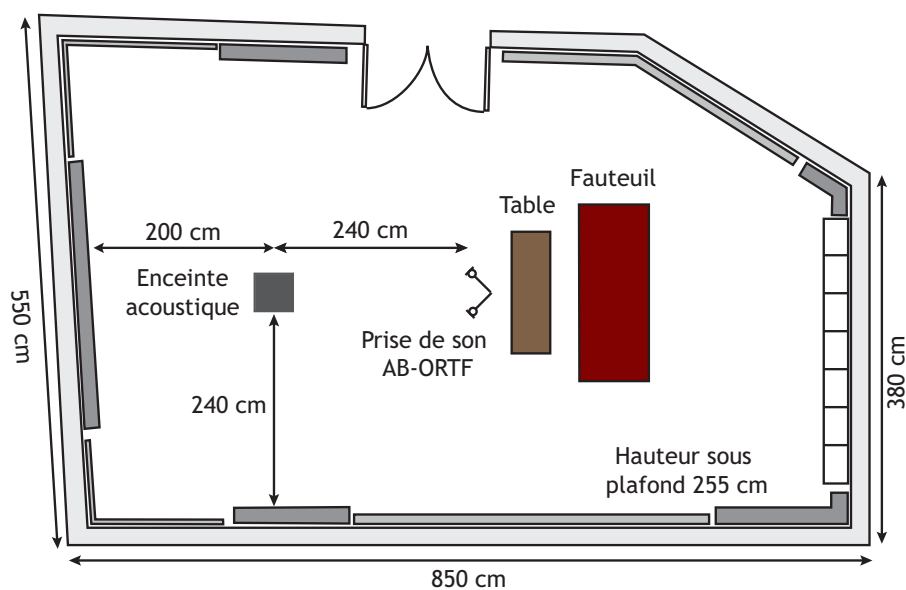


Figure C.1 – Plan de la salle d'écoute Mosquito avec les positions de l'enceinte et de la prise de son AB-ORTF.

Dans cette salle d'écoute, Lavandier [2005] a enregistré plusieurs morceaux de musique mais aussi des signaux techniques tels que du bruit blanc, du bruit MLS, et une série de sweep. Pour obtenir la RI de la salle, nous avons utilisé le sweep original généré par le module AURORA installé dans le logiciel Adobe Audition. Les caractéristiques du sweep généré étaient les suivantes :

- $f_1 = 10$ Hz
- $f_2 = 22050$ Hz
- Amplitude maximale = 8192 échantillons
- Durée : 4 s suivi d'un silence de 2s
- $f_e = 44100$ Hz

Nous avons utilisé la technique de convolution présentée par Farina [2000] pour récupérer la RI de la salle. La première étape consiste à calculer le filtre inverse qui correspond (retournement temporel du sweep original avec une modulation d'amplitude). Ensuite la convolution était réalisée entre ce filtre inverse et l'enregistrement stéréophonique du sweep. Le résultat de la convolution correspond à la RI stéréophonique de la salle. La figure C.2 montre la représentation temporelle de la RI de la salle Mosquito.

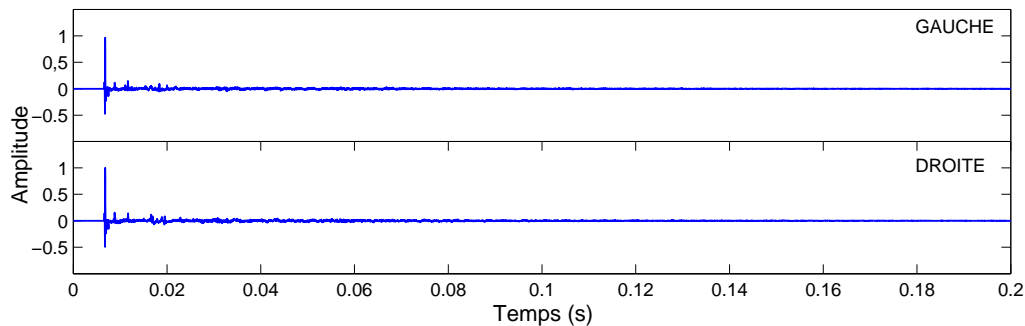


Figure C.2 – Représentation temporelle de la RI stéréophonique de la salle Mosquito.

Annexe D

Analyse Multidimensionnelle

Cette annexe apporte des précisions concernant l'analyse multidimensionnelle. L'analyse MDS est un outil statistique permettant de représenter un corpus de stimuli dans un espace subjectif à faible nombre de dimensions [Borg et Groenen, 1997]. Le résultat de l'analyse MDS de la matrice de dissemblance mesurée lors d'un test perceptif est une configuration spatiale dans laquelle les stimuli sont représentés par des points tels que les distances entre les stimuli coïncident le plus possible avec leurs dissemblances. L'espace obtenu permet donc de déterminer le nombre et la nature des dimensions représentant les critères de jugement des auditeurs.

Deux types d'analyse sont possibles. L'analyse dite métrique proposée par Torgerson [1952] considère que les données de dissemblances se comportent comme des mesures de distances entre chaque stimuli. La supposition que les dissemblances se comportent comme des distances peut être trop restrictive. Kruskal [1964] a développé l'analyse multidimensionnelle non métrique qui conserve uniquement l'information d'ordre entre les stimuli.

Pour la représentation, plusieurs modèles d'analyse sont disponibles en fonction de l'analyse des données de dissemblances que l'expérimentateur a choisi d'effectuer. Le modèle classique MDSCAL représente les dissemblances comme des distances dans un espace euclidien. Pour pouvoir analyser les données individuelles de dissemblance, le modèle INDSCAL permet de prendre en compte le fait que les auditeurs utilisent les mêmes dimensions pour différencier les stimuli mais ne les pondèrent de la même manière. Le modèle CLASCAL permet de trouver des groupes d'individus dont les jugements diffèrent du corpus principal d'auditeurs. Le modèle permet de pondérer les jugements de ce groupe d'individu. Les deux modèles INDSCAL et CLASCAL permettent d'obtenir davantage d'information que le modèle classique MDSCAL mais nécessitent d'avoir une matrice de dissemblance complète par auditeur. Le modèle MDSCAL permet de représenter les dissemblances représentant les jugements d'un auditeur "moyen".

Méthode utilisée

Dans les différents tests perceptifs, nous avons utilisé un programme de MDS classique MDSCAL élaboré par Torgerson [1952]. C'est un modèle métrique de MDS qui nécessite d'analyser une matrice de dissemblance, complète et symétrique. Pour ce modèle les distances entre les points de l'espace sont déterminés pour qu'elles correspondent au mieux aux valeurs de dissemblances. Nous avons aussi utilisé l'algorithme SMACOF (Scaling by Majorizing a Complicated Function) qui permet d'optimiser la proximité entre les données expérimentales de dissemblances et les distances de l'espace perceptif.

Nombre de dimensions

Pour définir le nombre de dimensions de l'espace perceptif, Borg et Groenen suggèrent de tracer l'évolution du stress en fonction du nombre de dimensions. La fonction stress est définie comme une fonction qu'il faut minimiser pour que les dissemblances estimées et les distances dans l'espace MDS coïncident au mieux. Cette fonction est généralement décroissante de manière monotone ce qui signifie que le stress diminue avec le nombre de dimensions ajoutées à l'espace. Pour déterminer le nombre de dimensions, il faut d'après Borg et Groenen repérer un "coude" sur la courbe qui symbolise un point où la diminution du stress est moins prononcée. Ce point correspond au nombre de dimensions à choisir pour la représentation et indique que l'ajout de dimensions supplémentaires ne diminuerait pas le stress de façon conséquente.

Le nombre de dimensions obtenu par la courbe de stress représente seulement une indication. Il faut pouvoir leur donner une interprétation. L'interprétation d'une représentation MDS signifie faire le lien entre des propriétés géométriques et une information représentative des objets symbolisés par les points. Le but est d'identifier un critère significatif qui a permis aux sujets de différencier clairement les stimuli et qui conduit à leur position sur une dimension. Ainsi, l'expérimentateur reste libre de définir le nombre de dimension de l'espace perceptif en fonction de l'interprétation qu'il est possible d'en faire.

Nous présentons les courbes de stress des différents tests d'écoute que nous avons effectués (figures D.1 , D.2, D.3 et D.4). Pour chaque test, nous avons fait apparaître le nombre de dimensions choisi pour représenter l'espace perceptif par une croix sur la courbe de stress.

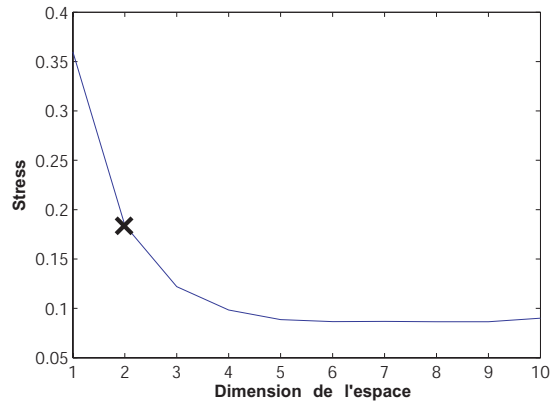


Figure D.1 – Détermination du nombre de dimensions lors de notre test sur 12 enceintes réalisé pour la validation de la méthode de comparaison à une référence.

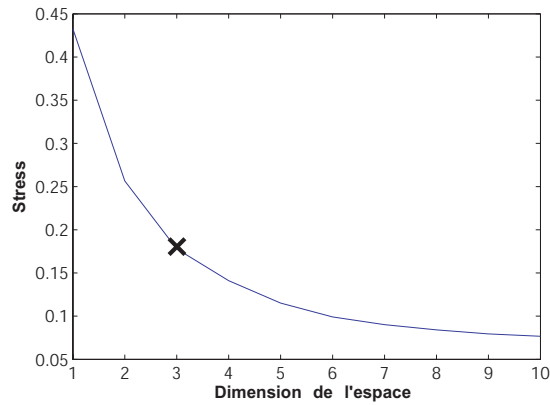


Figure D.2 – Détermination du nombre de dimensions lors de notre test sur 37 enceintes réalisé pour la validation de la méthode de comparaison à une référence.

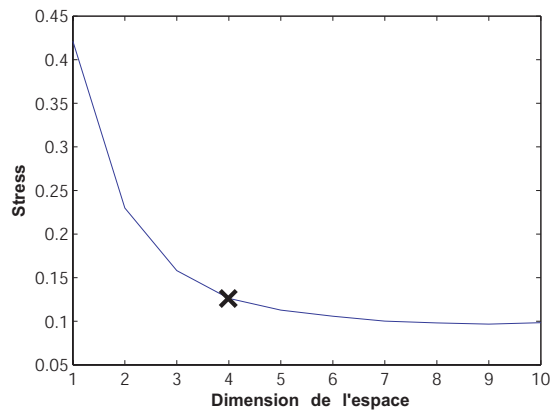


Figure D.3 – Détermination du nombre de dimensions lors de notre test sur 31 enceintes réalisé pour évaluer la distorsion non-linéaire.

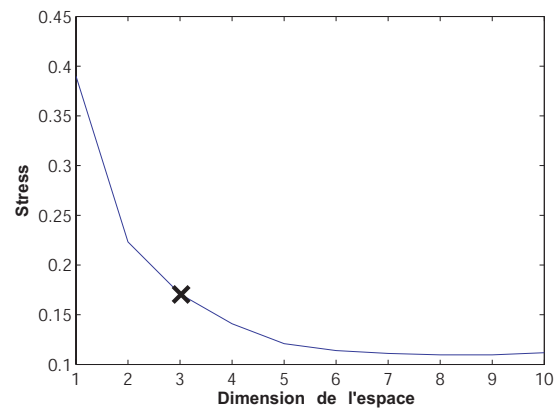


Figure D.4 – Détermination du nombre de dimensions lors de notre test sur 19 enceintes réalisé pour évaluer la distorsion non-linéaire.

Annexe E

Mesure de la distorsion du casque utilisé pour les tests d'écoute

Pour retransmettre les signaux que nous avons créé, nous utilisons un casque électrostatique Stax SR Lambda Professional. Étant donné que nous cherchons à caractériser la distorsion non-linéaire, il est important de pouvoir évaluer celle introduite par le casque. Nous avons utilisé la méthode d'identification de Novák [2009] pour identifier les non-linéarités générées lorsqu'un sweep est envoyé au travers du casque. La figure E.1 présente le protocole expérimental pour enregistrer le son rayonné par chaque voie du casque.

Les caractéristiques de la restitution et de l'enregistrement sont les suivant :

- Sweep : $f_1=100\text{Hz}$, $f_2=7000\text{Hz}$.
- Enregistrement et restitution synchrone : Convertisseur audio Fostex VC-8 et une carte son RME DIGI9652 à une fréquence d'échantillonnage.
- Oreille artificielle : B&K type 4152, microphone B&K 4144.

Nous avons mis au point un banc de mesure permettant de suspendre le casque et de garder une distance entre les écouteurs de 16 cm. La pression ainsi appliquée sur l'oreille artificielle correspond à la pression exercée par l'arceau du casque dans des conditions d'utilisation normales. Les écouteurs du casque Stax étant des écouteurs supra-auriculaire, nous avons placé entre l'écouteur et l'oreille artificielle une plaque percée en son centre pour assurer un couplage représentatif d'une utilisation sur une tête. Le diamètre du trou correspond au diamètre du microphone de l'oreille artificielle (1 pouce). La figure E.2 présente le banc mis au point pour cette mesure.

Pour évaluer la distorsion introduite par le casque, nous avons estimé les différences de niveaux entre les harmoniques non-linéaires 2 et 3 par rapport à l'harmonique 1 pour différents niveaux de restitution. Les résultats de cette identification sont donnés dans le tableau E.1.

Les valeurs de différence de niveaux sont relativement élevées pour les niveaux de restitution que nous avons testés. L'harmonique 2 est au minimum 60dB au dessous de l'harmonique 1 et

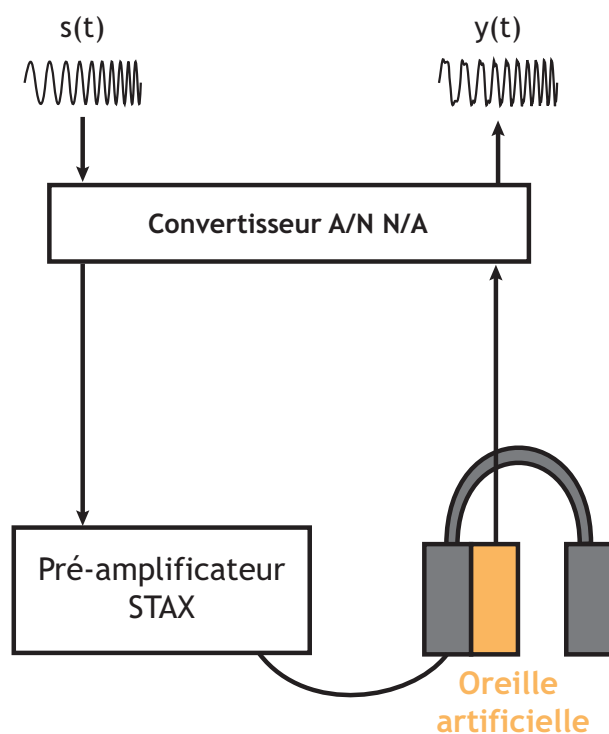


Figure E.1 – Protocole pour l'identification du casque Stax. L'enregistrement est fait avec une oreille artificielle et le convertisseur permet une mesure synchrone entre l'entrée et la sortie du système.

	GAUCHE			DROITE		
Niveau oreille artificielle (dB)	88.8	94.8	99.7	88.8	94.8	99.7
Rapport H2/H1 (dB)	-70	-64	-60	-72	-66	-61
Rapport H3/H1 (dB)	-83	-88	-87	-82	-88	-87

Tableau E.1 – Valeurs des niveaux entre harmoniques pour différents niveaux de restitution mesurées sur les deux voies du casque Stax SR Lambda Professional.

la différence entre les harmoniques 1 et 3 dépasse 80 dB pour tous les niveaux testés.

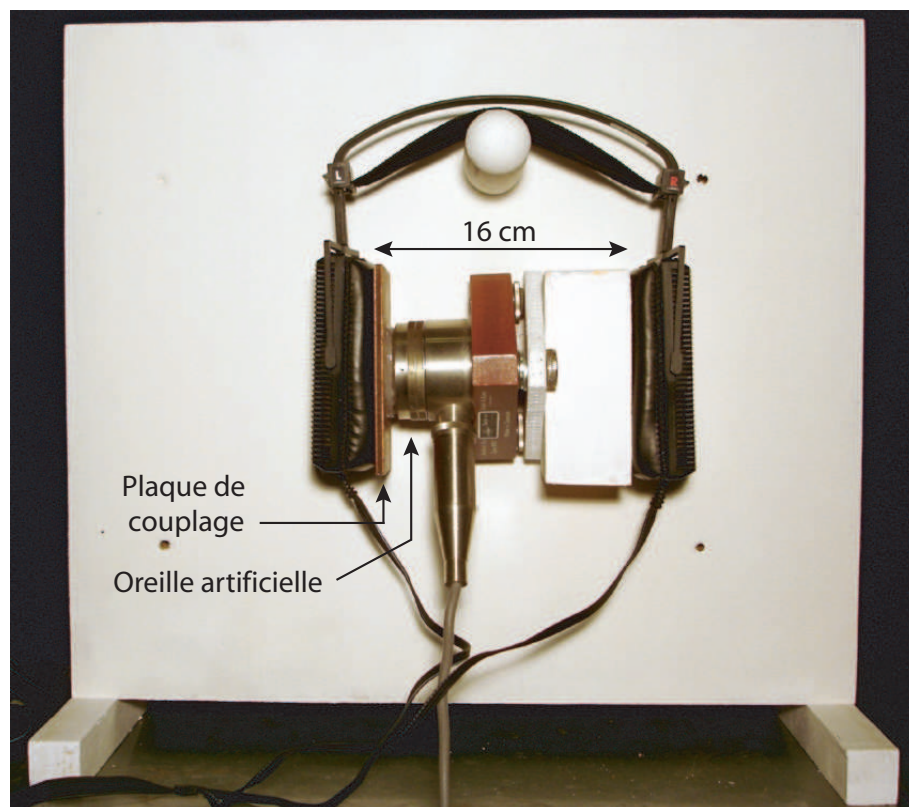


Figure E.2 – Photographie du dispositif expérimental pour évaluer la distorsion introduite par le casque.